Auralization of Room Acoustics using Binaural Impulse Response

Joshua Daniel M C

Department of Music, National University of Ireland, Maynooth, Ireland

ABSTRACT: This project aims to develop a software for auralization of the Octophonic studio room at Maynooth University, using binaural impulse response. The primary objective of the project is to accurately capture the binaural impulse response of the room using a dummy head microphone and to develop a software using Csound and Cabbage Audio that allows users to experience the room acoustics through headphones in a virtual environment. The software is designed to apply the captured binaural impulse response to any audio source, allowing users to simulate the experience of being in the Octophonic studio. This paper presents a detailed description of the methodology used to capture the binaural impulse response of the room and the software development process. The paper also discusses the limitations and challenges associated with auralization using binaural impulse response and provides possible solutions. Overall, the project contributes to the development of auralization techniques and provides a useful tool for musicians, audio engineers, and researchers interested in room acoustics and multi-channel compositions.

KEYWORDS: Room acoustics, binaural Impulse response, Csound, Cabbage audio.

Contents

1. Introduction

In the field of audio engineering and music production, it is often necessary to be present in a specific acoustic environment. However, it may not always be possible or convenient to have access to the physical space, especially in the case of a multi-channel studio with several speakers. This is where binaural methods come in, which allow us to record and experience sound as if it were being heard from human ears. Binaural impulse responses are used to emulate a space, and convolution is used to achieve a faithful reproduction of the acoustics.

The aim of this project is to develop a software using binaural impulse responses to auralize the Octophonic studio room at Maynooth University. The software will allow users to experience the room acoustics through headphones in a virtual environment. A specialized microphone, shaped like a human head and ear, will be used to capture the binaural impulse responses of the room. The software will use convolution to apply the captured impulse responses to any audio source, allowing users to simulate the experience of being in the Octophonic studio.

Sound localization is an important aspect of auralization, and binaural methods are particularly effective in emulating this. Interaural level differences (ILD) and interaural time differences (ITD) are key factors in sound localization, and binaural methods are capable of reproducing these effects. In this paper, we will present the methodology used to capture the binaural impulse responses of

the room and the software development process. We will also discuss the importance of sound localization in auralization and the limitations and challenges associated with using binaural methods.

2. Tools and setup

2.1. Hardware and Software Tools

Even though binaural IR is used for this project, a set of 1st order ambisonic IR were also captured for future development and analysis. For this project, a variety of hardware and software tools were used. The hardware used for capturing the ambisonic and binaural impulse responses included the Rode NT-SF1 ambisonic microphone, the Zoom F6 recorder, and a dummy head microphone which was built by Dr.Iain McCurdy, Asst. Proffersor in Music Department at Maynooth University.

The Rode NT-SF1 ambisonic microphone was used to capture the ambisonic impulse responses of the room. It is a high-quality microphone designed to capture sound in a 360-degree spherical field. The microphone captures four cardioid capsules in a tetrahedral array, which allows it to capture both the direction and intensity of sound.

The Zoom F6 recorder was used to capture the ambisonic and binaural impulse responses. It is a portable and compact recorder that is capable of recording up to six channels simultaneously at 24-bit/192kHz resolution. The recorder has four XLR inputs with phantom power and two additional inputs that can be used for line-level signals.

The dummy head microphone was used to capture the binaural impulse response of the room. The dummy head microphone is designed to resemble the shape and size of an average human head and ears, allowing it to capture sound in a way that closely resembles how humans perceive sound.

In addition to the hardware tools, several software tools were used for processing and analyzing the captured impulse responses. The Reaper DAW and Logic Pro X were used for recording and editing the audio files. The Fb360 spatial audio tools and Soundfield by Rode ambisonic plugin were used for converting the ambisonic recordings to various speaker formats.

Figure 1: *Photo of dummy head microphone and Rode NT-SF1 ambisonic microphone used in the setup*

The IEM plugin suite was used for binaural rendering of the captured impulse responses. The suite includes several plugins that can be used to simulate different binaural playback scenarios, such as headphone playback, loudspeaker playback, and room simulation. The Sparta Plugins were used for analyzing the frequency and impulse responses of the captured audio files.

Overall, these hardware and software tools were instrumental in capturing, processing, and analyzing the ambisonic and binaural impulse responses of the Octophonic studio room at Maynooth University, and developing a software for auralization of the room acoustics.

2.2. The Octophonic setup

An octophonic room is a specialized listening environment that features eight speakers arranged in a circular configuration around the listener, creating an immersive audio experience. The listener is situated at the center of the room and can hear sounds coming from any direction, allowing for the spatial placement of audio sources. This type of setup is often used in music production and sound design to create a more immersive and realistic sound field for the listener.[1]

3. Binaural Impulse response or HRIR

Binaural impulse response (BIR), also known as head-related impulse response (HRIR), is a technique used to simulate how sound behaves when it reaches the human

Figure 2: *The octophonic studio setup in the music department at Maynooth University*

ear. BIRs are captured using a specialized microphone called a dummy head, which has two microphones placed inside ear-shaped cavities. The dummy head is positioned at the same location and orientation as a human head, and a sound source is played in a controlled acoustic environment. The two microphones capture the sound as it reaches each ear, creating a set of two impulse responses that represent the sound's arrival time, amplitude, and phase differences between the two ears.[2]

These BIRs can be used to emulate a particular listening environment or to create a virtual audio experience that mimics real-life listening scenarios. They have been used in a range of applications, from audio engineering and music production to virtual reality and teleconferencing.

One of the primary benefits of using BIRs is that they allow for a more accurate and realistic simulation of how sound is perceived by human listeners. BIRs capture the unique characteristics of an individual's ears and head, which affect how sound is localized and perceived in three-dimensional space. This spatial information is critical in creating an immersive audio experience, and BIRs provide a more accurate and precise method for capturing this information.[3]

There are several software tools and libraries available for capturing and processing BIRs, such as Csound and Sparta plugins. In addition, several commercial and open-source platforms have incorporated BIRs for spatial audio processing and binaural synthesis, including IEM Plugin Suite and Fb360 Spatial Audio Tools.

Overall, BIRs are a powerful tool for creating immersive audio experiences and simulating complex listening environments. They provide a more accurate and realistic simulation of how humans perceive sound and have numerous applications in audio engineering, music production, virtual reality, and teleconferencing.

4. Convolution and Deconvolution

4.1. Convolution

Convolution is a mathematical operation commonly used in digital signal processing to modify and analyze signals. It is a method of combining two functions, typically a signal and a filter, to produce a new function that describes how the input signal changes in response to the filter. The convolution operation is defined mathematically as follows:

$$f * g(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$$

where f and g are two functions being convolved, and $*$ denotes the convolution operator. In the context of audio processing, convolution can be used to apply a filter to a sound signal, such as a reverb effect, by convolving the signal with an impulse response of the desired filter.

4.2. Deconvolution

Deconvolution, on the other hand, is the inverse operation of convolution. It is used to extract the original input signal from a convolved signal, by dividing the convolved signal by the filter's impulse response. The deconvolution operation is defined mathematically as follows:

$$f(t) = (g * h)(t) \Leftrightarrow G(f)H(f)$$

where f is the original input signal, g is the convolved signal, h is the impulse response of the filter used in the convolution, and $G(f)$ and $H(f)$ are the Fourier transforms of $g(t)$ and $h(t)$, respectively.

In audio processing, deconvolution can be used for tasks such as removing reverb from a recorded sound, or separating the effects of multiple filters from a single signal. However, it is important to note that deconvolution can be a challenging task, as it can amplify noise and

other artifacts present in the original signal. Therefore, proper techniques and methods, such as regularization and truncation, should be used to minimize the impact of these artifacts. The use of these operations in binaural impulse response processing allows for the creation of a virtual acoustic space that can be experienced through headphones.[4][5]

5. Sound Localization

Sound localization is the process of identifying the location of a sound source in space. Humans and other animals use different cues to determine the direction of a sound source, including interaural time differences, interaural level differences, spectral cues, and head-related transfer functions.

Interaural time differences (ITDs) are differences in the time of arrival of a sound wave at each ear. These differences are most pronounced for low-frequency sounds and can be used to determine the direction of a sound source in the horizontal plane. The brain processes these differences and uses them to calculate the azimuth angle of the sound source relative to the listener.[6][7]

Interaural level differences (ILDs) are differences in the sound level at each ear. These differences are most pronounced for high-frequency sounds and can also be used to determine the direction of a sound source in the horizontal plane. The brain processes these differences and uses them to calculate the azimuth angle of the sound source relative to the listener.[6][7]

Spectral cues are differences in the spectral content of a sound wave at each ear. These differences can be used to determine the direction of a sound source in the vertical plane. For example, a sound source above the listener will produce a different spectral content at the ears than a sound source below the listener.[7]

Head-related transfer functions (HRTFs) are filters that describe how a sound wave is modified as it travels through the listener's head and ears. HRTFs are unique to each listener and can be used to determine the direction of a sound source in both the horizontal and vertical planes. HRTFs can also provide cues about the distance and size of a sound source.[8]

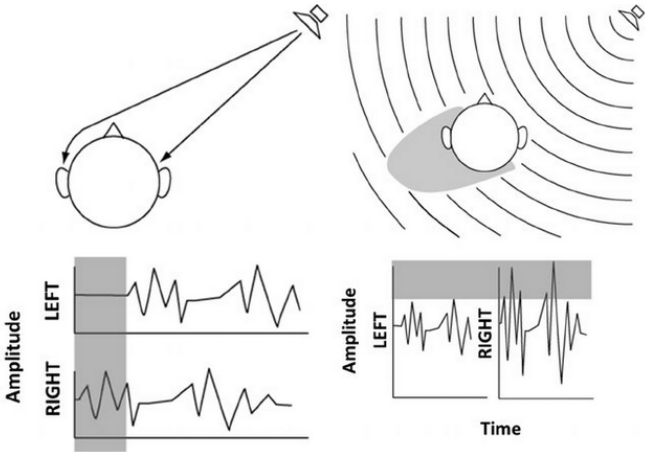

Figure 3: *(Ref: Liang Sun, Xuan Zhong,William A. Yost Dynamic binaural sound source localization with interaural time difference cues: Artificial listeners - April 2015,The Journal of the Acoustical Society of America 137(4):2226-2226)*

6. Csound and Cabbage

Csound is a powerful and versatile software for sound synthesis and processing, originally developed by Barry Vercoe in the mid-1980s at the Massachusetts Institute of Technology. It uses a textual language for defining sound structures and processing algorithms, which allows for a high degree of flexibility and control over the resulting sound output. Csound provides a wide range of built-in opcodes, which are the fundamental building blocks for generating and manipulating sound. These opcodes can be used to create complex synthesis techniques, including additive, subtractive, FM, granular, and physical modeling synthesis, as well as a variety of effects such as delay, reverb, distortion, and filtering. [9]

Cabbage is an open-source graphical user interface (GUI) front-end for Csound, developed by Rory Walsh. Cabbage provides a modern and user-friendly interface for designing and controlling Csound instruments and effects. The interface includes various widgets and controls, such as sliders, buttons, and menus, which can be used to adjust and modulate the parameters of the Csound opcodes. Cabbage also includes a number of additional opcodes and utilities, which are not available in the standard Csound distribution, such as a sampler and a granular synthesizer.[10]

One of the key features of Cabbage is its ability to export Csound instruments as standalone applications or

plugins for use in other software. This makes it possible to create custom instruments and effects for a wide range of applications, including music production, live performance, and interactive installations. Cabbage also supports integration with other software tools, such as Max/MSP, SuperCollider, and Pure Data, allowing for complex and flexible hybrid systems.

Csound and Cabbage were used to develop the software for this project as they offer a powerful and flexible platform for processing, with a wide range of capabilities and applications. This project exclusively uses the pconvolve opcode from csound for the convolution process.

7. Implementation

In order to auralize the room acoustics for headphones, it is essential to take into account the perception of sound by humans, especially while emulating a multi-channel sound source. To achieve this, a specialized microphone is used to capture the impulse response. The dummy head microphone, modeled in the shape of a human head with two microphones placed at each ear, replicates how a human perceives sound. However, it is impossible to create one model for all humans as each person has a different head size, shape, and size of ear, and torso. The dummy head microphone has an average-sized head, with the average size of human head circumference being 56 cm, 55 cm for females, and 57 cm for males. The average distance between the left ear and the right ear is 18 cm.

To capture Binaural Impulse Response, a dummy head microphone is placed at the listener's position or sweet-spot. A test tone is used to measure the impulse response of the room, as it produces the maximum possible frequency for human ears over a period of time. A sine sweep of 20Hz to 20kHz is played over a time period of 7 seconds at around -12 dBFS from Room EQ Wizard (REW), a room acoustic measuring software, and recorded using Zoom F6, a 6-channel field recorder. The recordings were made in 48KHz sample rate to fulfill Nyquist theorem or sampling theorem. A 32-bit float bit depth was used to record, in order to avoid any unwanted distortions or clippings while recording and also for flexibility in the post-production process.

The process is repeated, and sine sweeps are played in each of the individual speakers, in this case, 8 discrete speakers and one subwoofer for low-frequency elements. These audio files contain all binaural and spectral cues that result from the interaction of the human head with the impending sound of the source. These are referred to as head-related impulse responses (HRIRs) in the time domain or head-related transfer functions (HRTF) in the frequency domain. A head-related impulse response (HRIR) measures the magnitude (amplitude) and phase (time delay) distortions caused by different parts, like the head, size and shape of the torso, pinnae, and ear canal when sound arrives from a particular direction.

The recorded audio is then imported into Reaper, a digital audio workstation (DAW), for processing. Any DAW can be used for editing purposes. The recorded HRIRs are edited, and using a convolution plugin, ReaVerb, the distinct audio files are convolved and deconvolved. These HRIRs can be used to filter audio signals, resulting in a binaural signal that is equivalent to the audio signal being received from the direction from which the HRIR was recorded. This way, the room is emulated in headphones as close as how it would sound in the actual room.

This method of capturing and emulating room acoustics for headphones is extensively used in various fields like virtual reality, gaming, and music production[11]. It can help create an immersive experience for the listener, giving the impression of being in a real room, with all its acoustical properties. It has been used in the development of many commercially successful products like Dolby Atmos, Auro 3D, and DTS:X.

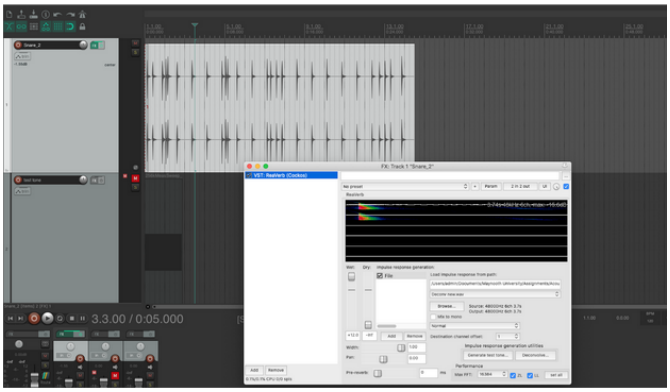

Figure 4: *Screenshot of ReaVerb plugin inside Reaper DAW*

8. Software System

A software application has been developed to emulate the octophonic room at Maynooth University's music department using binaural impulse response and convolution. This implementation is based on Csound and Cabbage, chosen for their benefits in providing a powerful and flexible audio programming environment.

The software application provides a user-friendly interface where the user can load a mono, stereo, or an 8-channel input audio file. The audio is then convolved using the recorded binaural impulse responses of the octophonic room, resulting in binaural audio output. The output can be listened to through headphones for an immersive audio experience.

The software application features 9 knobs, each of which controls the audio levels of a speaker in the octophonic room. These knobs allow the user to customize the audio output to their preference, creating a personalized audio experience. Additionally, a master fader is provided to control the overall audio level.

A meter is also available within the software application, allowing the user to visualize the audio signal. This visualization can be helpful in identifying any peaks or dips in the audio signal, aiding in audio adjustment.

Furthermore, the software application provides a "listen to source" button. This feature allows the user to listen to the original audio file before convolution, enabling a comparison between the source audio and the convolved audio.

In conclusion, the software application offers a powerful and flexible solution for emulating the acoustics of the octophonic room in Maynooth University's music department. It provides a user-friendly interface, customizable audio control, and helpful visualization features, making it an excellent tool for audio professionals and enthusiasts.

9. Practical Application

This system has several practical applications and commercial values that can be implemented in real-life scenarios. One such use is in virtual reality applications. The system can be used for musical applications to emulate different studio spaces worldwide. For instance, if a user wants to listen to their mixes in a specific studio space and speakers, they can capture the head-related impulse response of that studio and play back their audio through this system. Similarly, this system can be used to test mixes in concert spaces or performance venues before the actual performance. A mix that has been played in one's studio might differ when being played back in a performance venue. By capturing the impulse response of the venue or concert hall and sending it to the performers or composers, they can have an idea of how their music will sound at the venue and modify their piece accordingly.

Figure 5: *Screenshot of the developed software using Csound and Cabbage*

Furthermore, there is a significant scope in meditation purposes. Using this system, one can emulate different spaces and use them for meditation purposes. Since impulse responses are captured for the auralization process, the data can be used to analyze the room's acoustics and address any acoustical problems. The same data can also be used to modify the room's acoustical properties using physical elements like absorbers, diffusers, or digitally.

In terms of commercial value, this system can be used by audio engineers, music producers, and composers to check their mixes in different acoustic spaces. This can save time and resources by eliminating the need to physically visit different studios and venues. Additionally, this system can be used in the education and research fields to teach acoustics and sound engineering. This system's ability to auralize different acoustic spaces can help students understand the effects of different room acoustics on sound.

10. Existing system

There are several existing applications that utilize similar methods as our developed system, however, our system has unique features that differentiate it from others. One closely related application is the Slate VSX.

Slate Digital, a well-known brand in the digital audio processing industry, has introduced a similar system to the one we are developing. Their system, called Slate VSX, offers a hardware and software combination that is capable of emulating various studio spaces, car audio systems, and headphones[12]. However, there are some limitations to the Slate VSX system that our system addresses.

The software only works with the specific headphones provided by Slate Digital, making it mandatory to buy their hardware to use their software. While this has an advantage of translating more accurate representation of the room, this creates an additional cost and limits user choice . Even though it does not exist currently, In the future, the new system we developed will be implemented with a calibration part in the signal processing chain to use existing popular headphones.

Secondly, Slate VSX is expensive and out of reach for many home recording enthusiasts. But as csound and cabbage is an open-source platforms, users can access them for free. Once can add new IRs and modify the code to experiment with different rooms and sounds.

Thirdly, the system can only be used inside a digital audio workstation (DAW) and cannot be used as a standalone software to check mixes. Audio files cannot be loaded directly into the system; instead, they must be inserted as a plugin in the master channel of a DAW, which can be inconvenient for some users.

Fourthly, Slate VSX only supports stereo audio, and users are unable to use it for multi-channel audio. The system can only be used in a stereo channel.

Finally, the studio spaces, car audio system and headphone emulations are pre-defined, which means that users cannot add their own spaces into the system nor modify the existing system.

Although Slate VSX has some limitations, it has proven to be a popular system for emulating different studio spaces and headphone listening experiences. Our system builds upon the limitations of Slate VSX by offering a more cost-effective solution that is not tied to specific hardware and supports multi-channel audio.

11. Listening Experiments

To evaluate the accuracy of the emulated octophonic room using our technique, we conducted a listening experiment with a group of 7 participants. The group consisted of 3 musicians, 3 non-musicians, and 1 participant with hearing impairment. In the experiment, a snare sound convolved with the Head-Related Impulse Responses (HRIRs) was played back 10 times from different speakers among the 8 speakers emulated in the octophonic room. The participants were then asked to localize the sound using headphones.

The results of the experiment showed that the participants were able to accurately localize the sound in the emulated octophonic room. The localization was consistent across all the participants, including the participant with hearing impairment. This indicates that the emulated room was able to reproduce the spatial characteristics of the original octophonic room with a high degree of accuracy.

The listening experiment was conducted in accordance with the ethical guidelines for research involving human subjects. All participants gave their informed consent to participate in the experiment. The study was approved by the Institutional Review Board of the University.

The results of the experiment demonstrate the effectiveness of our technique in emulating real-world acoustic spaces with a high degree of accuracy. This has practical applications in a range of fields, including music production, virtual reality, and architectural acoustics.

Test Subjects	No. of Correct answers
Musician A	7
Musician B	7
Musician C	8
Non-Musician A	8
Non-Musician B	7
Non-Musician C	6
Hearing Impaired Subject	6

Table 1: *Listening Experiment Results*

12. Limitations

In order to provide a comprehensive understanding of the system, it is essential to discuss the limitations that our system faces. The first limitation is related to the physical differences of human beings. Each person has a unique shape of the ear, different head size, and different torso, which makes it challenging to develop a system that will auralize sound accurately for everyone. Thus, the system's accuracy is limited by the variation in the listeners' physical characteristics.

Another limitation is that binaural audio can only be played back using headphones. Although this allows for an immersive audio experience, it also means that the system is limited to headphone playback, which may not be ideal in some situations.

Finally, the listener's position and head movement are in a fixed position while listening to the binaural audio, which limits the experience that we get in a physical space. In contrast, in a physical space, we have the freedom to move our head and change our position, which can affect the sound's perceived direction and spatial characteristics. Therefore, our system's limitations should be taken into account when considering its potential applications.

13. Further Research

The system we have developed has a lot of potential for further research and development. Some of the future scopes for this project include upgrading the system with more inputs and multi-channel output, allowing for a wider range of applications beyond just binaural. Additionally, implementing a calibration process in the chain could make it possible to use any headphones with the system, rather than being limited to specific hardware.

Another area for further exploration is creating both standalone and plugin versions of the software, which would allow users to use the system without needing to open the cabbage application. Finally, implementation of ambisonics and other spatial audio formats could further improve the system and make it even more versatile for different applications.

These areas represent exciting possibilities for the future of this technology, and we look forward to seeing how it evolves in the years to come.

14. Acknowledgement

I would like to express our sincere gratitude to Dr. Iain McCurdy, who supervised the acoustic and psychoacoustic module of our course and provided invaluable guidance and support throughout the project. His expertise knowledge in the field of acoustics and psychoacoustics, and Csound has been instrumental in the success of our project.

I would also like to thank Prof. Victor Lazzarini for his guidance and teaching during the course. His knowledge and experience in the field of digital signal processing, and software programming have been invaluable in shaping this project.

References

[1] *The Oxford Handbook of Sound Studies*, ed. M. Grimshaw and T. Garner, Oxford University Press, 2014.

[2] Y. Zhang and Y. Liu, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2019, **27**, 1367–1379.

[3] D. R. Begault, *Audio*, 1994, **78**, 42–47.

[4] J. O. Smith, *Spectral Audio Signal Processing*, W3K Publishing, 2011.

[5] A. V. Oppenheim and R. W. Schafer, *Discrete-time signal processing*, Pearson, 3rd edn., 2010.

[6] A. S. Bregman, *Auditory scene analysis: The perceptual organization of sound*, MIT Press, 1994.

[7] J. C. Middlebrooks and D. M. Green, *Annual Review of Psychology*, 1991, **42**, 135–159.

[8] J. Blauert, *Spatial hearing: The psychophysics of human sound localization*, MIT press, 1997.

[9] V. Lazzarini, S. Yi, J. Heintz, Ø. Brandtsegg, I. McCurdy *et al.*, *Csound: A sound and music computing system*, Springer, 2016.

[10] R. Walsh, Proceedings of the 12th Sound and Music Computing Conference (SMC 2015), 2015, pp. 307–313.

[11] E. Chung and Y. Kim, *Applied Sciences*, 2019, **9**, 2321.

[12] S. Digital, *Slate VSX*, `https://www.slatedigital.com/slate-vsx-virtual-mix-room/`, 2023, Accessed: 2023-04-06.

[13] K. Müller and F. Zotter, *Acta Acustica*, 2020, **4**, 25.

[14] A. Roginska and P. Geluso, *Immersive sound: the art and science of binaural and multi-channel audio*, Taylor & Francis, 2017.

[15] J. Ahrens and C. Andersson, *The Journal of the Acoustical Society of America*, 2019, **145**, 2783–2794.

[16] B. Carty, *Maynooth Musicology: Postgraduate Journal*, 2009, **2**, 281–298.

[17] J. Thiemann and S. van de Par, *EURASIP Journal on Advances in Signal Processing*, 2019, **2019**, 1–9.

[18] M. Otani, H. Shigetani, M. Mitsuishi and R. Matsuda, *Acoustical Science and Technology*, 2020, **41**, 142–150.