Binaural Rendering Of Ambisonics - A 360 Degree Surround Sound Technology

Anupam Godse
Mentor: Prof. Vinod K. Pachghare (College of Engineering, Pune), Mr. Kaustubh Ashtekar (NVIDIA Graphics Pvt. Ltd.)

College of Engineering, Pune

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.

PROBLEM STATEMENT

Design and implement 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).

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.

Recording and Encoding:
Recording is done with the help of special sound field microphone. It can be encoded and stored in various formats out of which B-format is most widely used format.

Decoding:
The decoder's primary job is to produce loudspeaker signals that create a good illusion of the required directional sound field.

Binaural Rendering:
Binaural rendering is process of producing a 2 channel headphone output (left and right).

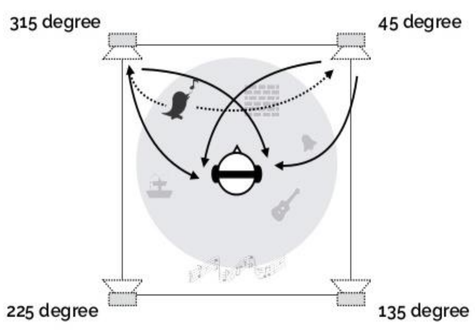
RELATED WORK

Disadvantages of Traditional approaches:
Traditional approaches had channel information stored for each speaker and hence had fundamental drawbacks:

1. Predefined layout and number of speakers.
2. Not suitable for mobile applications.
3. Not suitable for real time VR applications.
4. Jumping audio problem.

While traditional technique of surround sound had its limitations and disadvantages, Ambisonics, which was developed by Peter Fellget and Michael Gerzon 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.

Figure 1 Virtual ambisonics approach



RELATED WORK (cont…)

Virtual Ambisonics Approach:
Ambisonics require minimum 4 channels so it is impossible to render it over 2 channels for binaural output. Markus Noisternig, Thomas Musil, Alois Sontacchi and Robert Holdrich provided solution by this by introducing a virtual ambisonics approach explained in figure 1.
The approach solves this problem by processing input in following steps:

1. Render input over virtual array of speakers.
2. Apply Head related transform function to mono output of each speaker to generate binaural output.
3. Superimpose all the left and right signals into a single left and right output (binaural output).

SYSTEM DESIGN

The system is designed according to the virtual ambisonics approach. Various blocks of the system are explained below.

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.

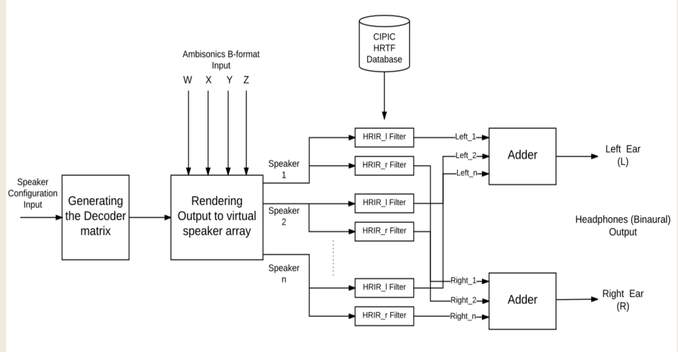
Rendering Output to Virtual Speaker Array:
This function will take the 4 channels (W, X, Y, and Z) and the decoder matrix as inputs and generate a mono output for each speaker of the speaker array.

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.

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.

Figure 2. illustrates the system architecture diagram of the system.

Figure 2 System Architecture Diagram



EXPERIMENTATION AND RESULTS

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.

Following are the results obtained:

Test file: An Ambisonics file with 4 channels shown in figure 2 and,
Output: Binaural 2 channel output shown in figure 3.

Figure 3 Ambisonics 4 channel input file

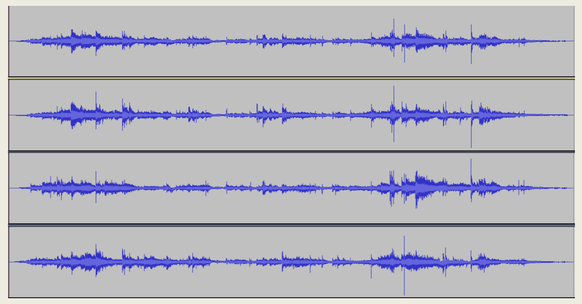
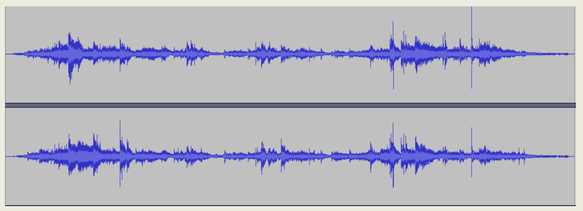


Figure 4 Binaural output file



CONCLUSIONS AND FUTURE WORK

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.

Future Work:
The system can be extended to incorporate head-tracking to improve localization effects and to make it suitable for real-time VR applications.