Binaural rendering of Ambisonics – a 360-degree surround sound technology

Anupam Godse, Vinod Pachghare, Kaustubh Ashtekar

Anupam Godse is Final Year Bachelor of Information Technology Student at Government College of Engineering, Pune, India. (phone: +918087915817; e-mail: godsean14.it@coep.ac.in).

Vinod Pachghare is an associate professor at Computer Engineering and Information Technology Department of College of Engineering, Pune, India. (e-mail: vkp.comp@coep.ac.in).

Kaustubh Ashtekar is a Senior Software Engineer with NVIDIA Graphics Pvt. Ltd., Pune, India (e-mail: kashtekar@nvidia.com).

**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.

*Keywords:*

360 degree audio, Ambisonics, audio in virtual reality, audio spatialization, binaural rendering, surround sound

# **INTRODUCTION**

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 listener’s 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 doesn’t render the audio signal for the predefined set of speakers but it can render audio on the ﬂy 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. **KEY TERMINOLOGIES**

## **Ambisonics B-format**

B-format is widely used format for recording sound field using Ambisonics technique. It has 4 channels: W, X, Y, and Z. W: Omnidirectional 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. [15]

## **Recording**

Recording is done with the help of special sound field microphone. It has one omnidirectional microphone (the W channel) and three figure-of-eight microphones (the X, Y and Z channels). It is made up of four cardioid capsules arranged in a tetrahedron, which can be combined as needed to provide the desired polar patterns. [15] There are also other types of microphones available for recording.

## **Decoding**

The decoders job is to produce loudspeaker signals that create a good illusion of the required directional sound field. [5] The Ambisonics format can be rendered on any speaker layout using the suitable decoder. [3]

## **The Virtual Ambisonics Approach**

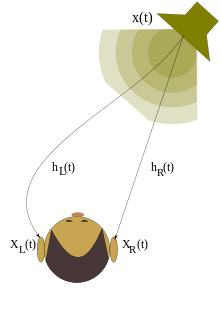
To transform the sound-field into the binaural audio, there is need to decode it first on a virtual array of speakers and then further applying HRTFs on each mono output of the speaker to generate a binaural signal from each speaker which can further be superimposed to get the final headphone output. [6] [7] [10] [16]

## **Binaural Rendering**

Binaural rendering is converting the output of speakers to headphone output (binaural left and right) by applying Head Related Transfer Functions (HRTFs) [6]

## **HRTF and HRIR**

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 modifies the original sound before it enters the ear. These modifications can be portrayed with the help of the HRIR’s, 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 virtual surround sound. The HRTF is the Fourier transform of HRIR. [12]



Head Related Transfer Function

**Figure 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>

HRTFs for left and right ear describe the filtering of a sound source (x(t)) before it is perceived at the left and right ears as xL(t) and xR(t), respectively as illustrated in figure 2.1. How an ear perceives a sound from a point in space is characterized by a Head related Transform function. To synthesize a binaural 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. [9] [12] [13]

# **RELATED WORK**

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 doesn’t 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 doesn’t 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, developed in the early 1970s by Peter Fellget [3] and Michael Gerzon [4] is a way of recording and reproducing surround sound in both horizontal and vertical surround, 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.

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, it is not limited to any number of channels, the more the number of channels the higher is the directional resolution [5]. The technique is based on a precise and unambiguous 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]

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 vector and the three main axes.

Equation (1) Fi = ½ \* [G1\*W + G2\*(X \* cos(α) + Y \* cos(β) + Z \* cos(γ)]

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:

Equation (2) B = C \* P

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:

Equation (3) P = pinv(C) \* B

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(θ)\*cos(φ), sin(θ)\*cos(φ), sin(φ)] 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]

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]

# **SYSETEM ARCHITECTURE DIAGRAM**

Figure 2 illustrates the system architecture diagram of the proposed system. Various blocks of the system are explained below.

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.

1. **Rendering Output to Virtual Speaker Array**

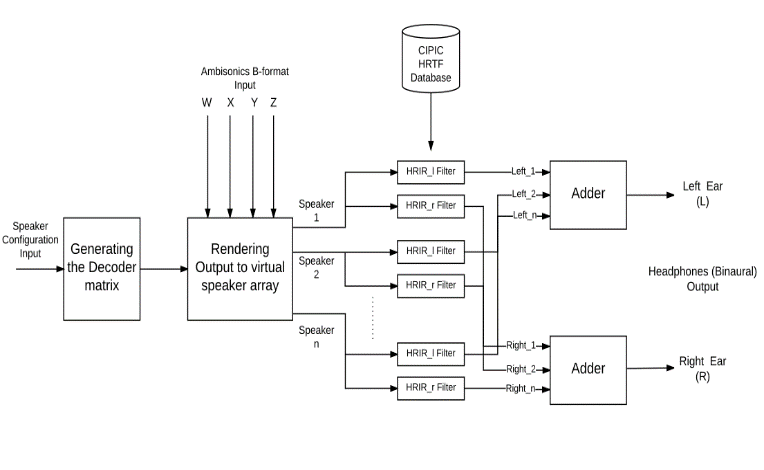
This function will take the 4 channels (W, X, Y, and Z) and the decoder matrix as inputs and generate a mono output for each speaker of the speaker array.

1. **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.

**4.4 Adder**

This is a simple adder which will superimpose all the outputs from each HRIR-l and HRIR-r filters and generate a single left and right final binaural audio.



# **CONCLUSION**

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. Furthermore, this can be extended to consider the head tracking coordinates from the VR devices and improve localization by applying rotations on the sound-field audio.

**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, <http://www.pspatialaudio.com/Ambisonics%202014%20article.pdf>.
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.



**Anupam Godse** is a Final Year Bachelor of Information Technology student at College of Engineering, Pune, India. He is also a member of communication subsystem in CSAT, an undergraduate satellite initiative at College of Engineering, Pune, India which developed SWAYAM, a passively stabilized pico-satellite, which was successfully launched into space by ISRO on 22 June 2016. He is currently doing an internship at NVIDIA Pune.



**Vinod Pachghare** is working as Associate Professor at College of Engineering, Pune (An autonomous Institute of Government of Maharashtra). Recently he was honored with Best Faculty of the year Award by CSI Mumbai chapter in TextNext 2018 at IIT Bombay.

He has published 03 books: Computer Graphics, Cryptography and Information Security, and Cloud Computing. He has to his credit over 30 technical published papers in National & International Journals and Conferences. He has worked as Chief Investigator for two major projects, "Wireless Intrusion Detection System" funded by AICTE, New Delhi and "Information Security Education and Awareness (ISEA)" project phase-II funded by Department of Information Technology, Government of India. He has worked as Principal Investigator for the R and D project " Intrusion Detection System", funded by Institution of Engineers (India). He has coordinated number of workshops, symposiums and faculty development.

He was worked as a member of Board of Studies (BOS), Computer Engineering, SPPU, Pune. He is a member of Board of studies for many autonomous Institutes in Maharashtra. He is also working as Member Advisory Board (Computer Engineering) for many colleges in Maharashtra. He has worked as a reviewer for books, conferences as well journals. He is a member of Computer Society of India (CSI), IEEE, IETE, ISTE, IAENG, IACSIT, Computer Science Teachers Association (CSTA). As a member of professional societies, he worked as session chair, reviewer, PC member, Board of studies members, advisory committee members in various Institutes.



**Kaustubh Ashtekar** is an engineer and has been working in the software industry for the last 11 years. His professional experience has mostly been in the field of IoT, developing low level and system software and for various types of embedded systems. He currently works for Nvidia Pune.