WHAM: WEBCAM HEAD-TRACKED AMBISONICS

M Dring College of Science and Engineering, University of Derby, Derby, UK
Dr B Wiggins College of Science and Engineering, University of Derby, Derby, UK

1 INTRODUCTION

This paper describes the development and implementation of a real-time head tracked auralisation platform using Higher Order Ambisonics (HOA) decoded binaurally based on open-source and freely available web technologies without the need for specialist head tracking hardware. An example implementation of this work can be found at: https://brucewiggins.co.uk/WHAM/.

To provide an immersive experience of 3D sound via binaural reproduction using headphones, it is widely acknowledged that the use of head tracking is beneficial to maintain the same auditory cues used in daily life (providing dynamic ITD, ILD and pinna cues). Listeners are shown to experience improved externalisation, source localisation perception within audio scenes and to minimise the number of front-back confusions[1,2]. The implementation of head tracked binaural representations are vastly simplified using Ambisonics, based on the spherical harmonic decomposition of a sound field at a single point[3]. YouTube, Facebook and Oculus (for example) have used this combination of formats (Ambisonics decoded binaurally) for 360 video and VR applications due to this flexibility. Although, the system is not without limitation, with spatial aliasing occurring above a set frequency which, for a human head size radius of correct reconstruction can be approximated to ~600Hz for 1st order Ambisonics, 1200Hz for 2nd order, 1800Hz for 3rd order, and so on. Lower orders do, however, offer the benefit of lower channel counts and good computational efficiency which is particularly important for implementation on mobile and lower power devices; the sweet spot of performance and efficiency currently being 3rd order Ambisonics. Various techniques to improve transparency at low orders whilst maintaining this computational efficiency are being actively researched[4] with time alignment of binaural filters at high frequencies (>1kHz), diffuse field equalisation and the assumption of a symmetrical head model (which halves the number of convolutions needed)[5] being good examples of these optimisations. Current implementations aim for correct and stable identification of source direction and flat frequency response for dry sources as the goal for these optimisations with 3rd order microphones and tools now becoming more readily available in order to both capture and synthesise auditory scenes. However, when auditioning more complex scenes including room response, work from Dring and Wiggins[6] has shown that improvements in spatial reproduction are apparent up to a much higher order. For applications such as in the auralisation of real or synthetically simulated rooms where accuracy, rather than efficiency is the priority, the use of higher orders than are possible using currently available tools and equipment is desirable.

Current 360 video/VR/AR implementation of Ambisonics to binaural are based on the assumption that anechoic head related transfer function (HRTF) data is used in the fixed viewpoint reproduction of the limited order Ambisonic B Format material that will contain both the audio sources and associated ‘place’ or room response (i.e. the entire sound scene). In order to implement head tracking, it is the B Format sound scene that is rotated which, due to the HRTF data used in the binaural conversion being anechoic, is equivalent to rotating the head. It is also the use of anechoic HRTF data that allows the assumption of a symmetrical head to not be too detrimental to the reproduction quality especially as the mismatch between the listeners and systems HRTF are likely a greater compromise. This limits the reproduction to the Ambisonic order of the captured/synthesised audio scene and/or available microphone technology. Previous work by the authors[6] has demonstrated how binaural scanning of a room (modelled in EASE) can be used to derive up to 35th

Ambisonic order reproduction of a single sound source in that space with equally high resolution room response. Crucially, if it is the head that is rotated, rather than a change of source position when scanning the room, accurate, very high order, head tracking can also be provided. Test subjects noticed improvements/differences up to around 20th order reproduction with good externalisation. Due to the asymmetrical nature of the captured room responses, the symmetrical HRTF assumption cannot be used when reproducing the sound field using this method, which is an important consideration.

After the system detailed in[6] was implemented and tested using an EASE room simulation to generate the BRIRs and Reaper combined with the MrHeadTracker head tracker for playback, the next stage in the project was to retest using real room data. To this end, a classroom at the Markeaton Street Campus of the University of Derby was binaurally scanned with a sound source at 45 degrees to the left of the listener providing all the data needed for the listening tests. Soon after this occurred, the COVID-19 lockdown happened and neither the University facilities nor equipment were available for use. This encouraged the development of a system using web-based technologies to allow for a more open and accessible platform which enables playback with head tracking without the need for specialist head tracking hardware. The system would need to be able to process the long, high order, Ambisonic to binaural filters in real-time and implement head-tracking using non-specialist hardware/software with the outcomes of this work being particularly well suited to the acoustics community providing higher quality auralisation of real or simulated rooms than is currently available.

2 METHOD

2.1 VHOA Ambisonics

Previous work has indicated that high order Ambisonics as the basis for a head tracked binaural representation of an acoustic scene can provide transparent reproduction above 20th order[6], with other studies showing that directional response can be adequately achieved above fifth order[7]. For an excellent and concise look at the process of using spherical harmonics (SH) and Ambisonics in the creation of binaural filters, the work of Politis & Poirer-Quinot[3] provides an excellent summary.

This project utilises only 2D, circular harmonics, which reduces the number of channels needed to implement the system, reduces the amount of binaural room impulse responses captured, but only allows rotation around the Z axis (rotation of the head) in terms of head tracking. Usually, HRTFs or BRIRs are captured at many points around the dummy head, which allows a source to be positioned at all points in the space. Using standard SH/Ambisonic transforms[3] head-rotation is then implemented by rotating the sound scene which re-pans or rotates the sources in that scene. If the head related transfer functions captured are anechoic, then rotating the sources and rotating the head in the sound scene are, in fact, equivalent. However, if the binaural filters contain room responses (i.e. they are BRIRs) then rotating the B format signals decoded in this way will actually reposition the sources in that room keeping the head static. This will lead to conflicting auditory and sensory cues and has implications for the auralisation of spaces using high order Ambisonics (higher order than there are Ambisonic microphones or software tools currently available).

A binaural decode using spherical harmonics/Ambisonics can be considered to be a transformation of the standard loudspeaker approach to Ambisonics, but with the speaker outputs replaced with HRTF pairs. This was well documented by McKeag and McGrath using 1st order Ambisonic recordings to feed head tracked binaural audio over headphones in 1996[8] with higher order spherical harmonic representation of HRTFs (up to 17th order) analysed by Evans et al. in 1998[9].

In general terms, the equation for a real-valued spherical harmonic, Y, for order n and degree m is shown below in equation (1):

$$Y_n^m(\theta,\phi) = N_n^{|m|} P_n^{|m|}\big(sin(\phi)\big)\begin{cases} cos(|m|\theta) & if\ m > 0 \\ sin(|m|\theta) & if\ m < 0 \\ 1 & if\ m = 0 \end{cases} \quad (1)$$

Where N is a normalization coefficient
P is the associated Legendre polynomial
θ and φ indicate the azimuth and elevation angles
n and m are the SH order and degree

For horizontal only systems (where $|m| = n$ and $\phi = 0$) the Legendre polynomial will evaluate to the value 1 with the 0th order harmonic represented by one signal and each subsequent order represented by a further two signals as shown in equation (2) below (using normalization scheme, maxN, where each harmonic is normalized to have a maximum value of 1):

$$\begin{aligned} Y_0^0 &= 1 \\ Y_1^1 &= \cos(\theta) \\ Y_1^{-1} &= \sin(\theta) \\ Y_2^2 &= \cos(2\theta) \\ Y_2^{-2} &= \sin(2\theta) \\ &\dots \\ Y_n^n &= \cos(n\theta) \\ Y_n^{-n} &= \sin(n\theta) \end{aligned} \quad (2)$$

The Ambisonic encoding (B) of a source (S) can be given by:

$$\begin{aligned} \mathrm{B} &= Y_n^m(\theta) \cdot S \\ \mathrm{B} &= [\mathrm{Y}_0^0(\theta), \mathrm{Y}_1^{-1}(\theta), \mathrm{Y}_1^1(\theta), \dots \mathrm{Y}_n^m(\theta)]^T \mathrm{S} \end{aligned} \quad (3)$$

These signals (known collectively as B Format) can be reproduced using a linear combination of plane waves from a number of discrete points/loudspeakers as shown in equation (4) where C is the encoding matrix and are the values of the spherical harmonics for the loudspeaker or, in this case, HRTF measurement positions and p is a column vector of loudspeaker (or measurement position) gains. L represents the total number of points/loudspeakers.

$$\begin{aligned} \mathbf{B} &= \sum_{i=1}^{L} \mathrm{Y}_n^m(\theta_i) \cdot g_i \\ \mathbf{B} &= \mathbf{C} \times \mathbf{p} \end{aligned} \quad (4)$$

So, for a first order horizontal system (3 circular harmonics) feeding four loudspeakers/positions, C would be a 3 x 4 matrix and p would be a 3 x 1 column vector of, currently unknown gains.

$$\mathrm{C} = \begin{bmatrix} Y_0^0(\theta_1) & \cdots & Y_0^0(\theta_4) \\ Y_1^1(\theta_1) & \ddots & Y_1^1(\theta_4) \\ Y_1^{-1}(\theta_1) & \cdots & Y_1^{-1}(\theta_4) \end{bmatrix} \quad (5)$$

In order to calculate the gains needed for each of the B-format input signals (which are the source, multiplied by the circular harmonic coefficients at the desired angle), the column vector p must be obtained, which will involve finding the inverse of C as shown in equation (6). It is this inverse matrix that becomes the decoding coefficients needed to produce the loudspeaker signals in order to reproduce the original signal, spatially, over the desired loudspeaker array. As C is rarely a square matrix, the Moore-Penrose pseudo inverse allows the required inversion to be realised.

$$\begin{aligned} \boldsymbol{B} &= \boldsymbol{C} \times \boldsymbol{p} \\ \boldsymbol{p} &= \boldsymbol{D} \times \boldsymbol{B} \\ &where\ \boldsymbol{D} = \boldsymbol{C}^+ \end{aligned} \quad (6)$$

Filter pairs for binaural reproduction can be calculated by weighting the binaural response pairs for each position with the corresponding decoding coefficient and summing, resulting in a pair of BRIRs for each B format channel (so, if we had a simple 1st order system with three circular harmonics, we would end up with 3 pairs of BRIR, one pair per channel). These can be convolved with each channel of an encoded source (as shown in equation (3)) to reproduce over headphones. The first three and last two channels Ambisonic BRIRs for an Ambisonic order N are shown in equation (7).

$$\begin{aligned}
\boldsymbol{AmbBRIR_0^0} &= \sum_{i=1}^{L} \boldsymbol{BRIR_i} \cdot \mathbf{D_{i,1}} \\
\boldsymbol{AmbBRIR_1^1} &= \sum_{i=1}^{L} \boldsymbol{BRIR_i} \cdot \mathbf{D_{i,2}} \\
\boldsymbol{AmbBRIR_1^{-1}} &= \sum_{i=1}^{L} \boldsymbol{BRIR_i} \cdot \mathbf{D_{i,3}} \\
&\dots \\
\boldsymbol{AmbBRIR_N^1} &= \sum_{i=1}^{L} \boldsymbol{BRIR_i} \cdot \mathbf{D_{i,2N}} \\
\boldsymbol{AmbBRIR_N^{-1}} &= \sum_{i=1}^{L} \boldsymbol{BRIR_i} \cdot \mathbf{D_{i,2N+1}}
\end{aligned} \tag{7}$$

where L is the number of source positions measured
N is the Ambisonic Order

One important point regarding this system is that the source will sound from the position of where the original source is located during measurement. Panning, or positioning a virtual source using an Ambisonic panner using the derived filters will actually be controlling the rotation of the head rather than the source position in the space. For this reason, it does not make sense to build up an entire scene of sources panned to different positions, but for the purposes of auralisation of a space, one source is able to give an impression of the space.

The process of capturing and reproducing using very high Ambisonic order is summarized as follows:

1. Capture 72 binaural room impulse responses (every 5 degrees) with a static source and rotating the head.
2. Calculate the pseudo-inverse of the circular harmonic gains for those head rotations (C^+ as shown in equation (6)).
3. Weight and sum the BRIRs for each rotation with the gains for that order's coefficient resulting in 2N+1 pairs of filters (where N is the Ambisonic order) as shown in equation (7).
4. Convolve these filters, per order, with a single Ambisonically panned source (B in equation (3)) making sure the panning angle matches the head rotation of the listener.

The system implementation is shown in block diagram form in Figure 1.

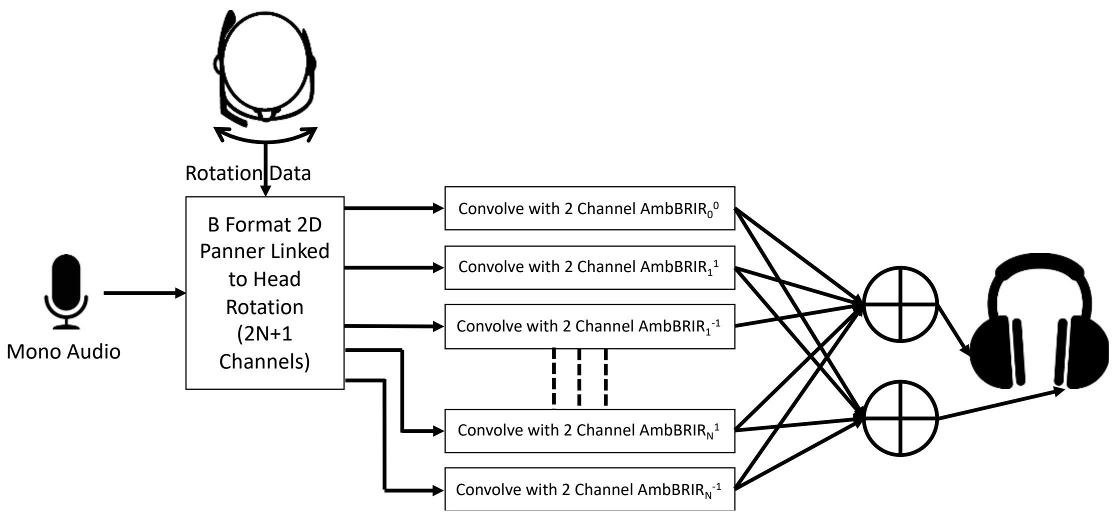


Figure 1: Block Diagram of the Webcam Head-tracked AMbisonics processing

2.2 Automated BRIR Capture

The natural development from previous work[9], which tested using simulated data, was to capture the binaural room impulse responses (BRIRs) of our test room (MS015 – University of Derby) to generate sufficient circular harmonics for 35th order rendering by rotating a head and torso simulator (HATS) every 5 degrees. These captures were initially intended for subjective testing to determine the parity between the auralisation of real and modelled spaces and if this had an effect on the test results regarding which order of Ambisonics could be used before the listener no longer perceived a difference. For this reason, the BRIRs represent a sound source (Dynaudio BM5 Mark 3) at 45 degrees off axis at a distance of 2.87m relative to the forward-facing listener seat position. To make the BRIR capture process efficient and with consideration of capturing rooms on a mass scale, a Python script (https://bitbucket.org/DrWig/wigware-reaper-scripts/src/master/WigET250-3D_Turntable.py) was written to automate the turntable rotation (ET250-3D) hosted in the Reaper DAW software and synchronised to the project timeline (see Figure 2 for an example of the scripts output during measurement). Using this, the user can instruct the script to rotate the turntable *x* degrees every *y* seconds and automatically stop the test after *z* seconds. The Reaper project contains 72 log sine sweeps, with the script forcing the 5 degree rotation at user defined time intervals for a user defined total duration, whilst capturing the signal return from the dummy head; the user need only to initiate the Reaper project to carry out the full test run automatically.

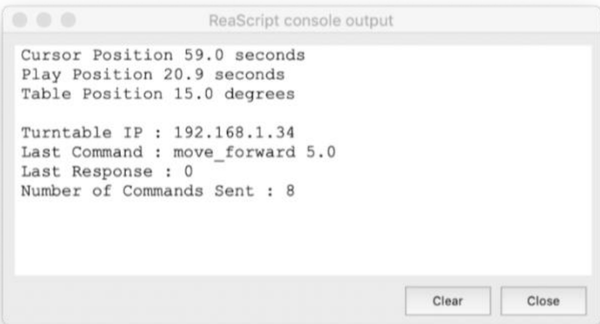


Figure 2: ReaScript Output Window While Running Turntable Automation Script

2.3 Face Tracking

Typical head tracking methods include the use of dedicated devices that attach to the listeners headphone band (e.g. Waves Nx Head Tracker[10]) or employ the use of an infrared camera and reflective markers positioned on headwear (e.g. TrackIR[11]), offering 6 degrees of freedom, making them useful for audio production and gaming applications. Low cost, DIY alternatives are also available and have been used by the authors to good effect in previous studies[9]. Studies have shown that head-trackers should demonstrate low-latency and angular resolution of 2 degrees to present stable binaural synthesis[12]. However, the draw-back in each of these head-tracking solutions is the requirement of a physical device which requires a financial outlay and if necessary, the competency to implement a build. Any resource created needed to include binaural listening with head movement to align closely to previous experiments conducted by the authors for continuity in research, therefore the solution to bypass a physical device was to implement webcam head tracking to provide essential positional data.

For the project to be realised the use of face-tracking software that could be embedded into the authors web page was required. Research led to Beyond Reality Face (BRF)[13], a state of the art browser based software development kit (SDK) that is capable of detecting faces in the webcam stream and tracking face movement using 68 facial landmarks. Figure 3 shows an example of the face markers captured via a webcam. Although this solution is not tracking an exact centre of the head, the high number of points relay significant data to follow the position of the head over a wide range of movements in front of the camera. The authors were further encouraged as this specific platform has been used in a marketing campaigns by a major brand to provide an augmented reality experience[14].

Figure 3: Webcam capture with Beyond Reality Face 68-point (blue dots) Face Tracking Overlay

The website calls the "BRFv4DemoMinimalWebcam.js" script which in turn activates the camera, determines the camera resolution and sets out the facial tracking points before returning the image to a user defined canvas area on the website. As is common in on screen previews (e.g. selfies) the image displayed is mirrored.

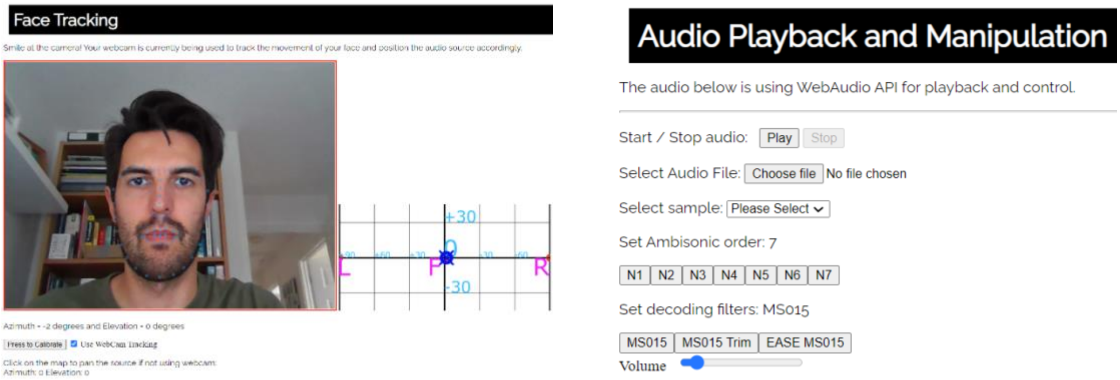
Whilst the computer's webcam might be fixed into position or mounted consistently nearby, the relative position of the listener cannot be guaranteed during even the shortest of sessions. It was deemed essential to implement a calibration button, whereby the vertical and horizontal position determined by the JavaScript are offset to 0 when the calibrate button is pressed, allowing the listener unlimited opportunities to reposition themselves during use of the resource.

2.4 JSAmbisonics – Resolving Asymmetry

In order for the real-time Ambisonics to binaural processing to occur in a browser, a number of third party library options are available such as JS Ambisonics[3] and Omnitone[16]. These libraries utilise the webaudio API[15] which allows for powerful real-time processing of the audio to happen using just a the browser. The choice of library to base the project on came down to the fact that the JS Ambisonics library allowed for the use of custom HRIR (or, as in our case, longer BRIR) data in the form of a JavaScript Object Notation (JSON) file created using a Spatially Oriented Format for Acoustics (SOFA) file. SOFA files can be created using the appropriate library for Matlab. The webaudio API is stated to support 'at least' 32 channels, which means most browsers support 32 channels (the minimum necessary to conform to the standard). JSAmbisonics currently supports up to 3rd order 3D Ambisonics (16 channels) and also up to 15th order 2D Ambisonics (2N+1 = 31 channels). Once a custom SOFA file has been loaded, the library will calculate the Ambisonic filters needed to implement the decoding (i.e. the process detailed in equations (5), (6) and (7)). However, the library assumes symmetrical filtering, where the head used for the anechoic Head Related Impulse Responses is assumed left/right symmetric. This, as mentioned, is not too detrimental an assumption for anechoic auralisation, but if BRIRs are captured, symmetrical filters would assume that the room response captured is left/right symmetric, which is never the case. For this reason, the JSAmbisonic library has been forked and updated to allow for asymmetrical filters to be used along with updated Ambisonic filter calculation and convolution code. This does mean that the channel count, in terms of the convolutions needed to implement the filtering, is doubled (2 per Ambisonic channel) which, due to the 32 channel limit in the WebAudioAPI, means 7th order Ambisonics is currently the implemented practical limit. Further development work should be able to overcome this issue in the future by processing left and right responses in separate streams. This forked version of JSAmbisonics can be found on Github at https://github.com/DrWig/JSAmbisonics.

3 THE WEBSITE

The fundamental purpose of the website is to provide a resource to audition VHOA BRIRs that can be accessed and controlled using resources common to all computers, eliminating the need for additional proprietary devices. The website currently has two sections, “Face Tracking” and “Audio Playback and Manipulation”, as shown in Figure 4.



(a) (b)

Figure 4: WHAM Website (a) Face Tracking Functions (b) Audio Playback and Manipulation Functions

Listeners do not have to enable the face-tracking functionality for auditioning to continue, disabling this feature reverts any desired rotation to movement of the puck located on the map to the right of the video window; making this usable even without presence of a webcam. Enabling the tracking feature provides the listener with the intended full experience of the site. Once active, the listener observes the puck movement follows the movement of the face captured by the webcam until an angle is reached where the 68 facial points are no longer determined by the BRF SDK. Whilst the puck will show movement in the vertical axis as both azimuth and elevation data are returned, the dynamic changes in the BRIRs respond only to horizontal changes in puck position; remembering height information will still be returned as the captured room responses are full 3D.

Changes to the auditioned audio can be made by the listener in the Audio Playback and Manipulation section of the website. Two audio samples (speech and drums) are provided by the site from the original JSAmbisonics library, however, listeners can also load audio files located on their own personal drives; with anechoic mono captures being recommended. The 3 BRIRs available to pass the audio files through are all associated to the room mentioned in section 2.2. Whilst listeners are currently limited to this selection of filters, they provide clear audible variations, highlighting the need for continued investigation by the authors started in their previous work[6].

Whilst the captured BRIR sets are capable of head tracking/rotation around the Z axis with up to 35th order spatial accuracy, the website is limited to orders of 1 through to 7 currently. At this stage, the importance to maintain asymmetry in the BRIRs has taken precedence over maximum order, with the limitation being imposed by the channel count capabilities of browsers’ implementation of the web audio API. Once rendered, listeners will receive an audio source positioned at 45 degrees counter-clockwise from straight ahead.

4 OBSERVATIONS

In its present form the website has achieved its primary purpose. The functions enable the users to experience binaurally reproduced 3D audio of VHOA BRIRs with head tracking implemented via a webcam stream. Having the webcam tracking provides the critical element for effective subjective listening, without the necessity of a physical head tracking device.

Unlike a physical head tracking device with 360 degree rotation, the webcam capture has a limited field of view (i.e. front facing). Although the user is at liberty to angle their head at a position of their choosing, with the most extreme being 180 degrees; observations show the BRF tracking points reach an approximate maximum of 40 degrees to the left or right before failure to track. The success of head tracking via the webcam is intrinsically linked to the success of the facial landmarks being recognised by the BRF SDK. To this end, good lighting and avoiding obstruction of the face are important. Positioning a light source in front of the face was recognised as enabling greater rotation angles to be achieved. On occasion, facial hair also caused problems with the accurate positioning of the facial landmarks. However, it has been noted that 40 degrees of head rotation does enable much of the natural movement that would be attempted whilst still viewing with the computer display.

As data is being passed from the webcam through the scripts embedded on the webpage a latency is present. Most physical head tracking devices seek a sub 30ms latency to maintain stable binaural synthesis. To determine a known consistent latency with this resource is not possible due to the variation in end user's system performance. However, using a Python script[17] to measure webcam only latency using visual methods, values of around 128ms at 30fps are reported (with 30fps equating to a time of 33.33ms between frames and the minimum measurable and timing resolution available using this visual method of latency measurement). Although the authors did not find the latency to be an issue whilst passing the webcam stream through the website, it exceeds the ideal value stated for physical devices.

5 CONCLUSION

Although this resource was never a planned part of the larger project, as in other aspects of life in 2020, adaptations to common practices once carried out with close social contact have been switched to more remote processes. The aim of this work, which was to implement the foundations for a test system using only a web browser, have been successful, and the software outcomes will be useful for others in acoustics who want to auralise spaces to high spatial accuracy with head tracking without the need for specialist head tracking hardware. The project is still in progress (see further work below), but is at a stage where others can make use of the outcomes. The 3rd party libraries and work that fed into this resource were invaluable with the changes necessary for asymmetric processing shared back to the community to enable others to benefit from the work.

In summary, a system able to deliver dynamically rotated auralisation of an acoustic space using 7th order BRIRs with asymmetrical filtering using only a web browser and integrated webcam, eliminating the need for additional specialist hardware has been successfully developed with the code needed to reproduce this made freely available.

5.1 Further Work

The site which is currently accessible is a development version where the concept and fundamental functionalities have been achieved. To continue to align with other active projects the site will need to function with higher order filters; ideally utilising a minimum of 20th order diffuse field equalised filters[6]. Maximising on the concept of remote interaction with listeners will also require listening test functionality and data collection to be embedded into the site, with responses being returned to the server for analysis and study.

Beyond site layout and instructional information, widening the audience appeal of the resource would see the integration of an upload function for locally stored SOFA files. Users would therefore be able to audition any audio sample through any personally captured BRIRs with head tracking for immediate and effective auralisation purposes.

To make objective comparisons between the performance of the site and physical head tracking devices it would be recommended to implement a process to measure the latency of the entire system. The outcome of this could inform future development of the site and lead to subjective investigations to determine a webcam head-tracking latency perception threshold.

6 RESOURCES

- WHAM website: https://brucewiggins.co.uk/WHAM/
- Automated Sweep and Turntable Rotation Python ReaScript: https://bitbucket.org/DrWig/wigware-reaper-scripts/src/master/WigET250-3D_Turntable.py
- Forked JSAmbisonics supporting 7th Order 2D Ambisonics with Asymmetrical Filters: https://github.com/DrWig/JSAmbisonics

7 REFERENCES

1. F. L. Wightman, D. J. Kistler, Resolution of Front-Back Ambiguity in Spatial Hearing by Listener and Source Movement, The Journal of the Acoustical Society of America 105, 2841-2853 (May 1999)
2. S. Werner, G. Gotz, F. Klein, Influence of Head Tracking on the Externalization of Auditory Events at Divergence between Synthesized and Listening Room Using a Binaural Headphone System, AES 143rd Convention, New York, New York, USA, October 18-21, 2017.
3. Politis, A., Poirer-Quinot, D. JSAmbisonics: A Web Audio library for interactive spatial sound processing on the web. *Interactive Audio Systems Symposium*, York, UK, 2016
4. McCormack, L. and Politis, A., 2019, March. SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods. In *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*. Audio Engineering Society.
5. Wiggins, B. Paterson-Stephens, I., Schillebeeckx, P. (2001) The analysis of multi-channel sound reproduction algorithms using HRTF data. (2676 downloads) 19th International AES Surround Sound Convention, Germany, p. 111-123.
6. M. Dring, B. Wiggins, THE TRANSPARENCY OF BINAURAL AURALISATION USING VERY HIGH ORDER CIRCULAR HARMONICS, Reproduced Sound 2019 - Institute of Acoustics, Bristol, UK, Vol. 41. Pt. 3 2019, p. 165-173
7. Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, B. Rafaely, (2017) Spectral equalization in binaural signals represented by order-truncated spherical harmonics. J. Acoust. Soc. Am. 141(6)
8. McKeag, A., McGrath, D. (1996) Sound Field Format to Binaural Decoder with Head-Tracking. 6th Austrailian Regional Convention of the AES, Melbourne, Austrailia. 10 – 12 September. Preprint 4302.
9. Evans, M. J. (1), Angus, J. A. S. (2) and Tew, A. I. (2) (no date) ‘Analyzing head-related transfer function measurements using surface spherical harmonics’, *Journal of the Acoustical Society of America*, 104(4), pp. 2400–2411. doi: 10.1121/1.423749.
10. Waves.com (2020), *Wave Nx Head Tracker*. [online] Available at: https://www.waves.com/hardware/nx-head-tracker [Accessed 30/09/2020]
11. Naturalpoint.com (2020), TrackIR 5. [online] Available at: https://www.naturalpoint.com/trackir/products/ [Accessed 30/09/2020]

12. Romanov, M.; Berghold, P., Implementation and Evaluation of a Low-cost Head-tracker for Binaural Synthesis. *AES 142nd Convention*, Berlin, Germany, May 20-23, 2017.
13. Marcel Klammer (2020), Beyond Reality Face SDK - v5.1.5 (BRFv5) - Platform: Browser. [online] Available at: https://github.com/Tastenkunst/brfv5-browser [Accessed 15/06/2020]
14. Tastenkunst (2012), *Nike Free Face uses Beyond Reality Face*. [online video] Available at: https://www.youtube.com/watch?v=4Y-caQiIDug [Accessed 27/09/2020]
15. W3.org. (2020). *Web Audio API*. [online] Available at: <https://www.w3.org/TR/webaudio/> [Accessed 16 October 2020].
16. Google (2020). *OMNITONE*. [online] Available at: <https://googlechrome.github.io/omnitone/> [Accessed 16 October 2020].
17. Perrytsao (2018), Webcam-Latency-Measurement. [online] Available at : https://github.com/perrytsao/Webcam-Latency-Measurement [Accessed 15/10/2020]