Hearing Buildings: A Program for Aural Representation of Architectural Space

COLIN RIPLEY
Ryerson University

Among the critical fault-lines within architectural practice and discourse is that which privileges the optical, conceiving of architecture as primarily a visual art form. Despite the multi-sensorial, embodied nature of our being in space, architectural discourse is largely silent where senses other than the visual are impacted.

This emphasis on the visual within architectural practice and discourse has been the case at least since the renaissance, when the increased reliance on drawing as a technique leading to the production of buildings shut out, to a certain extent, the other senses. The work by Marin Mersenne and Athanasius Kircher, contemporaries of Descartes, who adopted the latter thinker's theories of optical lenses to develop theories of acoustic lenses (which took, naturally, the scale of buildings) appears to have been an anomaly, and one outside of the architectural mainstream. Even this work suffers from the problem of representation, as the acoustic ideas can only, of course, be rendered in visual terms. The ensuing publication of architectural picture-books inevitably added impetus for this transition of architecture into a visual art form, as the image of a building took on unprecedented importance. Certainly by the nineteenth century this paradigm shift was complete. Thus we have a situation in which the introduction of a new technology of representation led to an ontological shift in the understanding of the object being represented[1].

Modern technologies of representation have done little to reverse this shift. Indeed, despite for example Marshall McLuhan's recognition of the multi-sensual possibilities of electric media in the form of the "Five Senses Sensorium",[2] and of the tactile and physically engaged nature of computer use (and to a lesser extent television watching), one could make the case that, to date, computational technologies in architectural practice have tended to increase the reliance on the visual. One could argue, for example, that the ever-increasing reliance on computer models over and above the building of physical models results in a shift away from the haptic and textural qualities of the traditional model. Further, since computer models are typically presented in the form of either still or animated images, the interactive bodily nature of the contemplation of a basswood model, for example, is lost. High-quality, "photo-realistic" rendering of materials has similarly shifted a contemplation of texture from a haptic and conceptual exercise, to one that is primarily visual. Even movement through a space which has not yet been built, which once depended on conceptualization and narrative, taking most often the form of a guided tour through a set of plans or other drawings, has with the computerized walk-though become a self-referential visual experience. Further, it is a disembodied experience: while previously one's body might have moved in one's imagination through the space, one now watches the scene from the point of view of a camera moving through the space. One does not imagine one is the camera.

Despite their tendancy towards a disembodied reading of architectural projects, these methods remain extremely valuable tools for architectural production, provided one is aware of the conceptual shifts which these tools engender and of what is lost in the process. However, there remain potentialities inherent in computer technology, not fully exploited in architectural practice, for the more fully embodied representation of building projects. For example, recent research seeks to adapt computer game engines for the presentation of architectural models. This research seeks to exploit the hands-on, tactile nature of computer use (and particularly of computer gaming) in order to break the spectator/spectacle divide inherent in traditional architectural animations (this is of course the same distinction as that between a film-goers experience and that of a gamer). In principle, the user of such a game-based system should come to feel as though he or she is to some extent physically present in the architectural model, just as an experienced CAD operator can come to feel physically present in a drawing. Anyone who has tried to view a CAD drawing on a computer screen as someone else pans and zooms through the drawing will understand this principle.

A further step in this same direction would be the addition of gravity to such systems. Indeed, many game engines do already incorporate a gravity component, at least for the characters in the game (although the environment, ie the building, remains weightless). The addition of gravity would, of course, tend to enhance the embodied experience of the architectural game. Further, the extension of the gravity system to the architectural environment has obvious potential benefits for architectural design, and could be a particularly useful tool for architectural education.

One other way in which computers can aid in developing a more fully embodied representation of building projects is through the representation of sound. In current practice, sound is generally viewed within architecture in one of two ways. The first and most typical takes an instrumental view of sound, which sees it as an element to be controlled using acoustic engineering. This is the view of sound taken, for example, in the design of concert halls, meticulously designed for their acoustic properties and often tunable to meet the requirements of a performance. This is also the view taken on a more prosaic level when an architect concerns him or herself with noise control – for example, specifying appropriate STC ratings for suite demising walls in an apartment building, or planning for sufficient soft surfaces in a busy restaurant to ensure an appropriate reverberation time.

The second approach typically taken towards sound is a metaphorical approach: this is, in essence, Schelling's notion of architecture as "frozen music". In practice, this typically means using musical structure to derive architectural form. Indeed, one of the few works published in the last half-century on the relationship of architecture and music from within the architectural community is *Pamphlet Architecture 16*, edited by Elizabeth Martin, which takes as its title *Architecture as a Translation of Music*[3]. That is, music must first be translated into a visual terms in order to be considered architecture. This approach is less common than the instrumental approach; we can see it, for example, in the window patterns at La Tourette by Iannis Xenakis, or in Daniel Liebeskind's *Chamberworks* series of drawings. This is also the approach commonly taken by architectural students when working with musical themes.

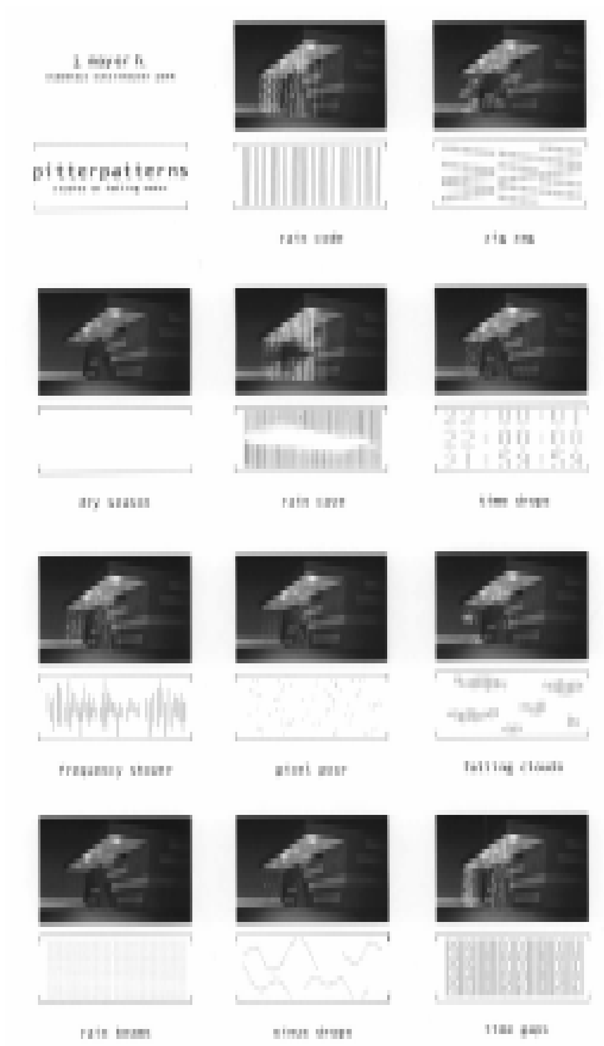
In short, architecture remains primarily a visual medium – not because buildings exist solely in the visual realm (they don't), but because architects draw. Even those few architects who do consider the sounds that their buildings make are limited by this problem of representation. For example, when determining the sound patterns of water drops falling from the computer-controlled roof drains of his *Charnhauser Park Town Hall*, architect Jurgen Mayer H. was limited to visual representations of the sound(Figure 1). In order for sound to play a more complete and direct role in architectural design – that is, for architects to recognize that all buildings and all spaces make noise and have acoustic properties and that these properties can be manipulated for architectural as well as programmatic and functional ends – it will be necessary to first develop a means of representing sound within buildings. In other words, we should be able to represent our virtual spaces not only visually, but also aurally, allowing clients, designers, students etc. to hear a proposed building as well as see it.

Just as the development of 3-D CAD systems (or, for that matter, of perspective drawing) produced shifts in architectural form, we would expect that the incorporation of the ability to represent sound in architectural practice would produce corresponding shifts in design. This would, at the very least, bring acoustics more fully into architectural discourse.

ELECTROACOUSTIC ARCHITECTURAL REPRESENTATION SYSTEM (EARS): DEVELOPMENT PROGRAMME

Given the multimedia capabilities of current computers, the development of a system for the representation of acoustical properties of architectural spaces should not in principle be a difficult task. In fact, most if not all of the basic components of the

Figure 1. Pitterpatterns. Image courtesy Jurgen Mayer H.



system are already available or have been significantly developed. The task will be to combine and integrate these components, which are found in such diverse fields as Sound Art, Performance Art, Computer Gaming, and Acoustical Engineering.

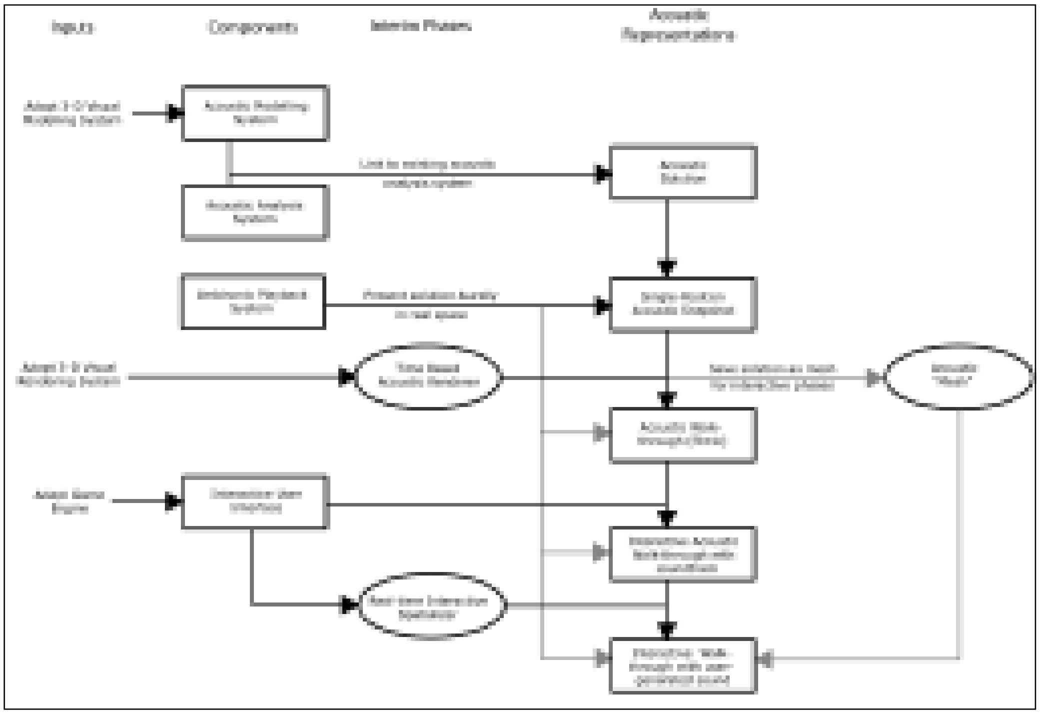
The end product of this programme will be an interactive system which allows users to move at will through the space of the building, hearing sounds produced by the building (and by other people potentially in the building), and also making their own sounds, which would then be re-presented to them in real time in accordance with the acoustic properties in play at their given location in the building. The system would be made up of four essential components (Figure 2).

First, an **Acoustic Modelling System** would have to be developed that could model, in addition to the standard properties of materials, their acoustic properties. Since the intention of this project is to move a consideration of sound into the architects' offices, and not simply be in the realm of the acoustic engineers, this would be developed as a plug-in modification to a popular, commercially available product. Ideally, the same system should be used for acoustic and optical modelling, although differences in the physical properties of light and sound coupled with the necessity, at least in the current technological environment, to keep the acoustic model as resource-light as possible complicates this idea. Since components that are of a scale less than about one wavelength are essentially transparent to sound, and since the audible wavelengths for the human ear range from about fifty feet down to about three-quarters of an inch, the acoustic model can have considerably less detail than a visual model intended for a high-quality rendering. If one is willing to do away with very high notes, the model can be further simplified: the highest note on the standard piano has a wavelength in air about equal to the depth of a 2x4.

As with visual modelling, the basic characteristics of a space that would have to be modelled in acoustic terms are geometry and material or surface properties. It is obvious that the shape size and configuration of a room, as well as the reflective, absorptive, and transmissive characteristics of components of the room will affect the way the space sounds as much as the way it looks. This too is complicated by the fact that materials will interact differently with sound at different frequencies; in principle, however, this is no different from the variable interaction with light at different frequencies, that is, colour. Systems used for applying colour to materials should be easily modified to deal with pitch.

An'**Acoustic Analysis System** or *sound-field modeller* would then have to be found or developed capable of reading the model information and determining the acoustic properties of the space. Acoustic analysis programs exist and are commonly used by acoustic engineers, but will need to be integrated with the modelling software. Such programs also tend currently to be both expensive and somewhat cumbersome to use, both of which are issues which must be overcome if this system is to be regularly used, again, by architects in the course of their practice.

Figure 2. EARS Project Flowchart



There are essentially two main algorithms used in the acoustic analysis of spaces; most products available on the market use a combination of the two systems[4]. The first of these systems, familiar to architects who use visual rendering programs, is ray-tracing, in which the various sounds reaching the receiver either directly or by reflection off the geometry of the space are calculated. At each reflection, the properties of the reflected sound must be altered to take account of the properties of the reflecting material, and the calculation must be done for each position within the space (unless the position of the listener is static and known), making this a slow calculation.

The second algorithm commonly used is that of virtual sources. Rather than trace all the reflected rays for each position in the space, each sound source is paired with a mirror image, reflected in each surface. The virtual sources then receive their own virtual sources, and so on, until the desired level of accuracy is reached. Each virtual source is modified in order to take account the properties of its'“mirror”. The resultant sound at any source is then just the sum of all the *direct* sounds heard from the real and virtual sources, with no reflections needing to be considered. This algorithm has clear advantages when dealing with interactive systems, as the majority of the calculations can, at least in principle, be done in advance, delivering something akin to a “rendered mesh” to the presentation.

Neither of these algorithms deals particularly well with diffraction of sound – its ability to bend around obstacles. Diffraction, of course, is not an issue at human scale for light, so there are no models in the world of visual rendering which would be particularly helpful. What’s more, since diffraction plays a relatively small role compared to direct sound and reflection in determining the acoustic properties of closed spaces such as concert halls, developing systems for dealing with diffraction has not been a priority for acoustic engineers. It will, however, be necessary to include diffraction in order to develop a realistic model for the rendering of non-closed architectural spaces.

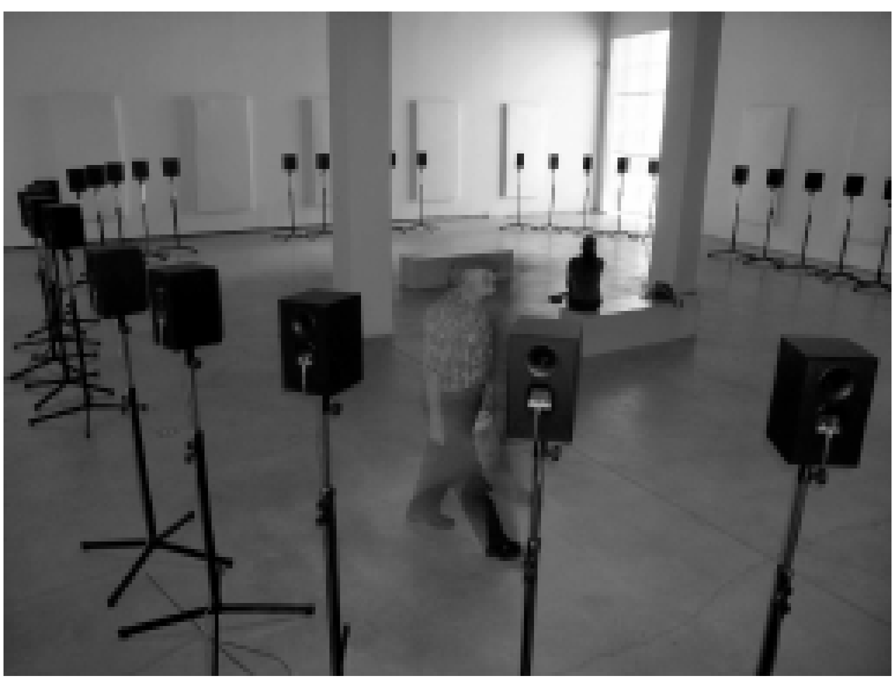
Finally, both of these algorithms deliver a static sound-field model of a space. In order to develop the system into full interactivity, it will be necessary to develop a time-based system which takes account of moving sound sources. This will certainly be needed if one is to allow, for example, clients to'“move” through the space, hearing their own voices.

A real-world **Ambisonic Playback System** will have to be integrated with the above. In principal, this would consist of an acoustically absorptive room in which is placed an array of loudspeakers, along with appropriate electronics to parse the sound into as many channels as is required for true ambisonic (that is, spatially fully real) playback. Such a space has recently been developed by Arup Acoustics in their *SoundLab*[5]. Arup's room, like the vast majority of acoustic design tools currently in use, is intended to aid in the design of acoustically sensitive spaces such as performance halls. Arup's room has been calibrated by measuring the acoustic properties of a large number of performance spaces; a proposed design can be compared acoustically, aurally, to over a hundred important halls, using a ten-speaker ambisonic system.

The principle is straightforward: by adjusting the relative intensity of a sound coming out of two speakers, the sound can be made to'"appear" to come from any position on a line between the two speakers. This is the typical stereophonic effect. One could, in principle, get 360 degree sound placement – that is, the ability to locate sound at any point around a plane - through the use of three speakers arranged in a triangle. Again in principle, this system could be extended to full spherical location of sound through the addition of a fourth speaker, with the four speakers arrayed in a tetrahedral fashion. The relative inability of humans to locate the height of sounds allows the vertical axis to be underdeployed in such a scheme.

In practice, ambisonic environments tend to use more than four speakers. A wide variety of such systems exist. Probably the clearest system in terms of geometrical articulation is the Morrow Sound Cube[6], which used eight speakers arrayed at the vertices of a cube, with the receiver or listener at the centre of the cube. Eight speakers are also the norm in electroacoustic performance, although most commonly the speakers are set in a circle, usually somewhat above the ear level of the audience (this is probably due more to the logistics of setting up an eight-channel system in what is usually an improvised space than to any ideal arrangement). Janet Cardiff's work of sculptural sound art *40-Part Motet* uses, not surprisingly, forty speakers to reproduce the sound of a choir singing in Salisbury Cathedral. The sound art of Bernhard Leitner uses a large number of speakers, in order to make audible lines, arcs, and circles of sound[7]. Home theatre systems, which represent one end of the ambisonic spectrum, use five, six, seven, or more speakers, distributed in a very specific manner.(Figure 3)

Figure 3. An Ambisonic Environment: Janet Cardiff's 40 Part Motet *at the Power Plant, Toronto. Photo courtesy of The Power Plant*



Clearly, the addition of more speakers than the minimum four (or three, to take for the moment just the planar localisation) has the advantage of increasing the potential resolution of the resulting sound field. The ideal case of an infinite number of speakers arrayed in a circle—or one continuous ring-shaped speaker—would of course give an ideal localisation. Smaller systems, even three-speaker systems, are capable of giving ideal localisation under three conditions related to the receiver: the receiver must be infinitely small, non-moving, and in a well known, pre-determined location. Since human hearing – and particularly the ability to locate sound – results from the presence of *two* receivers (two ears), the first of these conditions cannot be met, although the ears are close enough to reasonably approximate a point; resultant sound from a few speakers will localise reasonably the same over the area in which the two ears are located. However, it is clear that sound will localise differently if heard from the centre of an ambisonic space, or if heard from a position near one of the speakers. Sound directionality is determined primarily by the time difference with which the sound signal reaches the two ears, with the difference in sound intensity as recorded at the two ears playing a secondary role. In most circumstances, where the distance between the ears is small relative to the distance from ears to sound source, the difference in intensity is negligible. If however the ears are close to the sound source, as may be the case in an ambisonic environment, the intensity factor will play a larger and potentially distorting role. Increasing the number of speakers will thus have the effect of increasing the area over which this acceptable similarity of localisation is maintained, although it can be maintained only in areas which are relatively distant from all speakers.

Home theatre systems are an interesting case in that they have other media-related properties effect the speaker array. A 5.1 system, for example, such as Dolby 5.1, makes use of six speakers: Left Front, Right Front, Left Surround, Right Surround, Subwoofer (the .1), and Centre. Centre is used in home theatre situations mostly for dialogue or other non-spatialised sounds. In the newer 6.1 systems, a “Back” channel is added. Whether a centre channel is required in the EARS system is unknown at this point.

Finally, an **Interactive User Interface** will need to be integrated with these systems. Users will then be able to experience a prospective building both visually and aurally, not from a static position, but as an actor in the game. Not only will participants be able to move about at will in the building, seeing images and hearing sounds appropriate to that space, but they will also be able to hear *themselves* as though they are actually in that space. The result will be an acoustically augmented space, similar to spaces found in the work of sound artists such as David Rokeby, whose “Echoing Narcissus” project returns the speaker’s voice to him or herself, altered[8].

Phase I of EARS will be the production of acoustic renderings of individual spaces, using computer-generated sounds placed within the space by the system. Phase II will extend these renderings in time in a manner akin to typical architectural walk-throughs. This phase will produce acoustically accurate “films” of a building project, still using computer-generated sounds, and presented in a linear fashion.

Phase III will pursue interactivity in two parallel streams. On one side, the rendering system developed in phase II will be combined with a computer gaming system in order to allow free movement (and hearing) within the space. Audio systems for computer games typically are comprised of a number of components. For example, the audio system for Jurassic Park II[99] is comprised of four components: ambient sounds, such as bird calls, which are manifested as a grid of point sources, in order to approximate a regular distribution; localised sounds, such as the noise of frogs, located in a marshy piece of terrain; event sounds, such as the roars of the attacking dinosaurs; and external sound, such as instructions, beeps, etc. These sounds are deployed over the typical 5.1 or 6.1 home theatre system n accordance with their location in the game world, as well as that of the gamer; the effects of the geometry of the game, however, is for the most part neglected. In principle, an EARS-based project would carry two varieties of sounds: sounds made by the building, which are static in terms of position; and sounds made by other people in the building, which may be moving, or may be able to be approximated by a mesh. The receivers would be allowed to vary in position. On the other side, a third variety of sound will be added to the system: that is, sounds generated by the user in real-time and played back as though in the virtual space. Such systems for real-time sound manipulation and playback also already exist, and are currently used by a number of performance artists.

As I mentioned above, many of the components required for this system can be found in avant-garde practices in music and sound art. It is in the nature of such avant-garde artistic practices to act as "fundamental research", finding their way into architectural practice with a lag of about a generation, and with unpredictable results. It is probably too much to hope that a system as proposed here would do much to counteract six centuries of visual bias in architectural representation, in an increasingly visual world, but, as they say, it ain't over till the fat building sings.

NOTES

[1]For a further discussion of vision and visual modes of representation, see for example Jonathan Crary, *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century* (Cambridge, Mass.: MIT Press, 1990), Hal Foster, ed., *Vision and Visuality*, vol. 2, *Discussions in Contemporary Culture* (Seattle: Bay Press, 1988), Karsten Harries, *Infinity and Perspective* (Cambridge, Mass.: MIT Press, 2001), Alberto Pérez Gômez and Louise Pelletier, *Architectural Representation and the Perspective Hinge* (Cambridge, Mass.: MIT Press, 1997). These are just a few examples of a wide literature on this subject.

[2] Marshall McLuhan, "Inside the Five Senses Sensorium," *Canadian Architect* (1961).

[3] Elizabeth Martin, *Architecture as a Translation of Music*, *Pamphlet Architecture; 16* (New York: Princeton Architectural Press, 1994).

[4]Jens Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, Rev. ed. (Cambridge, Mass.: MIT Press, 1997).

[5] See the Arup SoundLab web site, http://www.arup.com/acoustics/soundlab/arup_soundlab.htm.

[6] For information on the Morrow Sound Cube, see the Charles Morrow Company web site, http://www.cmorrow.com.

[7] See Bernhard Leitner, *Geometry of Sound* (Stuttgart: Reihe Cantz, 1997).

[8] For information on the work of David Rokeby, see David Rokeby, *David Rokeby* (Oakville, ON: Oakville Galleries, 2004). Also see David Rokeby's web site, http://homepage.mac.com/davidrokeby/home.html.

[9] Stephan Shutze, "The Creation of an Audio Environment as Part of a Computer Game World: The Design for *Jurassic Park - Operation Genesis* on the Xbox[tm] as a Braod Concept for Surround Installation Creation," *Organized Sound: An International Journal of Music Technology* 8, no. 2 (2003).

SELECTED REFERENCES

Blauert, Jens.

Spatial Hearing: The Psychophysics of Human Sound Localization. Rev. ed. Cambridge, Mass.: MIT Press, 1997.

Crary, Jonathan.

Techniques of the Observer: On Vision and Modernity in the Nineteenth Century. Cambridge, Mass.: MIT Press, 1990.

Foster, Hal, ed. *Vision and Visuality*. Edited by Hal Foster. Vol. 2, *Discussions in Contemporary Culture*. Seattle: Bay Press, 1988.

Harries, Karsten. *Infinity and Perspective*. Cambridge, Mass.: MIT Press, 2001.

Leitner, Bernhard. *Geometry of Sound*. Stuttgart: Reihe Cantz, 1997.

Martin, Elizabeth. *Architecture as a Translation of Music*, *Pamphlet Architecture; 16*. New York: Princeton Architectural Press, 1994.

McLuhan, Marshall. "Inside the Five Senses Sensorium." *Canadian Architect* (1961): 49-54.

Pérez Gômez, Alberto, and Louise Pelletier. *Architectural Representation and the Perspective Hinge*. Cambridge, Mass.: MIT Press, 1997.

Shutze, Stephan. "The Creation of an Audio Environment as Part of a Computer Game World: The Design for *Jurassic Park - Operation Genesis* on the Xbox[tm] as a Braod Concept for Surround Installation Creation." *Organized Sound: An International Journal of Music Technology* 8, no. 2 (2003): 171-80.