Auralizing Acoustic Architecture

A Multi-channel Ambisonic Listening Room for Architectural Design

Brady Peters[1], Benjamin Lappalainen[2], Alexandra Fiori[3]
[1,2,3] University of Toronto
[1] brady.peters@daniels.utoronto.ca [2,3] {benjamin.lappalainen|alexandra.fiori}@mail.utoronto.ca

There is a need for architectural design tools that enable designers to understand how spaces sound. This project aims to develop a method by which architects can gain experiential acoustic feedback on existing and proposed projects through the use of ambisonic spatial auralization. This research proposes that the creation of a fully three-dimensional soundfield can be a tool for architectural acoustic design. The use of ambisonics has largely been limited to virtual reality applications; however, with the growing support in a variety of popular software, opportunities for using spatial audio as a design tool are beginning to make themselves apparent. This paper reports on an experimental setup for a 12-channel, speaker-based auralization system that plays recorded and simulated ambisonic tracks. Novel uses for this setup are proposed.

Keywords: *Architectural Acoustics, Acoustic Simulation, Auralization, Ambisonics*

INTRODUCTION

All architecture has an acoustic quality. Whether architects consider it or not, all spaces have sound and react to the sounds that we make. Unfortunately, the ability to design, to explore, or to communicate sonic experience is not within the ability of most architects. A wide variety of acoustic simulation software exists today, providing architects and engineers with tools that allow them to predict the acoustic properties of an environment. Computational and parametric models can generate 3D models that can be used as input to acoustic simulation software; however, the tables and numerical values given by simulations offer little in the way of a perceptual or experiential feedback on the acoustic qualities of space. Much current research in acoustic architecture looks at how geometry and materiality can be specified to create specific acoustic conditions, which is of course very important, as it is these qualities that determine the acoustic qualities of architectural spaces; however, to enable sound to be a driver for architecture, designers must be able to gain a perceptual understanding of the spaces they are proposing.

Computed acoustic simulation can create "auralizations", the acoustic equivalent of "visualizations." Much progress has been made in the generation of auralizations from these simulations, allowing a user to listen directly to their virtual acoustic environment. Today, state-of-the-art algorithms for room acoustic modelling are mostly able to provide plausible, but not entirely authentic auralizations of real-world environments when strictly operating from a 3D model of the space as an input (Brinkmann et al. 2019). The perceptual validity is typically assessed using methods similar to that described in the Spatial Audio Quality Inventory (SAQI), which is a common vocabulary used to describe the perceptual qualities of auralizations (Lindau 2015). While there are certainly benefits that come with improving realism, it's important to consider the practical use for auralizations in architectural design. Interviews conducted with users of auralization brought up the point that due to the subjective nature of sound, clients listening to simulated auralizations generally don't perceive subtle differences that may invalidate an auralization according to SAQI criteria (Thery et al. 2019). For auralizations to be realistic the spatial orientation of the sound sources is critical, and while spatializing sound using headphones is possible, this approach has limitations on the understanding of the spatiality of the soundfield. Ambisonics mimic the way humans localize sounds in front of, behind, above, and below us, and provides the ability to reproduce or create a fully 3D soundfield that is an immersive representation of an acoustic environment, existing or yet to be built. To examine perceptual validity, most acoustic researchers rely on binaural audio delivered through headphones rather than ambisonic speaker setups (Thery et al. 2019). Our research seeks to evaluate changes in room design variables that are perceptually audible on an ambisonic speaker setup.

The primary goal of this research is to develop a method and workflow for the auralization of ambisonic B-format sound files that can be listened to on a 12-speaker ambisonic sound system. Using as much free and open-source software as possible to aid in the accessibility of these techniques, this workflow aims to be well-defined such that it can easily be reproduced by others. Once this system is set up, the second goal is to identify and describe the methods by which architectural designers can use this system productively to create better sonic environments. We develop three design strategies to create and experience 3D architectural sound using our ambisonic system. The first strategy is to record existing urban and architectural soundscapes using a specialized ambisonic recorder. The second is to use acoustic simulation software to generate ambisonic impulse responses and auralizations of virtual environments. To do this, we use Rhinoceros 3D to create a 3D model of an enclosed environment and the acoustic simulation software Odeon to create auralization files from the 3D model. Our auralization workflow will be capable of taking in either a B-format impulse response from the simulation software and convolve this in the audio software with a chosen anechoic sound (thus simulating the effect of the sound being played at the designated source location in the virtual space), or simply play the B-format auralization directly. We will compare the results of these processes for perceptible differences. The third strategy is to use audio software to create new architectural acoustic soundscapes and use the ambisonic system to test these spatial audio propositions.

BACKGROUND: AURALIZATION

Auralization is "the technique of creating audible sound files from numerical (simulated, measured, or synthesized) data" (Vorlander 2008). There are a variety of ways of collecting the data required to produce a reasonably accurate auralization of an environment. The main component that is required to produce an auralization is an impulse response. The impulse response is the output when a dynamic system is effected by a brief input signal. In acoustics the impulse response is like an acoustic fingerprint of a space; it contains all of the information of reflections and energy absorption across all frequencies. An impulse response is valid for a single source-receiver location pair. Computational simulation methods allow for more flexibility simply due to the ease of calculating impulse responses for a variety of source-receiver pair locations. Once an impulse response is produced, it can be convolved with a dry sound, which is any sound recorded in an anechoic chamber. Anechoic chambers provide an entirely nonreflective, non-echoing environment in which purely direct sound can be recorded without any room effects. The convolution of this sound with the impulse response produces a simulation of what the listener at the receiver position would hear.

After this data is generated, the sound is played through a listening system. Typical this is headphones and loudspeaker arrangements. Binaural representations are typically generated for headphones, which use the auralization data in combination with a Head-Related Transfer Function (HRTF) to accurately account for the additional audible effects of a listener's head, torso, and ear structure on the sound (Vorlander 2008). Generalized HRTF datasets are used in applications such as video games to provide binaural sound to users without requiring personalized measurements, although the perceived accuracy of these auralizations depends on how closely the characteristics of the listener's ears match the HRTF in use. HRTFs may also be used with stereo loudspeaker arrangements in order to achieve the perception of surround sound. Loudspeaker arrangements using a greater number of speakers positioned around the listener don't require the HRTF processing step. Most research around simulated auralization has been focused on comparing and increasing the analytic and perceptual accuracy of acoustic simulation techniques (Postma and Katz 2016). While continued research in this field has significant merit, many existing techniques are already able to produce perceptually valid auralizations. Auralizations for the purpose of acoustical design are typically used in a demonstration context for showing the audible results of the design to a client (Thery et al. 2019).

BACKGROUND: AMBISONICS

Ambisonics was developed specifically for the film and music industry, however, it has rarely been used as a tool in architecture. The evolution from stereo format to ambisonics can date back to the Paris Exhibit of Electricity in 1885, when Clement Ader used multiple sets of telephone transmitters and receivers to play sounds from other areas of the expo; and, in the 1920s, a binaural sound system was developed by Harvey Fletcher and his team at Bell Telephone Laboratories, better known today as headphones (Malham and Myatt 1995). Little progress was made until Disney's famous Fantasia movie sparked a new revolution for sound in cinema as the sound designers, with the help of Leopold Stokowski, developed a nine-track recorder that used nine separate synchronized optical recordings (Malham and Myatt 1995). Cinema and sound design for films began a chain-reaction of events which lead to the rapid development of sound systems and how audiences experienced movies through the aural architecture/sound designs of the film. Although the idea has been around for multiple decades, it has never been commercialized until recently due to advances in technology, relating to virtual reality, 360-degree videos, game designs, and cinematic experiences in theaters. Ambisonics is a method of recordings, mixing and playing back sounds in a 360-degree soundfield. British engineer Michael Gerson developed the technique during the 1970s (Gerzon, 1980), see Figure 1. It recreates the three-dimensional soundfield by allowing the listener to hear sounds in front, behind, below and above them.

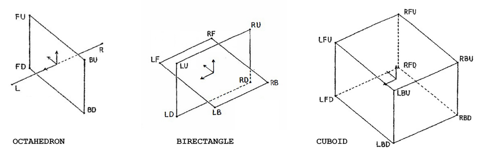

Figure 1
Gerzon's examples of loudspeaker arrangements with the listener at the center: [Left] Cuboid; [Center] Octahedron; [Right] Birectangular; (LF) left-front speaker, (RF) right-front speaker, (LB) left-back speaker, and (RB) is the right-back speaker. (from Gerzon 1980)

Ambisonics is a surround sound format that allows for positioning of sound sources both horizontally and vertically relative to the listener. Ambisonic listening setups utilize loudspeakers typically positioned on the surface of an imaginary sphere and facing the center, with the listener's head positioned at the center (Lee 2007). The primary distinction between ambisonics and other multichannel formats is that channels in an ambisonic file do not correspond directly to speaker signals. They contain a representation of the sound field using components of the pressure gradient in a set amount of directions, depending on the order to which the sound field is to be reproduced. The encoded file, which is referred to as a B-format file, must be decoded before it can be played on a speaker system. The decoding step utilizes a linear combination of the ambisonic channels to generate the signal to be played on each speaker.

In theory, an infinite number of spherical harmonics are required in order to perfectly recreate a given sound field. In practice, a finite length of the series is used based on available hardware and software resources. For a chosen finite order n, (n+1)2 channels are required to represent the sound field. First-order ambisonics are therefore represented by 4 channels, second by 9, etc. To get the full effect when listening to a given ambisonic order, the number of speakers in the setup must at least equal the number of ambisonic channels. As such, practicality and cost may limit the use of higher order ambisonics in loudspeaker setups at present. Additionally, the area in the center of a speaker-based ambisonic system in which the sound field can be recreated accurately (the "sweet spot") increases in size as the order of the ambisonic playback increases (Frank 2014). An important distinction when discussing the practical use of B-format files are the differences between existing standards regarding channel ordering and normalization. The two most common conventions are Furse-Malham (FuMa) and ambisonic Exchangeable (ambiX). For first-order ambisonics, FuMa uses a channel order of WXYZ, whereas ambiX uses WYZX ordering (Nachbar et al. 2011). The omnidirectional W channel is also 3dB stronger in ambiX than in FuMa. The ambiX channel ordering is also referred to as ambisonics Channel Numbering (ACN), where channels are simply referred to numerically (W corresponds to 0, Y to 1, Z to 2, etc.) (Blue Ripple 2020). While these differences are subtle, they are incredibly important when decoding the file. ambiX seems to have the widest use today, although most software for the processing of ambisonics supports both conventions.

The primary benefit of ambisonics is that the required decoding step means a single source file can be decoded to a wide variety of listening layouts without loss of information. A sound designer who wishes to perform vertical placement of sounds doesn't need to worry about loss of information if the listener only has a stereo speaker setup - the sound field will be recreated as accurately as possible with the given listening setup. With virtual reality applications entering the mainstream market, ambisonics have managed to gain traction through binaural renderings for headphones (a known and common listening configuration). Low-cost ambisonic microphones and greater support for B-format recordings have also contributed to the recent success of ambisonics. Binaural renderings utilize a virtual set of loudspeakers positioned around the listener, which somewhat circumvents the practical limitations of higher order Ambisonics. The limiting factor for Ambisonic order then becomes the number of audio channels supported by a given playback system, and this limit is rarely reached due to the low perceptual discrepancy between orders past a moderate threshold. Much of the research on acoustical simulation utilizing ambisonics contains rather sparse details on the practical setup and implementation of the listening system, instead focusing on more theoretical aspects. What follows is a description detailing the hardware and software implementation of a speaker-based auralization system.

THE AURALIZATION LISTENING ROOM

A 9 ft. x 13 ft. room for ambisonic auralization has been set up, utilizing twelve Yamaha HS5 speakers configured as shown in Figure 2. These speakers are controlled using a MOTU 16A audio interface, which performs the digital-to-analog conversion of the 12 audio channels sent from a laptop connected via USB. Yamaha HS5 speakers were chosen for their exceptional price-performance ratio and relatively small size. The MOTU 16A interface has 16 analog outputs and was chosen based on the minimum requirements of our twelve-speaker listening configuration. It also allows for future expansion of the layout up to 16 speakers, which would allow for an accurate representation of up to third-order ambisonics. The current speaker layout can accurately represent up to second-order ambisonics. The room is outfitted with acoustic absorbing panels on the walls around the listening area in order to reduce the effect of room acoustics as much as possible. While the panels help to "deaden" the sound in the room, sound dampening could certainly be improved, particularly in the lower frequencies which was easily noticeable when speaking in the room. As this is a small, rectangular space, room modes are perceptible. Speakers were hooked up to the analog outputs of the receiver with channel numbering starting from the center-front-up speaker, with numbering proceeding clockwise around each ring from the top to the bottom, see Figure 3.

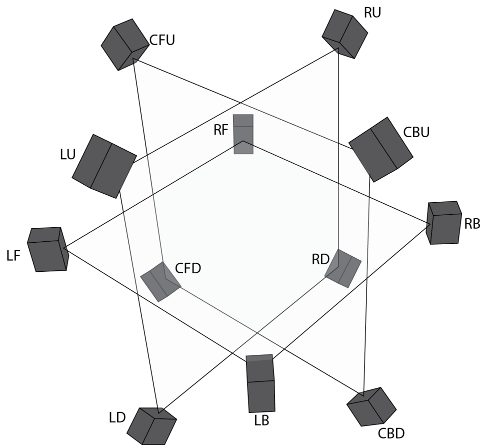

Figure 3
The 12-loudspeaker arrangement: (LF): left front, (RF): right front, (CFD): center front down, (CFU): center front up, (LU): left up, (RU): right up, (LD): left down, (RD): right down, (LB): left back, (RB): right back, (CBD): center back down, (CBU): center back up

Figure 2
The Ambisonic Auralization Lab

AMBISONIC ENCODING AND DECODING WORKFLOW

This section outlines the process by which ambisonic audio from a variety of sources can be listened to on the 12-speaker system described above. The entire encoding and decoding process is performed using the Reaper Digital Audio Workstation (DAW). Reaper was selected due to its free evaluation period, low cost for educational and personal use (60 USD at present), multi-platform support, and powerful architecture for routing audio through virtual channels on a per-track basis. Like most other DAWs, it supports all versions of the VST standard, allowing the use of an enormous selection of free and commercial plugins. All plugins described below are part of the ambiX v0.2.10 ambisonic plugin suite and mcfx v0.5.9 multichannel audio plugin suite, which are both free to download and use (Kronlachner 2020).

The audio interface is first selected as the output device to expose the available output channels to Reaper, see figure 4. A track is added to the project, and the software channels 1-12 in the track are assigned to hardware output channels 1-12. All plugins from the AmbiX and mcfx plugin suites were installed - note that ambiX provides pre-compiled downloads for order 1, 3, and 5 versions of the plugins. If other ambisonic orders are desired, those will need to be compiled. Details on the compilation process are outlined in the ambiX Github repository. The base track with configured inputs and outputs was then copied multiple times and modified, with each track providing a method of listening to or generating ambisonic audio from different types of source files.

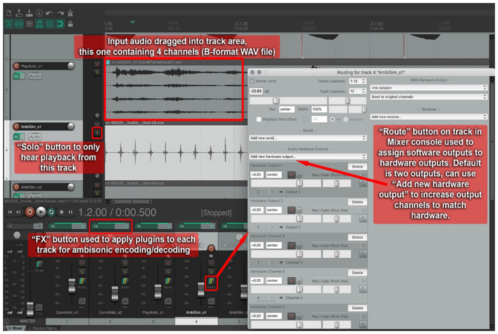

Figure 4
Annotated Reaper interface

The first track allows for any mono anechoic sound to be convolved with an ambisonic impulse response, see figure 5. This is performed utilizing two plugins - the first is a multi-channel convolution reverb "mcfx_convolver" which takes an impulse response and anechoic recording and convolves the two. The output of this plugin is ambisonic in B-format and is used as the input for the second plugin "ambix_-decoder" which decodes the file to channels necessary for the speaker layout based on a decoder matrix. To generate a decoder matrix, you need the exact azimuth and elevation of every speaker in the layout. Ideally, all speakers should be the same radial distance from the listening point, although delay can be added to specific channels to compensate if speaker distances vary. MATLAB's higher order ambisonic (HOA) demo functions was used to generate the decoder matrices for 1st, 2nd, and 3rd order ambisonics. The second track, see figure 6, allows for a recording in B-format to be played back. The setup is the same as the first track, except the convolution plugin is disabled.

The third track allows for simulated positioning of an anechoic sound around the listener, which is helpful for testing and validation of the setup and workflow, see figure 7. A positional encoding plugin "ambix_encoder" is used to encode the mono source to B-format and has settings to adjust azimuth and elevation of the sound as well as the motion rate if you wish to automate the movement of the sound around the listener. The output of this plugin is sent to the decoding plugin to be played on the speaker layout. This track has been essential in the validation of the hardware and software elements of the listening system. The sound was positioned at all combinations of azimuths and elevations in 45-degree increments and played back. Channel volumes were monitored at each of these settings to ensure that the sound was being played through the correct speakers, and the setup was listened to in order to confirm the perceptibility of the sound positioning.

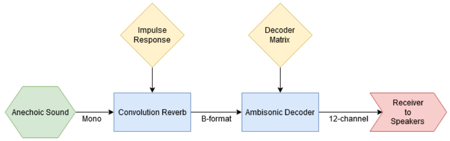

Figure 5
Track 1 flow diagram.

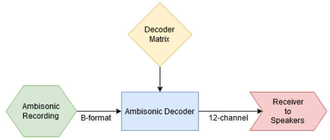

Figure 6
Track 2 flow diagram.

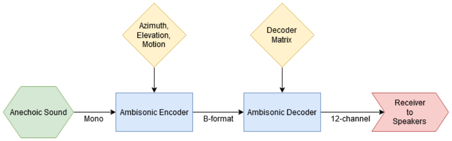

Figure 7
Track 3 flow diagram.

AMBISONIC RECORDING

The second track described in the previous section requires a B-format file. One method of obtaining such a file is to use an Ambisonic microphone such as the Zoom H3-VR, which we used in our testing. This microphone can capture 1st order recordings directly to B-format WAV files (Zoom 2020). Great care should be taken when obtaining recordings of spaces to ensure that the microphone is sufficiently isolated from vibrations that could adversely affect the captured audio. The easiest way to do so is to mount the microphone on a sturdy tripod. This microphone also has the benefits of being very small, light, and portable, allowing for it to be used in a variety of environments.

AMBISONIC SIMULATION

Auralizations can be produced in acoustic simulation software using a 3D digital model of a space and selected source and receiver positions. We used Odeon for this research. Odeon exports 1st and 2nd order ambisonic auralizations and impulse responses. If recordings and simulated auralizations of a space are obtained with identical positions for the source-receiver pair, the audio can be compared both analytically and perceptibly to examine differences between the auralizations. The comparison of simulated auralizations with real recordings is already the focus of many research projects, and is typically done with the intention of comparing acoustic simulation software or developing novel acoustic simulation methods that more accurately resemble real environments. This is extremely valuable research, but is only beneficial to designers by informing them on which acoustic simulation software is the best to use for simulated auralizations. This research aims to create a listening environment in which the spatial acoustic qualities of different architectural designs can be listened to, studied, and reflected upon.

PERCEPTUAL OBSERVATIONS

Preliminary perceptual testing of the auralization system was performed, primarily using the simulated positioning of a mono sound (Track 3 as described previously). Perceptual determination of horizontal positioning seemed to work quite well, even when using 1st order Ambisonics. As expected, 2nd and 3rd order Ambisonics slightly increased the perceptual positional resolution of the sound. This could also be seen objectively by looking at the volume levels in the various speaker channels - higher Ambisonic orders showed less “spill” of volume into neighbouring channels. However, vertical positional determination was quite difficult. Given that humans naturally have better horizontal positioning in comparison to vertical positioning, it’s hard to determine whether this was the fault of the auralization system or a failure of human perception in the absence of additional sensory cues, such as visuals.

DISCUSSION AND CONCLUSIONS

The speaker-based auralization system outlined in this paper has several notable benefits. Users can turn their head as they normally would to better assess sounds coming from the system, whereas a headphone auralization system would require head-tracking to perform in a similar manner. Additionally, no HRTF is required when using a speaker-based system; since the user is already listening to the sound as if they were sitting in the environment that it was recorded. Our setup also has the benefit of utilizing a single piece of software for encoding, decoding, and playback, providing great ease-of-use. There is minimal to no adjustment required once the system is set up - new source files can simply be dropped into the existing tracks and immediately played back. The biggest requirement is the up-front setup, when the decoder matrix must be generated and all of the plugins must be installed. Working within a DAW geared towards music production allows for many great features that can be used for comparison and testing purposes. More tracks can be set up similar to those described in the previous section to perform side-by-side tests with different ambisonic orders by “soloing” tracks back-to-back, allowing users to flip between different settings with virtually no delay to obtain the most accurate perceptual comparisons between auralizations.

The primary goal of this project was the development of a method for generating and listening to spatial audio. In this paper we have communicated a method for the auralization of ambisonic audio using an accessible software and hardware system. This 12-speaker system uses the Reaper DAW for the majority of the software workflow and MATLAB to generate the decoder matrix for our speaker configuration. The goal of this system is to facilitate the design of architecture that is concerned with the spatial qualities of sound. As all architecture is spatial and contains sound, this could potentially be of benefit to a great number of projects. The testing of this system for specific projects is beyond the scope of this paper, which is focused on the specification of the design tool.

While much future work can be done using this system in its application on new spatial audio projects, there are also several areas of research that should be carried out with regards to the system itself. Confirming the perceptual validity of the system is important, and this could be done by comparing recordings with one's own experience of an existing environment. This would involve using simulations of a modelled existing environment to compare the validity of the simulations to real recordings. An upgrade to the auralization system would be the addition of a VR headset paired with a game engine to allow for a user to have both visual and audio cues while experiencing auralizations, and building on this, it would be particularly interesting to examine the perceptual resolution of sounds when comparing the auralizations with and without VR. A perceptual comparison of headphones with the speaker-based system when using VR would be valuable given that virtually all current VR systems rely on headphones. After this system has been used on a few architectural projects, this knowledge could be used to develop a set of guidelines for the design of architecture with spatial acoustic qualities. A benefit would be to give designers a knowledge base of how the adjustment of various qualities of a space will affect the acoustic properties of that space, and what it might mean to have acoustic qualities that are directional, or multi-directional, and how that could improve communication, experience, or well-being.

REFERENCES

Brinkmann, F, Aspock, L, Ackermann, D, Vorlander, M and Weinzierl, S 2019, 'A round robin on room acoustical simulation and auralization', *Journal of Acoustical Society of America*, 145, pp. 2746-2760

Cushner, J, Myrbeck, J and Young, R 2012, 'Using ambisonic technology in entertainment and design', *Protocol: The Journal of the Entertainment Technology Industry*, 17(4), pp. 16-20

Frank, M 2014 'How to make Ambisonics sound good', *Forum Acusticum Proceedings*, Krakow

Gerzon, M 1980 'Practical Periphony: The Reproduction of Full-Sphere Sound', *Audio Engineering Society 65th Convention*, London

Lindau, A 2015, *Spatial Audio Quality Inventory (SAQI). Test Manual. v1.2*, Technische Universität Berlin, Berlin

Malham, D and Myatt, A 1995, '3-D Sound Spatialization using Ambisonic Techniques', *Computer Music Journal*, 19, pp. 58-70

Nachbar, C, Zotter, F, Deleflie, E and Sontacchi, A 2011 'ambiX - A Suggested Ambisonics Format', *Ambisonics Symposium*, Lexington

Postma, B and Katz, B 2016, 'Perceptive and objective evaluation of calibrated room acoustic simulation auralizations', *Journal of Acoustical Society of America*, 140, pp. 4326-4329

Thery, D, Boccara, V and Katz, B 2019, 'Auralization uses in acoustical design: A survey study of acoustical consultants', *Journal of Acoustical Society of America*, 145, pp. 3346-3458

Vorlander, M 2008, *Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Springer-Verlag, Berlin

[1] http://www.ambisonia.com/Members/ricardo/ASD.htm/

[2] https://www.blueripplesound.com/b-format

[3] http://www.matthiaskronlachner.com/?p=1910

[4] https://www.zoom-na.com/products/field-video-recording/field-recording/zoom-h3-vr-handy-recorder