Proc. of the EAA Joint Symposium on Auralization and Ambisonics, Berlin, Germany, 3-5 April 2014

AURALIZATION AS AN ARCHITECTURAL DESIGN TOOL

Matthew Azevedo[1], Jonah Sacks[2]

[1]Consultant in Acoustics,
Acentech Inc.
Cambridge, MA USA
mazevedo@acentech.com

[2] Sr. Consultant in Acoustics,
Acentech Inc.
Cambridge, MA USA
jsacks@acentech.com

ABSTRACT

Auralization provides a valuable tool that allows architects, building owners, and other decision-makers to directly experience the aural implications of design decisions and allows them to make more informed choices. Standard numerical metrics are difficult to relate to aural phenomena without significant practice and frequently fail to capture acoustical issues that are essential to the basic functionality of spaces. Consultants at Acentech have been using auralizations of full soundscapes including many independent sources as design and communication tools for a variety of projects including atria, lecture halls, theaters, and performance spaces. These auralizations have included natural speech and electro-acoustic reinforcement, crowd activity, interactions between PA systems and room acoustics, HVAC noise, wall and window transmission, and the subjective effects of sound masking. In general, clients find the experience of listening to their as-yet unbuilt spaces to be exciting and useful. Though most are not trained listeners, they typically move quickly past the “wow” stage and into critical listening and candid discussion of the different acoustical treatments presented and of the overall sound of the space. This helps architects and project owners to feel connected to the acoustical aspect of the design, and it helps the team to agree on design decisions that may have significant implications regarding cost and aesthetics.

This paper presents several case studies of projects where auralization was an integral part of the design process. Additionally, it describes a rapid auralization design and development process using a MaxMSP-based real-time ambisonic convolution platform.

1 INTRODUCTION

Over the last ten years, consultants at Acentech have worked to develop and refine techniques for using auralization as part of the design process for a variety of buildings and spaces. Much of this work has focused on the auralization of activity sounds in large public spaces such as atrium-type lobby spaces in universities and cultural institutions. The relevant features for auralization of this type of space are:

- overall loudness of activity noise,
- ease of close conversation in the presence of this noise,
- intelligibility of a PA system, and
- transmission of sound among adjoining spaces.

By presenting auralizations of an environment over loudspeakers in an acoustically appropriate meeting room, it is possible for a group of listeners, such as a project design team and owner, to experience the auralization together and to accurately judge many aspects of the auralized space’s acoustical character, such as loudness, reverberance, and speech intelligibility.

To create a perceptually realistic soundfield, it is necessary to include multiple source locations and to convolve each source with time-incoherent anechoic sound material. Our early techniques required completing separate convolutions for each source and then combining sources and adjusting their levels in an audio editor. This time-consuming process was replaced by a MATLAB routine that performed all of the convolution and combination while maintaining relative level calibration throughout. But, this process still required that all audio to be presented in the auralization be pre-rendered, resulting in a more time-consuming revision and calibration process and leaving no possibility of level adjustment among sources during playback.

The most recent innovation to this process uses software created with the MaxMSP development environment to perform all of the required convolution in real time during the presentation. This allows for easy toggling on and off of individual sources, level and timbre manipulation of individual sources, and switching among various architectural design conditions, all during the presentation.

This paper presents case studies of several auralizations that employed these techniques in different ways to demonstrate relevant information about projects and to provide direct experience of the acoustics of project spaces.

2 A REAL-TIME CONVOLUTION PLATFORM FOR AURALIZATION DEVELOPMENT

We conceptualize auralization development as four main processes:

- modeling,
- impulse generation,
- source material selection and convolution, and
- playback calibration.

Typically, our modeling begins in SketchUp and is then exported to CATT Acoustic where material properties are added. The CATT model is then processed with TUCT, CATT’s room analysis subprogram which generates impulse responses.

Because CATT did not provide convolution and summing of different anechoic audio material with multiple sources, we developed a MATLAB script in 2008 which convolved source material with first-order ambisonic (B-Format) impulses responses for each sound source in the model, and then summed and rendered the resulting audio as either a four channel (quadraphonic) or five channel (5.1 surround) WAV file using simple decoding filters generated by CATT Multivolver. This process was effective, but revisions and calibration were time-consuming since any changes involving source material required repeating the entire convolution and summing process. This placed auralization out of reach for most project budgets.

2.1 Auralization with pre-rendered audio: Museum Atrium

The first major step in the early development of our auralization program took place in 2003 through 2005 during design of a major new wing at the Museum of Fine Arts Boston, which opened to the public in 2010. This work was presented previously at the EAA Symposium on Auralization in Espoo, Finland in 2009 [1] and at Internoise 2009 in Ottawa, Canada in 2009 [2]. The auralization presented a familiar architectural acoustics problem and solution: excessive loudness in a large, reverberant public space, addressed with the inclusion of large areas of acoustically absorptive material. The design team and museum trustees heard a predictive auralization of their future space during a fundraising banquet, complete with 500 talkative diners, a live swing band, and an amplified speech by the museum's director. Participants experienced the difficulty of speaking with their table-mates at this busy function in the absence of acoustical treatment and registered relief when the recommended treatment was added in the virtual environment. The auralization included two source locations for diners, a third for the band, and a fourth for public address loudspeakers. Three architectural design conditions were presented: no acoustical treatment, a small amount of acoustical treatment, and the recommended amount of acoustical treatment. All audio was pre-rendered and played back as four-channel wav files. Anechoic sounds were gathered from various sources, including original recordings of conversation and clearing of dishes made in our nearly-anechoic presentation room, speech from museum audio tour recordings, and studio recordings of a swing band.

Figure 1: The Ruth and Carl J. Shapiro Family Courtyard at the Museum of Fine Arts Boston, MA USA.

The Ruth and Carl J. Shapiro Family Courtyard opened in 2010, and the response from the museum and the public has been overwhelmingly positive. The courtyard performs acoustically as designed and supports a wide range of uses. Subjectively, it sounds remarkably similar to the auralization: live but well-controlled even when full of activity.

This auralization, though successful, was time-consuming to produce. Minor adjustments to relative levels required repeating much of the process. Such adjustments are often necessary in this kind of auralization, as different source material is recorded and calibrated at different levels. Final relative level balancing is often an iterative process, checked with the aid of a sound level meter, CATT output data, and various reference sources. When balancing, for example, a live band with activity noise from diners, we assume a role similar to that of a live sound board operator, adjusting the band to a subjectively appropriate level. The difficulty of making such adjustments was a source of frustration to us for several years.

2.2 Towards a more flexible and efficient auralization development platform

Two points drove the redevelopment of our earlier auralization process. First was a desire to reduce the time and cost required to create an auralization presentation, so that we could make wider use of auralizations in our consulting practice. Second was a need to maintain design flexibility for as long as possible in the development process, so as to respond efficiently to client requests for changes.

The first two steps of our auralization process, modelling and impulse generation, are by far the most time intensive. For complicated auralizations with many sources, processing the model can require several CPU-weeks of processing time. While multicore CPUs and smart model design which accommodates parallel processing significantly reduces total processing time, processing a complicated model still frequently requires tying up multiple computers for several days. Because of this, we desired a platform where the model only needs to be processed once, and most needed tweaks or adjustments can happen downstream.

Convolution in MATLAB is a fairly quick process. However, our process of pre-rendering the auralization meant that any change, no matter how small, required a full repetition of the convolution and summing steps – not a trivial matter when making final preparations on the day of a client meeting!

These issues pointed towards a platform that could perform the convolution and ambisonic decoding in real-time, while also providing flexible options for additional processing.

2.3 Real-time multichannel convolution using MaxMSP

We currently produce all of our auralizations using the MaxMSP programming environment published by Cycling ’74 for all steps following the generation of impulse responses in TUCT. These include convolution, ambisonic manipulation, and the final presentation. Max was originally created as a graphical, object-oriented programming platform for musical composition, and has been steadily upgraded to include a robust set of audio (MSP, Max Signal Processing) and video (Jitter) processing tools. Max also makes graphic user interface (GUI) implementation trivial, since many of the graphical function objects are directly useable as UI objects as well. One of Max’s strengths is that it is an open platform that allows independent development of additional function objects (called "externals"); this extensibility is critical when using Max as an auralization tool since many of required functions, such as convolution and ambisonic encoding and decoding, are not part of the base MaxMSP object library.

The most important tools for auralization in Max are the convolution externals in the HISSTools library, written by Alex Harker and Pierre Alexandre Tremblay of the Huddersfield Immersive Sound System (HISS) research group at the University of Huddersfield, UK. The HISSTools Impulse Response Toolbox provides simple yet powerful objects for real-time convolution in Max, as well as tools for measuring and manipulating impulse responses. HISSTools can be freely downloaded from the HISS website [3].

We use two libraries for ambisonic processing in our auralizations, *Ambisonics Externals for MaxMSP* written by Philippe Kocher and Jan Schacher at the Institute for Computer Music and Sound Technology at the Zurich University of the Arts, Zurich, CH and *Ambisonic tools for Max/MSP* written at the Center for New Music & Audio Technologies at the University of California at Berkeley, Berkeley, CA, USA. Both of these libraries are freely available at the websites of the respective organizations [4, 5]. These tools allow decoding the convolved ambisonic signals for playback on arbitrary speaker arrangements and also allow sound sources to be placed and moved within an ambisonic soundfield.

Using MaxMSP as a development platform has dramatically reduced the time required to create a complex auralization and allows us to retain the flexibility to modify the auralization all the way up to, and in some cases during, the final presentation to our clients. This in turn leads to a more responsive and deeper engagement with our clients resulting in better communication between all parties and more successful projects.

3 CASE STUDIES IN AURALIZATION-DRIVEN DESIGN

3.1 Multiple concurrent sources: Concert Hall

The new General Academic Building at the University of Massachusetts Boston, currently in design, will include a 400-seat music recital hall to be used for solo recitals, chamber music, jazz combos, orchestra, vocal chorus, and other musical ensembles. The design includes acoustical variability in the form of curtains at the lower walls. An auralization presented several different types of performance, including solo piano, jazz band, symphony orchestra, and vocal chorus, heard at three audience positions, with various curtain configurations. For the large ensembles, two source locations were included on stage to provide some left-right spatial spread. Stereo anechoic material was used where available, and in other cases monaural material was used. The source material was taken from the Denon compact disc "Anechoic Orchestral Music Recording" (1995) and from Wenger Corporation's "Anechoic Choral Recordings" (2004). The auralization helped the university music faculty to gain confidence in the design, particularly its acoustical variability.

Figure 2: CATT Acoustic model of the University of Massachusetts recital hall, Boston, MA, USA

We have begun experimenting with using anechoic recordings of individual instruments made by researchers at Aalto University [6], and hope to use these to render orchestral music with improved spatial accuracy in future projects.

3.2 Natural source, sound reinforcement, and HVAC system noise: Multifunction Hall

Margery Milne Battin Hall at the Cary Memorial Building in Lexington, Mass. is a multifunction auditorium built in 1928. The hall's users were unsatisfied with its sound reinforcement system, and the building's HVAC system was loud and inefficient. Acentech was asked to design an upgraded PA system and to assist the mechanical engineer in quieting the ventilation system.

To demonstrate the predicted effects of our various design recommendations, and to help the client prioritize their use of limited funds, we created and presented an auralization. This auralization included a natural sound source on the stage, the HVAC background noise, the full existing sound system (12 sources), the full proposed sound system (21 sources), and room responses with and without acoustically absorptive curtains. Several options were included for HVAC noise. The current noise was presented based on the system noise level and spectrum measured in the auditorium, and two proposed upgrades to the HVAC system were presented based on levels and spectra calculated according to ASHRAE guidelines. Since the auralization is being processed in real time, the background noise, PA system, and curtains can be toggled during playback with a continuously running source. This allowed for seamless comparisons between the different options and made the differences much easier for the clients to understand.

Figure 3: Battin Hall, Cary Memorial Building, Lexington, MA USA

Because the convolution takes place with first order ambisonic impulse responses, each source requires four channels of real-time convolution. While we have successfully had over 50 channels of convolution running concurrently on a quad-core CPU, the 140 channels of convolution that would be required to generate the responses of all 35 sources in real time is not practical. To reduce the number of concurrent convolution channels, we grouped all of the loudspeaker sources for the existing PA system and the sources for the proposed PA system, which will be fed the same anechoic source material, into a single TUCT run and then exported a summed impulse response of the entire group. This allowed us to present all 35 sound sources while only requiring 16 concurrent channels of convolution.

3.3 Custom source material and integration of measured impulse responses: Aquarium Exhibit Hall

One of the New England Aquarium's most popular exhibits is the Shark and Ray Touch Tank, where guests can pet small sharks and stingrays. The space has a naturally high background noise level due to the tank's pumps and the electric dryers for guests to use after washing their hands. When combined with the added noise of as many as 100 excited elementary school students, the hall containing the Touch Tank can be extremely loud. The New England Aquarium was interested in ascertaining the noise impact on the animals in the tank, as well as exploring potential room treatments to reduce the overall noise level for the comfort of their staff and guests.

Figure 4: The Touch Tank at the New England Aquarium, Boston MA

For the auralization of this space to be successful, it was critical that appropriate audio source material was used. Since the audience for this auralization was intimately familiar with the modelled environment the sounds presented had to match their expectations in order to be convincing. To achieve this, we recorded many sounds within the space, including the background sound of the unoccupied space, the hand dryers, and the sound of excited children. The speech of one of the educators who directs patrons in safely touching the animals was transcribed and re-recorded anechoically. Since we were also concerned about the noise impact on the animals in the water, we also recorded the background sound underwater using a hydrophone, and measured the transfer function impulse response between a microphone just above the water and a hydrophone in the tank.

The auralization was very successful, but in an unusual way: in this case we demonstrated that additional acoustical treatment would not make a substantial difference in the loudness of the space, and that the most effective option for quieting the exhibit would be to control the number of patrons in the space at one time. We also clearly demonstrated that the airborne noise generated by patrons in the exhibit was almost entirely masked inside the tank by the noise of the pumps that circulate the water. By making a modest investment in the auralization, the Aquarium was able to avoid making a large investment in room treatments that would have required shutting the exhibit down to install and then been ineffective in addressing their noise concerns.

This auralization served to highlight a common thread in all of our auralization work: the importance of appropriate source material in creating perceptual veracity (as opposed to parametric accuracy) in an auralization. By having access to a semi-anechoic space, we are able to record custom material for use in our auralization work to ensure that the character of the source material is appropriate to the expectations of our audience. By tailoring the source material to match what our clients hear in their current spaces, listening to the auralization can be focused on the parametric aspects of the presentation without being sidetracked by clients' being disengaged from the listening experience by the distraction caused by inappropriate audio content.

3.4 Complex soundscape with varied acoustics in a large space: University Atrium

The centrepiece of the new home of the Olin Business School at Washington University in St. Louis, MO is the Forum, an amphitheatre-like lecture and presentation space which extends through several floors of circulation space and is ultimately open above to a five storey glass, wood, and stone atrium featuring a café with seating all around the opening. Despite our presentation of calculated reverberation times and expected background sound and speech intelligibility levels, the client remained unsure that the extensive acoustical treatments we recommended were truly needed. After further discussion, it was decided that an auralization would be the best way for the client to make an informed decision regarding room treatments, and to develop appropriate expectations for the acoustical performance of this complex space. Of particular interest to the client was the level of café activity noise that would be audible in the Forum during a lecture presentation.

This auralization included a speech source and sound reinforcement system in the Forum, and four independent activity sources in the atrium. Many different types of anechoic source material were used for the activity sources, ranging from quiet studying with light footfall noise to boisterous conversation and the sounds of eating and tables being cleared. The level of ambient activity can be varied in real time, and for each activity level chosen by the operator appropriate activity samples are dynamically selected and fed into the convolution engines for the various modelled activity sources. A close-mic'd recording of a speech by an Olin School professor was used as the sound source in the Forum, which was both context-appropriate and helped to anchor the audience's sense of place in the modelled environment. This recording was made by the school for public relations

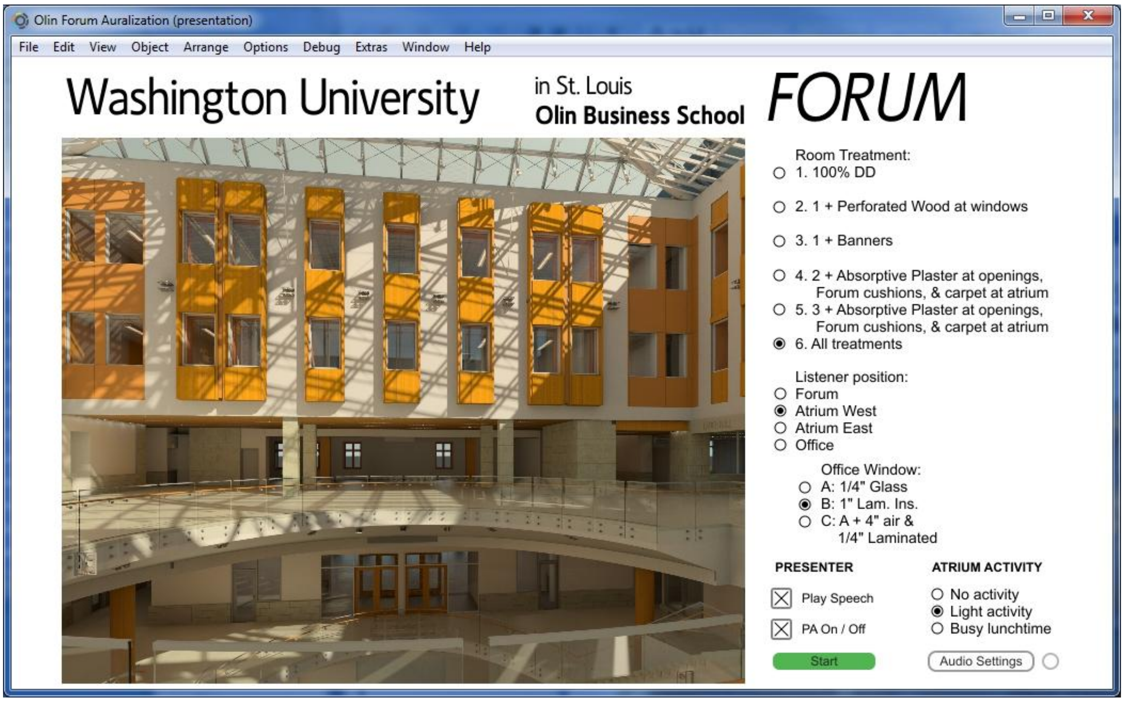

Figure 5: The GUI (Graphical User Interface) of the Olin Business School Forum auralization, constructed in MaxMSP.

purposes and, while not anechoic, was sufficiently dry to be effective for this auralization of a reverberant space. Moreover, the cardioid podium microphone used for the original recording is of the type which speakers will use in the built space, which, while compromising the timbre of the unamplified speech, better presents the full signal path of amplified speech.

Five varying levels of acoustical treatments were modelled, ranging from the initial design to fully treating every available surface. Five listener locations are available, including one inside a faculty office overlooking the atrium which incorporates the transmission loss of the window glazing and allows for three different glazing types to be auditioned. Of course, all of these parameters are freely variable at run time without a need to stop and restart the source recording. See Figure 5 for an illustration of the graphical user interface (GUI) used to select the various options available in this auralization.

In this case, school administrators in attendance were able to make the connection between the architectural design, the numerical descriptions of the room acoustics, and the percept of experiencing the described environment through the auralization in a way that they had not been able to from a written report. Having come into the auralization sceptical about the need for the extent of acoustical treatments that we'd recommended, their final decision was that fully treating the Forum and atrium was a requirement for the success of the project. They were also alerted to practical considerations regarding scheduling of Forum presentations and the need to control Café activity during these presentations.

3.5 Algorithmically generated, massively multichannel outdoor soundscape: 17th Century Churchyard

While not an architectural design problem *per se*, The Virtual Paul's Cross Project [7] represents the current cutting-edge of Acentech's auralization program. This project was a joint investigation led by Dr. John Wall at North Carolina State University which included the English, architecture, and linguistics departments at NCSU, the St. Paul's Cathedral archaeology staff, and acoustical consultants at Acentech. The goal of the project was to recreate the soundscape and visual surroundings of John Donne's 1622 Gunpowder Day sermon at Paul's Cross in the churchyard outside of St. Paul's Cathedral as it was prior to the Great Fire of London in 1666.

The Paul's Cross auralization includes the modelled acoustics of the open churchyard and its surrounding buildings, a 2 hour and 15 minute anechoically-recorded sermon, and an "artificially intelligent" crowd which is variable in real time from between zero to 5000 people and that "listens" to the sermon and then selects appropriate reactions from a custom-produced sample library. Birds fly overhead, horses trot past at the edges of the crowd, dogs bark, and the bells of St. Paul's mark the quarter hours. All of this is observable from twelve different listener locations, and crowd size and listener location are variable during playback without any interruption to the sermon.

The ability to generate the full soundscape in real time was absolutely critical to the success of this project. If we had been required to hand-arrange all of the crowd reactions and ambient sound events, we would have been faced with a task comparable to doing sound design for a feature-length film which would have required an amount of work drastically outside the project budget. By using statistical models to drive the auralization, we were able to specify model parameters and then automatically generate all of the ambient sounds, with that added benefit that no two listens through the sermon are exactly the same. Also, not needing to pre-render the audio substantially reduced the size of the data assets. As an example, the audio assets for the auralization total 2.6 GB of data. If all configurations of the auralization were to be pre-rendered, it would result in 240 GB of data, almost a hundredfold increase. As we have seen in other cases, real-time auralization allows the sermon to run continuously while model parameters are changed, resulting in a notably more enveloping and believable aural experience for the listener.

While triggering the more static environmental sounds of wind, animals, and church bells was fairly straightforward, generating a crowd that could track the dynamics of the sermon was a greater challenge. A rudimentary artificial intelligence (AI) was written in Max that listens to the sermon and then selects a sample from the library of crowd recordings of an appropriately intense response. The full crowd is made up of ninety independent instances of this AI, which are tiled across the soundfield of the churchyard. Each AI's behaviour preferences are randomized at runtime. This allows the impression of a very large crowd of independent listeners to be generated from a small amount of source material.

To fully model each node of the AI and environmental sound within CATT would have required 250 days of processing time and 400 additional channels of real-time convolution. Instead, a simplified approach was taken. The dry output of each AI and environmental sound generator was encoded into an ambisonic signal to give the direct sound the appropriate spatial and level cues in the final presentation. Then, the omnidirectional channels (W in B-Format parlance) of all of the AIs and effects were summed together and convolved with a set of impulses derived from an omnidirectional source in the center of the churchyard with the direct sound removed. Thus, the direct sound is presented in a spatially accurate way via the ambisonic encoding of the dry audience and effects, and the reverberant sound is still perceived as coming from reflective surfaces that are accurately placed around the listener. Since precise spatial localization was desired for the preacher and the church bells, they are auralized with dedicated sources with appropriate directional characteristics in the model. Even with the shortcuts for reducing the channel load of the ambient sound sources, the Paul's Cross auralization requires over 100 concurrent audio playback channels (but only 12 convolution channels, due to the simplifications described above) when the crowd size is set to its maximum.

The Virtual Paul's Cross auralization has facilitated a new level of conversation and inquiry into Donne's preaching and more broadly of the experience of the churchgoing public of 17th century England. We have been able to provide a direct experience of the way an unamplified voice interacts with an outdoor forum such as the historical Paul's Yard and of how speech intelligibility and loudness change for many listener positions and crowd sizes. Being able to switch between listener positions and crowd sizes immediately and in real time allows for much easier and more natural comparisons to be drawn from the auralization than would be possible from tabular data and opens the experience up to a wide range of interested listeners who lack the acoustical knowledge to gain a meaningful understanding of aural events from technical descriptions of them.

Figure 6: A rendering of Paul's Cross, ouside of St. Paul's Cathedral, London, UK

4 CONCLUSIONS

We have found auralization to be uniquely capable of bridging the "communication gap" between acousticians, architects, and project stakeholders in our consulting practice. Allowing parties to come together and listen as a group to the acoustical implications and possibilities of architectural design decisions facilitates understanding in a powerful way that allows for rapid and harmonious decision making. In many cases, auralization is not only the best way to communicate acoustical information, but can result in overall cost savings to a project by reducing the need for revisions during the design process or renovations to correct problems after construction.

In particular, a real-time auralization system allows for richer and more immersive auralizations, without interrupting the flow of the source material and an immediate response of the sound to interactions with the presentation GUI. A real-time system also results in substantial time and cost savings to the client, and allows both initial development as well as revisions as the project progresses to happen on a schedule consistent with the aggressive timelines of many architectural projects.

5 REFERENCES

[1] I. Pieleanu, "Auralizations of Public Atrium Spaces", *Proc. Of the EAA Symposium on Auralization, Espoo, Finland, 15-17 June 2009*, 2009.

[2] J. Sacks, "Auralization for Public Spaces," *Proc. Internoise 2009, Ottawa, Ontario, Canada, August 23-26 2009*, 2009.

[3] HISS: Huddersfield Immersive Sound System, (University of Huddersfield), http://www.thehiss.org/ , accessed 25 Nov. 2013.

[4] Ambisonics Externals for MaxMSP, (Institute for Computer Music and Sound Technology, Zurich University of the Arts), http://www.icst.net/research/downloads/ambisonics-externals-for-maxmsp/ , accessed 15 Jan. 2013.

[5] Ambisonic tools for Max/MSP, (Institute for Computer Music and Audio Technologies, University of California at Berkeley), http://cnmat.berkeley.edu/link/4415 , accessed 15. Jan. 2013.

[6] T. Lokki, J. Pätynen, “Applying Anechoic Recordings in Auralization”, *Proc. Of the EAA Symposium on Auralization, Espoo, Finland, 15-17 June 2009*, 2009.

[7] B. Markham, M. Azevedo, J. Wall, "Acoustical archaeology - Recreating the soundscape of John Donne's 1622 gunpowder plot sermon at Paul's Cross." *The Journal of the Acoustical Society of America* 133.5 (2013): 3581-3581.