Binaural Rendering Of Ambisonics - A 360 Degree Surround Sound Technology

A Project Report

Submitted by

Anupam N. Godse   111408016

in partial fulfillment for the award of the degree

of

B.Tech Information Technology

Under the guidance of

Prof. V. K. Pachghare

College of Engineering, Pune

AND

Mr. Kaustubh Ashtekar

NVIDIA Graphics Pvt. Ltd.

DEPARTMENT OF COMPUTER ENGINEERING

AND

INFORMATION TECHNOLOGY,

COLLEGE OF ENGINEERING, PUNE-5

May, 2018

DEPARTMENT OF COMPUTER ENGINEERING

AND

INFORMATION TECHNOLOGY,

COLLEGE OF ENGINEERING, PUNE

CERTIFICATE

Certified that this project, titled "Binaural Rendering Of Ambisonics - A 360 Degree Surround Sound Technology" has been successfully completed by

Anupam N. Godse   111408016

and is approved for the partial fulfillment of the requirements for the degree of "B.Tech. Information Technology".

SIGNATURE
Prof. Vinod K. Pachghare
Project Guide
Department of Computer Engineering
and Information Technology,
College of Engineering Pune,
Shivajinagar, Pune - 5.

SIGNATURE
Prof. Vandana Inamdar
Head
Department of Computer Engineering
and Information Technology,
College of Engineering Pune,
Shivajinagar, Pune - 5.

Acknowledgement

I express my sincere gratitude towards Prof. Vinod K. Pachghare for his constant help, encouragement and inspiration throughout the project. I am fortunate to have guide who gave me freedom to explore on my own and help throughout this process by providing timely review, constructive comments and valuable suggestions. Without his guidance his project wouldn't have been possible. I would also like to thank Mr. Kaustubh Ashtekar, NVIDIA Graphics Pvt. Ltd. for guiding me throughout the journey. I am very grateful for his expertise that helped in understanding all the crucial aspects related to the project. I am also thankful to all other resources such as research papers and members across communities that helped me to find relevant problems and their solutions.

Abstract

Recently ambisonics format has gained popularity as directional/spatial audio encoding format for 360 degree videos, virtual reality, etc., with major video distribution platforms such as Youtube and Facebook adopting it for 360 degree videos. One of the most important characteristics of ambisonics is that it does not require the layout of speakers to be predefined for encoding. Rather the encoded representation can be decoded for any given speaker layout, which provides users, the flexibility to choose any layout of speakers and decode the given ambisonics representation for the same. The first order ambisonics encoding of a sound field requires four channels of an audio stream and the directional information (spatialization) can be further improved by going for higher order ambisonics encoding with a larger number of channels. Rendering spatial audio requires a large number of speakers (6, 8 speakers for 5.1, 7.1 surround respectively) placed in a specific way around the listener. All this hardware setup can be replaced with a headphone and an ambisonics to binaural rendering software. Binaural rendering is based on the concept of creating the effect of a virtual speaker on headphones using Head Related Transfer Function (HRTF). The aim of this paper is to present the studies which focus on positives of ambisonics over the traditional surround sound techniques and the method for implementing the ambisonics binaural rendering system.

Contents

List of Figures

Chapter 1

Introduction

1.1 Ambisonics

There has been a tremendous success in the virtual reality applications due to the advancements in graphics. However, the virtual reality experience is incomplete if its audio part is ignored because then the experience lies far from reality. Therefore, audio is considered one of the important factors while evaluating realism in the virtual reality. Since then, Ambisonics has been in focus for audio implementation in virtual reality applications. Ambisonics is a method of recording and reproducing audio in full 360-degree surround. Ambisonics treats an audio scene as a 360-degree sound sphere around center point coming from different directions. Center point is where the microphone is placed while recording, or where the listeners sweet spot is located while rendering. [4] Traditional surround sound technology has several drawbacks. These techniques work only on the predefined array of loudspeakers to produce the output sound field which is the most important drawback of this technology. By contrast, Ambisonics doesnt render the audio signal for the predefined set of speakers but it can render audio on the y for any user defined speaker array. It not only works for static but also for rotating sound field i.e. it works for real time applications. [3] When the sound field rotates the sound tends to jump from one speaker to another when we use a traditional approach. Ambisonics uses a number of virtual speakers so the transformation is smooth even when the sound field is rotated. Traditional surround sound techniques are front biased but ambisonics distributes the sound evenly in 3D space. [3] Traditional techniques had difficulties in representing sound beyond the horizontal dimension. Whereas ambisonics works with the elevation as well, and thus the effect is more immersive. [3] Rendering of ambisonics file format over speakers requires a minimum of 4 speakers for the first order Ambisonics and the number of speakers required are more in case of higher order Ambisonics. [5] Further, we need Ambisonics technology to work for mobile devices as well where we cannot have loudspeakers and the sound should be rendered on headphones. This can be achieved using a virtual Ambisonics approach which has 2 steps 1) decoding sound over an array of virtual loudspeakers 2) using the corresponding pair of HRTFs to filter an output of each loudspeaker to get a binaural output. [16] Moreover, the output of this can be improved by incorporating the head-tracking so that output can be made more accurate with consideration of head-movements. [6] [7] [10] [16]

1.1.1 Ambisonics B-format

There are various encoding formats used for storing ambisonicsB-format is widely used format for recording sound field using Ambisonics technique. It has 4 channels: W, X, Y and Z.

W : Omni directional sound pressure.

X : Front-Back direction with respect to the listener.

Y : Left-Right direction with respect to the listener.

Z : Up-Down direction with respect to the listener.

1.1.2 Recording and Encoding B-format

Recording is done with the help of special sound field microphone. It has one omni-directional microphone (the W channel) and three eight shaped microphones (the X, Y and Z channels). The microphone constructed by arranging four cardioid capsules as shown in below figure 1.1, and to get the desired patterns these can be combined as needed. There are also other types of microphones available which are more sophisticated.

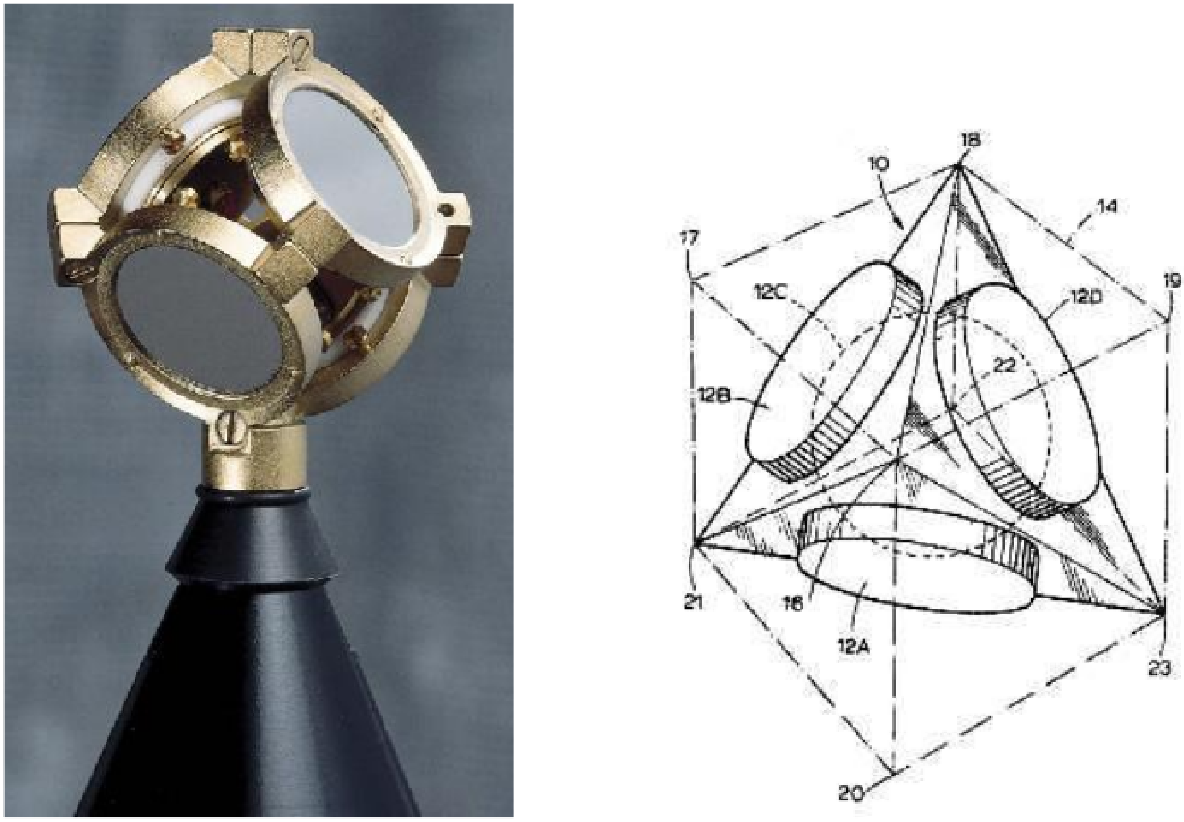

Figure 1.1: Microphone to record sound-field

1.1.3 Decoding

The decoders job is to produce loudspeaker signals that create a good illusion of the required directional sound field.[5] For decoding the ambisonics file on the desired array of available speakers, the decoder matrix is constructed using the available azimuths and elevation angles of the speaker positions which acts as a filter for the input. The Ambisonics format can be rendered on any speaker layout using suitable decoder.

1.2 Binaural Rendering

Binaural rendering is converting the output of individual speakers to headphone output (binaural left and right) by applying Head Related Transfer Functions (HRTFs) as shown in below figure 1.2 and then combining results into single left and right binaural output.

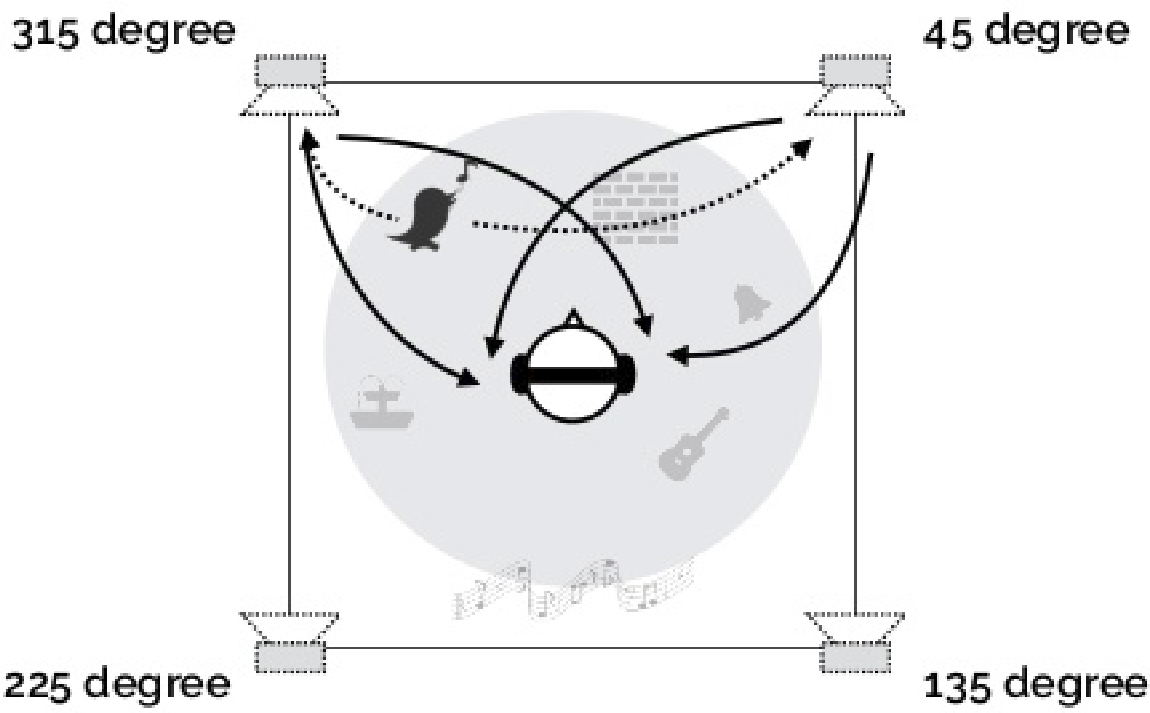

Figure 1.2: Binaural Rendering

1.2.1 HRIR and HRTF

HRIR (Head Related Impulse Response) Humans detect the sound source by taking derived cues from one ear and by comparing cues from both the ears.

The cues have two differences one is time difference and another one is the intensity difference between cues of both ears. The sound source interaction with the human body modify the original sound before it enters the ear. These modifications can be portrayed with the help of the HRIRs, the head-related impulse response, which locates the source location. HRIRs help to convert the sound so that it appears to the user to be played at the desired location. They are used to generate the sound effects in virtual surroundings. The Head Related Transform function is what we get when we perform the Fourier transform on Head Related Impulse Response.

1.2.2 CIPIC HRTF Database

The CIPIC HRTF Database is a public-domain database of high-spatial-resolution HRTF measurements for 45 different subjects. The database includes 2,500 measurements of HRIR for each subject. 25 different interaural-polar azimuths and 50 different interaural-polar elevations were considered for taking the measurements for each subject. [8]

1.2.3 The virtual ambisonics approach

To transform the sound-field into the binaural audio, there is need to decode ambisonics on virtual array of speakers and then further applying HRTFs on each mono output of speaker to generate a binaural signal from each speaker which can further be superimposed to get the final headphone output.

(Figure 1.1 Source: By The original uploader was Soumyasch at English Wikipedia - Transferred from en.wikipedia to Commons., CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=3848567)

How an human ear experiences a signal from a direction in space is characterized by a Head related Transform function. To synthesize a binaural

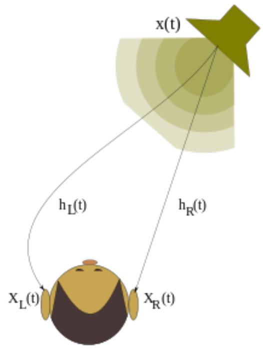

Figure 1.3: HRTF filtering effect

sound from a particular point in space a pair of HRTFs (left and right) can be used. In summary, HRTF is a transfer function which describes how a sound from a specific point in space will arrive at the ear. The filtering effect of the sound source can be described by the HRTFs for the left and right ears before the signal reaches the left ear and the ear as shown in figure 1.3 above.

1.3 System Overview

The following figure 1.4 gives the system overview. It takes 4 channels as input (W, X, Y and Z) and generated a 2 channel i.e. binaural headphone output.

Inputs: 4 Channels

W : Omni directional sound pressure.

X : Front-Back direction with respect to the listener

Y : Left-Right direction with respect to the listener

Z : Up-Down direction with respect to the listener

Output: 2 Channels

Left(l) : Left headphone output

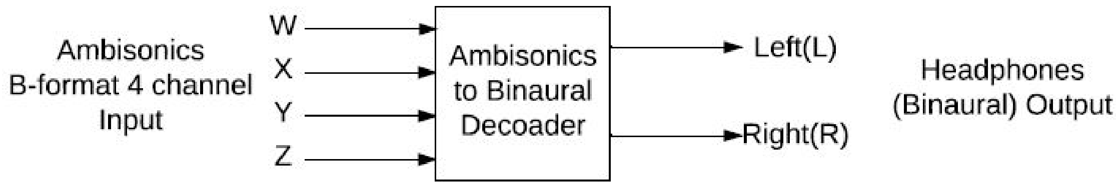

Figure 1.4: Ambisonics to Binaural Decoder System Overview

Right(r) : Right headphone output

1.4 Problem Statement

Given an ambisonic audio encoded file having 4 channels, W, X, Y and, Z, generate a 2 channel binaural output.

1.5 Objectives

Following are the objectives of the project:

1. The system should parse the ambisonic file to get the meta-data of the audio. For example, the number of channels, the size of the audio samples, the total number of audio samples, the sampling rate of the file etc.

2. Further, it should be able to extract the audio samples and construct the input matrix accordingly.

3. Next, the system should be able to build the customized filter from the angles of the sound input and the HRTFs extracted from the CIPIC HRTF database.

4. Finally, the system should be able to apply the filter on the input channels and generate the 2 channel binaural output

1.6 Motivation

Virtual reality has gained tremendous popularity in the recent years and surely will have great market in the future. It has many applications in gaming, movies, and, almost everywhere where there is involvement of video and audio. In the past years there has been great research on video aspect of virtual reality and there are many great technologies developed for the same. However, to get the immersive experience of the virtual reality, not only video needs to be have realism but, without audio, virtual reality is far from achieving its goal of immersive reality experience. Audio is as important aspect as video when it comes to realism in virtual reality. The environment which we are virtually rendering should have directional audio effect for it to be truly experienced and enjoyed by the users. Having flat sound would result in unpleasant experience to the users. So, developing a system to gain immersive audio experience is one of the major tasks of the virtual reality system as whole. Furthermore, developing the binaural system so that the need for the speakers and related hardware can vanish and audio realism can be experienced with just the use of headphones. The need of audio rendering system for virtual reality and making it flexible so that rendering should no more depend on the predefined layout of speakers drives me to undertake the development of the binaural rendering system as my final year BTech project.

1.7 Methodology

There are several ways to achieve realism using audio in virtual reality. Monophonic systems merely give distance information of audio, stereo systems add the front limited angle sector information and then comes the rendering on speakers with each speaker having channel information attached to it and therefore the layout has to be predefined. Increasing the number of speakers increase the immersive effect. Ambisonics is another way to achieve the same with much more flexibility and more immersive experience. So for developing the system ambisonics technology is used. Furthermore, as it is impossible to render ambisonics on just two speakers i.e. on earphones, it can be done by using virtual ambisonics approach. The virtual ambisonics approach is to render the ambisonics on many number of virtual speakers and then transform each speaker's output to the binaural output using the head related transform function(HRTF) filter which is extracted from the CIPIC HRTF database and then superimposing all the left and right outputs to get the final binaural output.

1.8 Organization of report

The report is carefully organized to guide the reader to gain the perspective of the needs and the approach used to develop the system. The report begins with giving the brief introduction to the ambisonics technology, then it explains binaural rendering and the related concepts, next it explains in depth the related research work in ambisonics and ambisonics to binaural rendering systems and what aspects of those research work are useful to develop the system. Later, it explains the system architecture diagram which gives the overview of the different blocks used in development of the system. Further, it explains how the experimentation was performed during development of the system and the corresponding results. The report concludes by giving the brief conclusion and the scope for the future work.

Chapter 2

Literature Survey

2.1 Traditional approaches and their drawbacks

Michael Gerzon has criticized the traditional surround sound approaches and also has given the criteria for the design of the surround sound systems [1]. The traditional quadraphonic systems never gave the optimum results. The aim of these systems was to duplicate the effect of 'original 4 track tape', but they failed to do so [3]. Peter Fellgett said that the existing techniques were inadequate in a number of ways like they restricted to a fixed number of speakers and the production needs of 4 channels to be available [3]. Moreover, these techniques rely on encoding the speaker channel information which can be rendered on predefined speaker layout only - 5.1, 7.1 - otherwise doesnt give the intended effect. In addition, the traditional surround techniques were limited to horizontal plane excluding the height attribute. These techniques are only suitable when the image is stable and doesnt suite well for real time applications due to audio scene rotations and the output jumps from one speaker to another as there is a fixed discrete predefined speaker layout. These existing approaches resulted in poor conditions even under ideal surroundings [3]. They suffered from hole in the middle effect and if the situation is less ideal it becomes unusable. For example, when the room is non-square or when the listener is not at the sweet spot [5]. The use of the 4th channel always degraded the localization quality, the mentioned hole in the middle effect. Thus only 3 channels were recommended and the use of 4th channel was still a question. This led to the addition of the periphonic (height) information [3].

While traditional technique of surround sound had its limitations and disadvantages, Ambisonics, which was developed by Peter Fellget [3] and Michael Gerzon [4] is the way to record, encode, store and reproduce the soundfield in 3D, which gave more immersive experience to the listener and provided full upward compatibility to any number of loudspeakers in the user defined configuration. The traditional approaches failed to give the intended immersive audio effects, they required a significantly higher number of channels to improve the sound quality, they required the speaker layout to be predefined and needed the listener to be present at only a particular position i.e. the sweet spot. These were the disadvantages of the traditional approach.

2.2 An Ambisonics approach

Monophonic reproduction merely provided information about direction and distance only. Then the stereo added explicit information for front sector not more than 60 degrees in width [3]. Apart from this, various techniques were suggested by using more loudspeakers, more channels, extending the directional information beyond 60 degrees. As these are separate ways to attain realism by improving various factors, Ambisonics aimed to combine these as an integrated whole and thus adding simplicity [2]. To record, to convey and to regenerate the accurate and repeatable surround sound with the perfect directional effect was the main aim of the ambisonics technology [3]. It is the technology for surround sound which aims not to make the loudspeakers audible as separate sources of sound but instead recreate the sound field [1, 2]. Ambisonics technique can be used with any number of loudspeakers with reasonable configuration thus providing for full upward compatibility. Moreover, the technique does has no limit to the number of channels, the more the number of channels the higher is the spatial resolution [5]. The technique is based on a higher accuracy and unequivocal specification of how the encoding should handle directionality in contrast with the quadraphonic approach which handled only 4 directional signals [5]. It defines encoding such that all the directions are equally covered in contrast to the traditional techniques [5]. Why ambisonics is good because it covers 360 degree information sound with a limited number of channels. 4 channels (first order Ambisonics) can be rendered on 4 or more speakers with user defined speaker layout. To improve the directionality effects further, Higher order Ambisonics can be implemented which will require a higher number of speakers i.e. higher the order, more the number of speakers will be required. This contributes to higher spatial resolution. Ambisonics, in contrast to traditional surround sound techniques, can create a smooth, continuous and stable sound field even when the sound field rotates and this is because it is not predefined for any particular speaker layout, thus suitable for real-time applications.

A Stanford paper by Cedric Yue and Teunde Planque emphasis on the importance of implementing ambisonics based 360 degree audio system for a quality feel of virtual reality. They rightly point out that a good amount of work has been done in the field of virtual reality focusing on graphics part of it and audio part is given less attention to date. [8]

2.3 The Virtual ambisonics approach

For making 360 degree audio suitable for mobile applications and rendering it over headphones, Markus Noisternig, Thomas Musil, Alois Sontacchi and Robert Holdrich introduces a virtual ambisonics approach in the paper named 3D Sound Reproduction using a Virtual Ambisonics Approach. This paper summarizes how the virtual ambisonics approach can be used to emulate the 360 degree immersive audio experience with headphones using HRTFs and the intermediate rendering over virtual speaker array. It explains how a convincing binaural sound reproduction requires filtering the sound sources with the HRTFs. Moreover, it suggests incorporating head-tracking for further improvements in localization. [6] [7] [10].

Angelo Farina and Emanuele Ugolotti described the software implementation of B-format encoding and decoding. They introduced the basic decoding equation which computes the feed Fi for specific speaker in loudspeaker array, where α, β and, γ are angles between the sound field vector and the X, Y and Z axes.

$$Fi = 1/2 * [G1 * W + G2 * (X * cos(\alpha) + Y * cos(\beta) + Z * cos(\gamma)] \quad (2.1)$$

They have specified the values of G1 and G2 gains in the paper for different regular array configurations. However, this equation works merely for regular or nearly regular shaped configuration of speakers such as square, hexagonal etc. and doesn't work well for irregular configuration of speakers. [14]
On the other hand, Markus Noisternig, Thomas Musil, Alois Sontacchi and Robert Holdrich, have introduced set of equations using the Morre Penrose

pseudo inverse method to work for irregular speaker configuration as well. If P is the vector denoting input to the sources, 1st order ambisonics B format is given as:

$$B = C * P \tag{2.2}$$

Now, as we already have B i.e. the Ambisonics B-format (W, X, Y, Z channels) and we need to regenerate P. Then P can be calculated as:

$$P = pinv(C) * B \tag{2.3}$$

Here C is the encoding matrix generated from the speaker configuration i.e. by considering azimuth and elevation of each speaker (each column represents one speaker) and pinv is the pseudo inverse. Each column vector in C matrix is

$$[0.7071, cos(\theta) * cos(\phi), sin(\theta) * cos(\phi), sin(\phi)] \tag{2.4}$$

where θ is azimuth angle and ϕ is elevation angle. Thus, P matrix has the mono output signal for each loudspeaker. on which HRTFs can be applied further to get a binaural output from each speaker which can be further superimposed to get final left and right headphone outputs. [6] [7] [10]

2.4 Development of Binaural System

After rendering over the virtual speaker array HRTFs are used to filter these signals of each speaker and converting mono to left and right signals for each speakers. When these all signals are superimposed we get a single left and single right (Binaural) headphone output signals.

2.5 Discussion

Shu-Nung Yao has described the customization and real-time implementation of binaural rendering by asking the listener to select the closest matching dataset from the database and finally presenting both the subjective and objective measurements for the experienced audio quality. [12]

In the presentation paper by Bruce Wiggins, he discusses the algorithm used by the Google and analyses the virtual ambisonics approach for binaural rendering with respect to inter-aural time, level and, spectrum differences. He has implemented 1 to 35th order ambisonics and carried our corresponding analysis and the results and conclusions are given in the paper. [13]

Reading various research papers on why did ambisonics technology was introduced points us to the drawbacks of the traditional approaches. Furthermore reading about papers on ambisoncs gives us insight of why is the technology useful and how it gives flexibility in the applications. It introduces us to the concept of Head related transform functions which are helpful to transfer speaker output to binaural output. Moreover, reading about virtual ambisonics approach gives us overview how to develop the binaural rendering system.

Chapter 3

Problem Statement

3.1 Problem Definition

Design a system which takes ambisonics B-format (4 channels, W, X, Y and Z) as an input and generates output for headphones i.e. binaural output (2 channels).

3.2 Outcomes

1. The system is expected to work with minimum latency i.e. it should generate binaural output without any time delay, this ensures that it works for real time applications.

2. The headphones output should clearly able to distinguish between sounds from different directions.

Chapter 4

Requirements Specification

4.1 Functional Requirements

1. The system must take an ambisonics B-format (4 channel) file as input and give 2 channel output.

2. The head-tracking device inputs should be taken into consideration for real time applications and output should be generated accordingly.

3. It should work for 360 degree videos, real-time high-end gaming and other Virtual Reality applications.

4.2 Non-Functional Requirements

1. **Usability:** The system should be easy to integrate with any Virtual reality application.

2. **Performance:** The system should be efficient to provide the output by reducing the number of computations.

3. **Latency:** The latency must be very low to be suitable for real time applications.

4.3 Hardware Requirements

1. **RAM:** It should have minimum 2 GB RAM.
2. **Hard Disk:** It should have minimum 40 GB of free space.
3. **Graphics Card:** System should have a sufficient graphics card to support the required Virtual Reality Application.

4.4 Software Requirements

1. **C++ Platform**
2. **MATLAB:** 5x or higher version
3. **CIPIC HRTF Database**

Chapter 5

System Design

5.1 System Architecture Diagram

The system implementation can be broadly classified into 3 steps:

1. Rendering of audio over array of virtual speakers.
2. Applying Head related transfer function to each virtual speaker output to transform it to binaural output.
3. Superimposition of all left and right outputs into a single left and right binaural output.

Figure 5.1 illustrates the system architecture diagram of the proposed system. Various blocks of the system are explained below.

5.2 Different functional system blocks

1. **Generating the Decoder Matrix** The speaker configuration (the azimuth and elevation for each speaker) will be taken as input and this function will generate a decoder matrix.
2. **Rendering Output to Virtual Speaker Array** This function will take the 4 channels (W, X, Y, and Z) and the decoder matrix as inputs

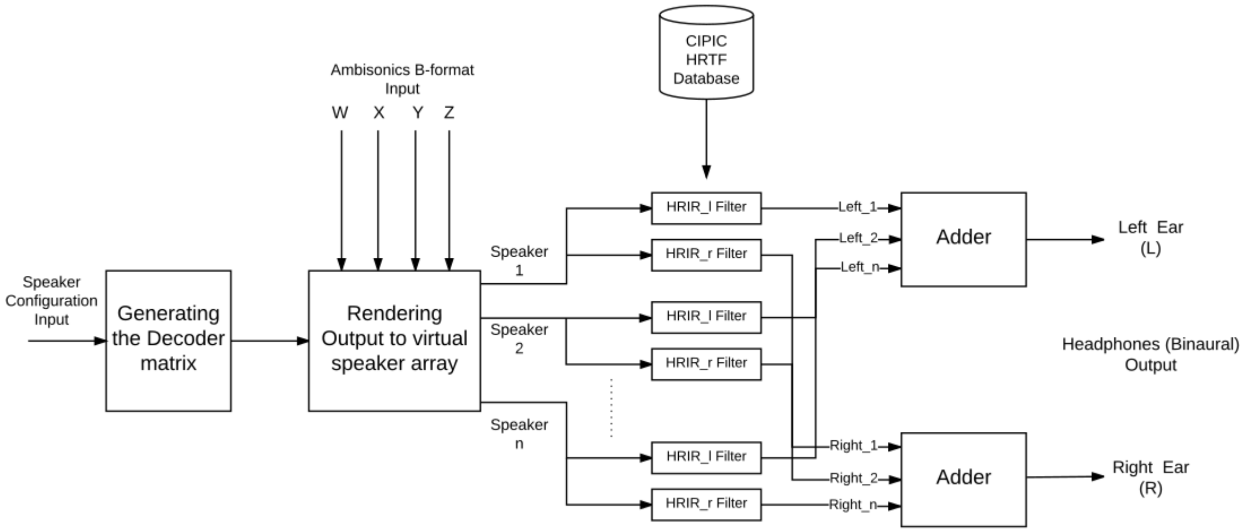

Figure 5.1: Ambisonics to Binaural Decoder System

and generate a mono output for each speaker of the speaker array.

3. **The CIPIC HRTF Database** This is the Database which has the HRIR pairs (left and right) for the range of azimuth and elevation pairs for each of the speakers.

4. **Adder** This block will superimpose all the outputs from each HRIR-l and HRIR-r filters and generate a single left and right final binaural audio.

Chapter 6

Experimentation and results

6.1 Experimentation

The system was tested on customized encoded ambisonics B-format file. The file has sounds for which all four directions viz. Front-left, front-right, center, back-left and back-right can be tested. This file was given as an input to the implemented ambisonics to binaural decoder system. The system output was a file with 2 channels i.e. the desired binaural output. Similarly it was tested on other ambisonics files from the Internet and desired results were obtained.

Furthermore, the system was also tested with inclusion of rotations in the sound field. The scene was rotated by +90, +180, etc. in the XY plane degrees and the output was taken. It was observed that the output given was correct. For example, when the rotation desired was 90 degrees the sound from front-left speaker was heard from front right direction, the one from front-right was heard from back-right direction and similar effect was observed for other speakers.

6.2 Results

The aim of the system was to take a ambisonics B-format file as an input and get a binaural output so that the user perceives directional effects in the headphones itself. When the system will be incorporated with some game or real time applications the input instead of .wav file would be just 4 channel streams of W, X, Y, Z channels which is anyway achieved in the current system by extracting them from the .wav file. Moreover, the head tracking co-ordinates will be taken and the input will be multiplied by the corresponding rotation matrix.

The results generated were with the minimum latency and thus the system can be used for real time applications in VR.

The current system has aim to take 4channel input and give binaural output and which was achieved successfully.

Following are the results in the graphical format:

Test File 1 : Customized ambisonics B-format encoded file
Test File 2 : An ambisonics B-format file from Internet
Test File 3 : An ambisonics B-format file from Internet

The results are as desired and the testing was successfully completed.

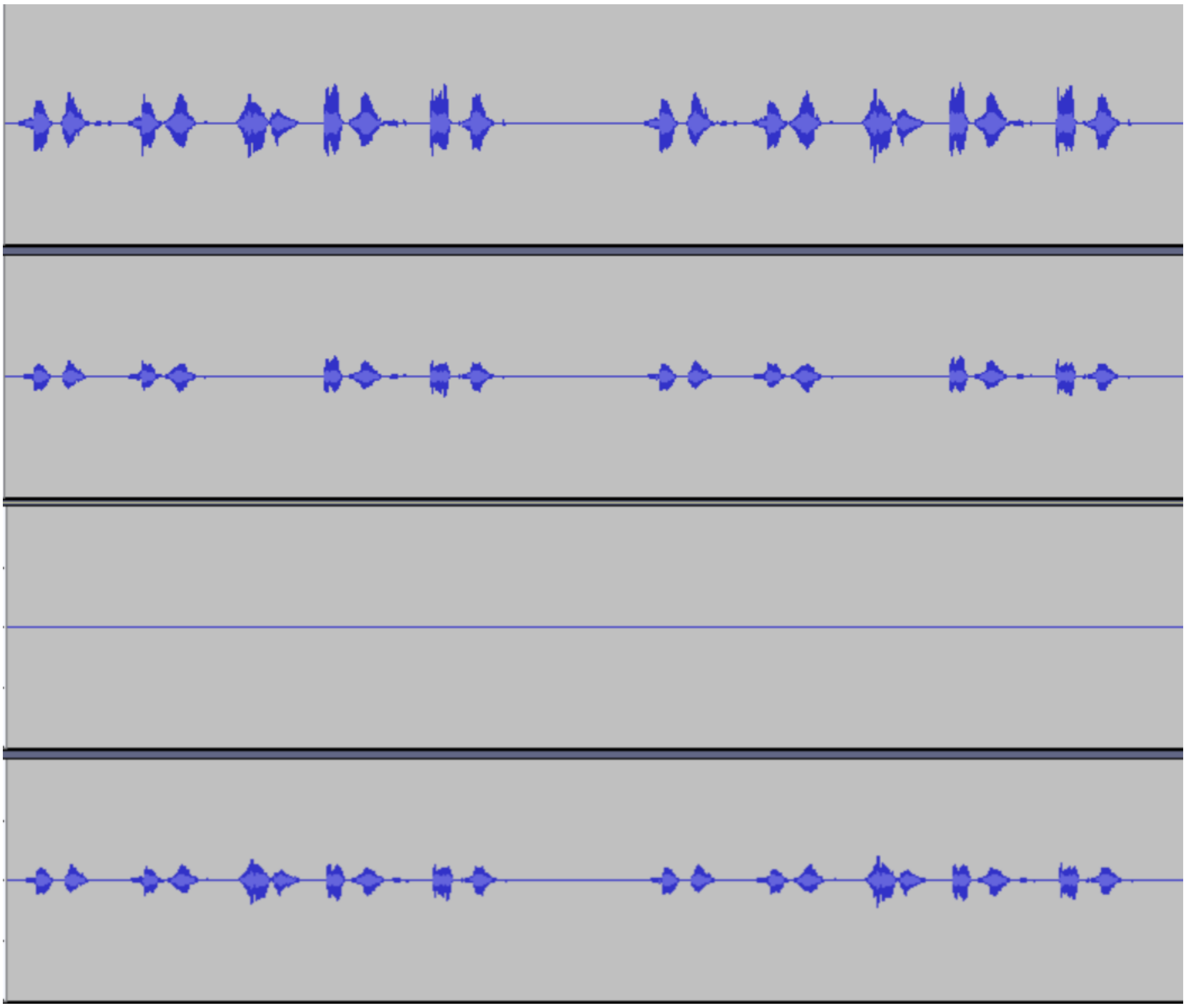

Figure 6.1: Test File 1

Figure 6.1 illustrates the amplitude time graph of audio file (Test file 1) which is the input file to the system with four channels W, X, Y and Z. The horizontal axis represents time and the vertical axis represents the amplitude of the audio sample.

Figure 6.2 illustrates the amplitude time graph of output file (Output file 1) which is the output file generated by the system with 2 channels. The horizontal axis represents time and the vertical axis represents the amplitude of the audio sample.

Figure 6.3 illustrates the amplitude time graph of audio file (Test file 2) which is the input file to the system with four channels W, X, Y and Z. The

Figure 6.2: Binaural Output of Test file 1

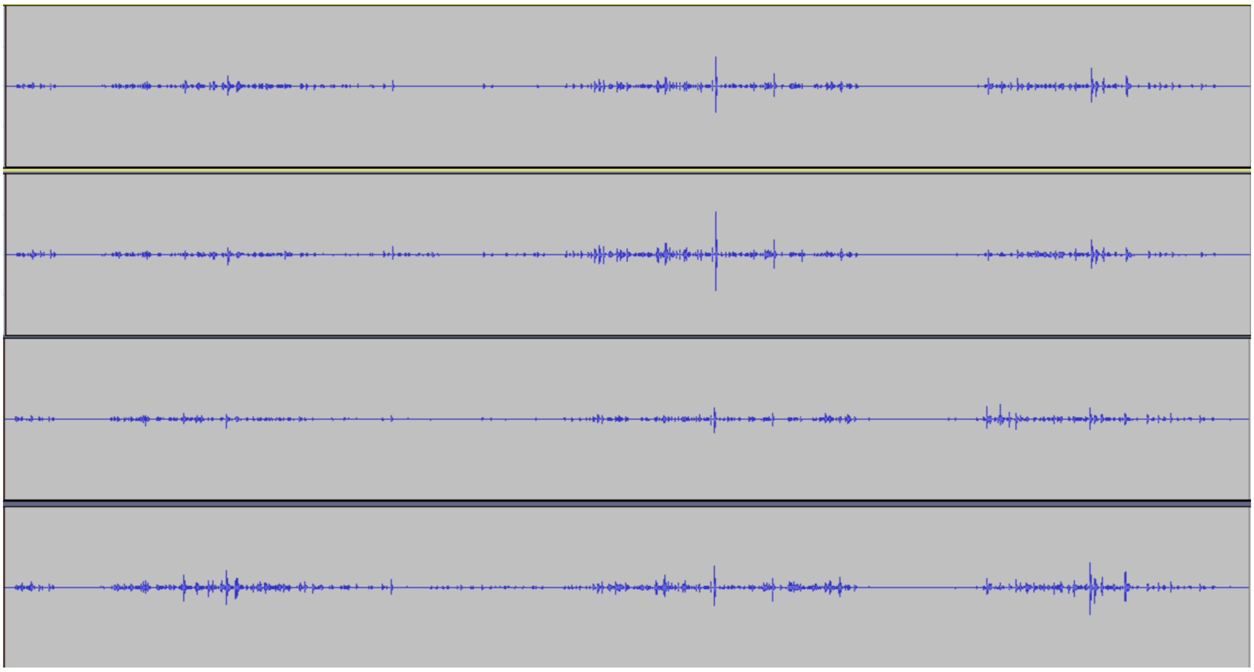

Figure 6.3: Test File 2

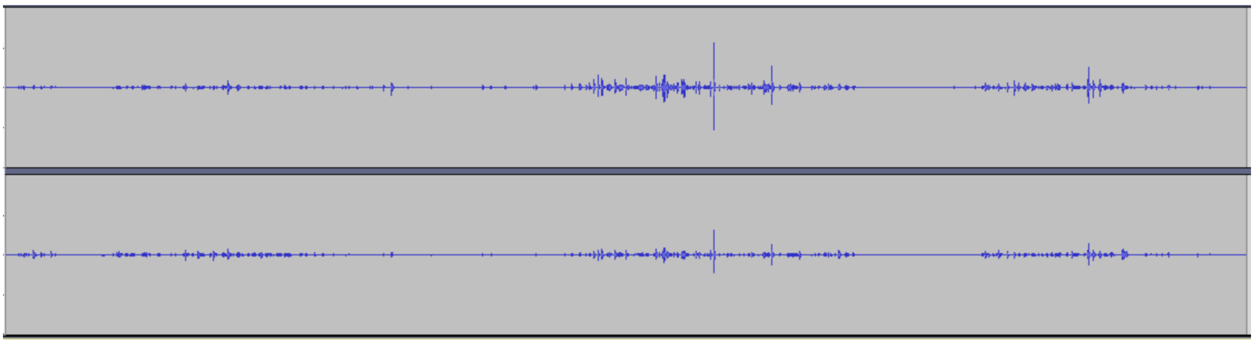

Figure 6.4: Binaural Output of Test file 2

horizontal axis represents time and the vertical axis represents the amplitude of the audio sample.

Figure 6.4 illustrates the amplitude time graph of output file (Output file 2) which is the output file generated by the system with 2 channels. The horizontal axis represents time and the vertical axis represents the amplitude of the audio sample.

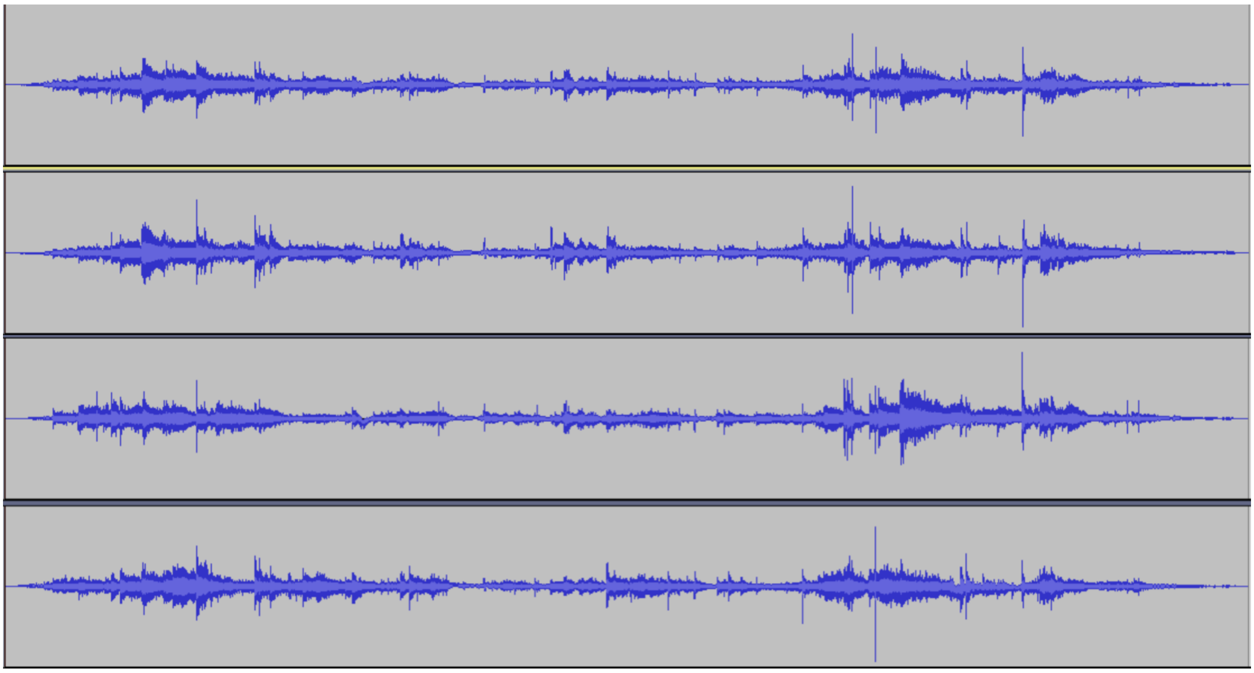

Figure 6.5: Test File 3

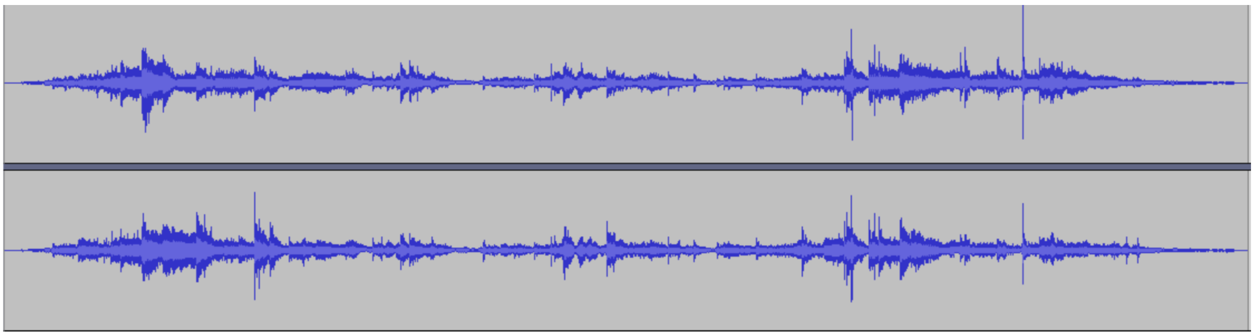

Figure 6.6: Binaural Output of Test file 3

Figure 6.5 illustrates the amplitude time graph of audio file (Test file 3) which is the input file to the system with four channels W, X, Y and Z. The horizontal axis represents time and the vertical axis represents the amplitude of the audio sample.

Figure 6.6 illustrates the amplitude time graph of output file (Output file 3) which is the output file generated by the system with 2 channels. The horizontal axis represents time and the vertical axis represents the amplitude of the audio sample.

The system performed very well to generate the binaural left are right output which renders on headphones. The system was as shown tested on two files one customized and another from the Internet and it gave the desired results. The system was also tested by rotating the sound-field with various angles and the results were logical for those rotations too.

Chapter 7

Contributions

Following were the outcomes of the implementation of the the project.

1. As the sequence of application of filters does not affect the final result, the system first combines both the filters i.e. the Head related transform function filter and the filter which renders ambisonics to virtual array of speakers.

2. As a result it has achieved 300 times performance gain in processing of input files.

3. The filter generation is made a one time task by generating the filter once by combining Head related transform function filter and the filter which renders ambisonics to virtual array of speakers and storing it in an array, this again helps in improved performance.

4. The system provides flexibility to the user to give file with any sampling rate, of any length and, with any size per data sample.

5. The system allows user to give file in any one of the FUMA or ambiX format of ambisonics B-format.

6. The design of the system is robust and has minimal latency which is perfect for real time applications.

Conclusion and Future Work

Conclusion

Traditional approaches had many drawbacks such as predefined layout of speakers and hence were not suitable for real time applications. When ambisonics was introduced it overcome the drawbacks and provided more flexibility to the users. The traditional approaches had fixed channel for each speaker and hence when the layout was changed, the channel information needed to be changed and therefore was not suitable for real-time applications. It is understood that indeed ambisonics has many advantages over the traditional approaches. It can also be used for the real-time applications by applying the appropriate rotations over the matrices. It gives the better audio effects compared to the previously used approaches and that is why the technology is adopted by Facebook, Google and many other companies which work in Virtual Reality area. It has a wide range of applications in 360-degree videos, high-end gaming, and other virtual reality applications. Combining ambisonics technology with the virtual ambisonics approach to generate the binaural output has advantages of eliminating the need for multiple loudspeakers and has an additional advantage of working well for mobile applications.

Future Work

The implemented system can be extended in following ways.

1. The implemented system work can be extended to consider the head tracking coordinates from the VR devices and improve localization by applying rotations on the sound-field audio. This will make it more suit-

able for real time applications such as gaming and other VR experiences. In gaming, the player's head co-ordinates will be tracked by Virtual reality headsets and then these co-ordinates will help to improve the effects at real-time.

2. Moreover, the system can be extended to support the panning to rotate sound-field when the pan is rotated to get the desired effects which may be suitable for other required applications.

3. Currently the HRTF is taken from the CIPIC HRTF database which was recorded for number of subjects. This can be extended to consider the HRTF of the user to get the most accurate effects as each person has different HRTF because of the different head structure of each person.

References

1. Michael Gerzon Surround Sound Psychoacoustics, Wireless World, December 1974, Vol. 80, pp 483-486.

2. Michael Gerzon, What's Wrong with Quadraphonics?, Studio Sound, Vol. 16 no.5, pp. 50, 51, 56, May 1974.

3. Peter Fellgett, Ambisonics. Part one: general system description, Studio Sound, Vol. 17 no. 8, pp.20-22, 40, August 1975.

4. Michael Gerzon, Ambisonics Part two: Studio Techniques, Studio Sound, Vol. 17 no. 8, pp. 24-26, 28, 30, August 1975.

5. M.A. Gerzon, Ambisonics in multichannel broadcasting and video, Journal of the Audio Engineering Society, 33, no. 11, 1985.

6. Noisternig M., Musil T., Sontacchi A., Holdrich R., A 3D Real Time Rendering Engine for Binaural Sound Reproduction, in Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA, 6-9, July 2003.

7. Noisternig M., Musil T., Sontacchi A. Robert Holdrich, 3D Binaural Sound Reproduction using a Virtual Ambisonic Approach, in Proceedings of VECIMS 2003 - international Symposium on Virtual Environments, Human Computer Interfaces, and Measurement Systems Lugann, Switzerland, 27-29 July 2003, pp 174-178.

8. Cedric Yue, Teun de Planque, 3-D Ambisonics experience for Virtual Reality, Stanford University, 2017.

9. V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, The CIPIC HRTF database, in Proceedings of the 2001 IEEE Workshop

on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575), New Platz, NY, 2001, pp. 99-102.

10. Noisternig M., Musil T., Sontacchi A., Holdrich R. A 3D Ambisonic based Binaural Sound Reproduction System, International Conference on Multichannel Audio Audio Engineering Society, 2003.

11. Davis, Larry S. and Duraiswami, Ramani and Grassi, Elena and Gumerov, Nail A. and Li, Zhiyun and Zotkin, Dmitry N., High Order Spatial Audio Capture and Its Binaural Head-Tracked Playback Over Headphones with HRTF Cues, Audio Engineering Society Convention 119, October 2005.

12. S. N. Yao, Headphone-based immersive audio for virtual reality headsets, in IEEE Transactions on Consumer Electronics, vol. 63, no. 3, pp. 300-308, August 2017.

13. Wiggins, B., Analysis of binaural cue matching using ambisonics to binaural decoding techniques, Presented at 4th International Conference on Spatial Audio, Graz, Austria, 7-10 September 2017.

14. Angelo Farina, Emanuele Ugolotti, Software Implementation of B-Format Encoding and Decoding, Audio Engineering Society Convention 104, May 1998.

15. Richard Brice, Ambisonics White Paper, Pspatial audio, October 2014.

16. G. Enzner, M. Weinert, S. Abeling, J. M. Batke and P. Jax, "Advanced system options for binaural rendering of Ambisonic format," IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, 2013, pp. 251-25.