Sound Localisation Performance in Humans under Ambisonic and Binaural Systems

Author: Thomas Jesson

Supervisor: Lee Davison

Introduction

This investigation sought to understand whether Virtual Reality and competitive eSports, alongside their need for resolution and movement, can provide a path to market for the niche format that is Ambisonics. By performing a series of specific tests on a series of human volunteers the speed and accuracy of each individual will be calculated when experiencing a Binaural soundscape derived from a game engine panning system compared to an Ambisonic soundscape.

Background

If a sound source is located off axis from 0˚ (centre front) a time difference will be experience by the ears of the listener based on its angle of incidence. The brain can then determine the location of the signal by identifying which ear received it first and calculating the additional distance the signal has to travel to the more distant ear ***(See figure 1).*** This is known as the Interaural Time Difference (ITD) and by using this method humans are capable of resolving direction down to just a few degrees of resolution.

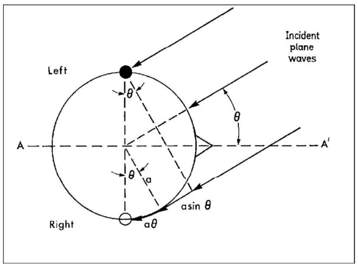


Figure 1: In this model the ITD is given by α(θ + sin θ)/c (where c = 340m/s for the speed of sound, and θ is in radians.)

Similarly, when considering intensity, the farthest ear lands in the heads 'sound shadow', which instigates what's known as the Interaural Level Difference (ILD). The combinations of these effects create what is known as a Head Related Transfer Function or HRTF; which is unique for every sound source and its angle of incidence, position and elevation.

It is easy to believe that the unique HRTF of each source, that act much like a fingerprint, would be enough to ensure the listener hears them correctly with Binaural reproduction, but there is a list of primary problems that can interfere with achieving the accuracy required. Such as:

- **The difference in listener head/ear size and shape (though common features are present).**
- **Head movements that we use to alleviate directional confusion is difficult to incorporate.**
- **Visual cues can have a strong effect on perception and are often missing when reproducing binaural.**
- **Headphone EQ and methods of mounting can affect perception.**
- **Errors in signal chains can cause phase and frequency response issues.**

Overall, the most reliable direction cues will always come from the HRTF of any individual listener but generalised functions can be used to create HRTF filters that may be used just at the cost of absolute accuracy in the reproductive process.

The Ambisonic system itself aims to provide a complete hierarchical approach to directional recording, reproduction, transmission and storage which can also be equally down sampled to mono, stereo, surround systems or even full 'periphonic' reproduction, which also includes height information. B-Format uses four signals that represent the velocity and pressure components of a sound field in any given direction [*shown in **Figure 2***]. There is a comparability with the sum and difference format of standard two channel stereo, this is due to B-Format being made up of an omnidirectional component (W) and three orthogonal figure-of-eight components (X,Y and Z).

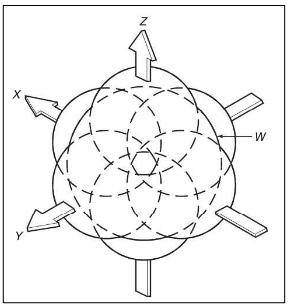


Figure 2 : W,X,Y and Z components of B-Format, representing three orthogonal velocity (figure-eight) components and the omnidirectional pressure component of a sound field respectively.

Methodology

Initially this investigation aimed to utilise the spatial audio lab at Solent University which consists of a HTC Vive VR Headset and a 25 channel Ambisonic speaker array that accepts audio encoded in third order. The crux of this would have involved connecting an audio engine like Wwise to Unreal Engine which would track and apply the intended localisation test signal to moving objects in a Virtual Reality game space. The audio engine would then need to be linked to the Digital Audio Workstation 'Reaper' , using virtual cables, which would then decode the azimuth and angles of the objects and signals to the speaker array using the ambiX (B-Format) v0.2.8 Ambisonic plug in.

This test would have then been repeated but the spatial audio technique would have been switched to the internal Binaural panning system built into Unreal Engine and its generalised HRTF filters, which would have been experienced over the VR Headset itself. Unfortunately, before this proposition could be investigated, the spatial audio lab at Solent University became unavailable due to global events.

A simplified localisation test was performed instead.

This would involve two Binaural mixes, reproduced over headphones, one being decoded from a Reaper plug-in that encodes the azimuth and elevation of the specified test signal to third order Ambisonics. The other would take place within Unreal Engine itself; with the azimuth and elevation of the test signal placed within the 3D space being encoded by the internal panning system, then decoded Binaurally.

Using a Tone Generator in Reaper ten 1,000Hz/1kHz tones were generated at 0dB in an uncompressed .wav file to guarantee fidelity and to ensure distortion wouldn't be present.

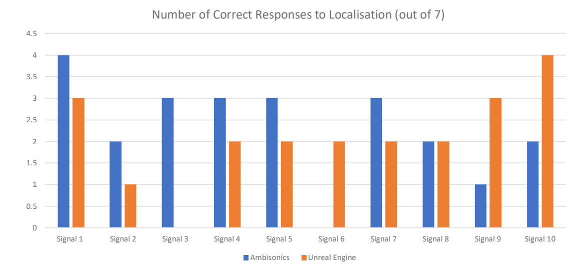
Each signal occupies its own track in Reaper (set up to encode 16 channel Third Order Ambisonics) with IEM's 'Coordinate Converter' FX plug-in being applied initially. This enables the input of the intended azimuth and elevation parameters with which the sound source would be positioned. The degrees are then converted to Cartesian Coordinates (X, Y and Z) by the plug-in.

Figure 3: IEM's Co-ordinate Converter plug-in

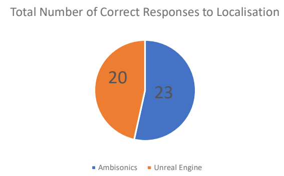
IEM's 'Room Encoder' is then used to position the test signal using the X,Y, Z coordinates which are placed into the 'Source Position' section of the plug-in. The same principles are then applied within Unreal Engine so the localisation performance of each can be compared as fairly as possible under the constraints of this investigation.

7 participants were presented with an azimuth and elevation graph and four options representing an azimuth and elevation parameter to choose from for each of the test signals. One of these options is the exact location the test signal is coming from and the other three consist of incorrect parameters.

Results



Third Order Ambisonics received more correct answers and has displayed higher levels of certainty amongst the panel when determining azimuth angles in particular. Situations involving height and direct angles are where the format encounters shortcomings. With Unreals panning system maintaining good levels of accuracy across the board, just with less instances of absolute precision with certain signals.



Conclusion

If the gaming industry continues to move towards Virtual Reality becoming the go to platform to experience interactive visual media, Ambisonics has presented an argument that its ability to provide improved levels of accuracy to a 3D space can only help to enhance realism and immersivity. Although the panning systems found in today's game engines can clearly still hold their own against a system with unmatched potential in spatial resolution. Therefore, it's believed that the experiences found in more common ways to play games (ie story driven role-playing games on pc and consoles) will continue to rely on these more traditional spatial systems, as they are still able to display high levels of accuracy.

References: F. Rumsey, T. McCormick. 2014./M.Gerzon, 1977./A. Roginska, P.Geluso.2017/G.Almeida, A.Ribas, J.Calleros. 2019/J.Blauret.1997