Genelec Aural ID

DR HYUNKOOK LEE tests a system that promises quick, personalised HRTFs for 3D audio listeners

What are HRIR and HRTF?

Three-dimensional (3D) audio is rapidly becoming a new production and delivery standard for music and film content as well as for various virtual and augmented reality (VR/AR) applications. To ensure high levels of immersion and realism in VR, 360° video must be accompanied by 360° binaural audio with a head tracking technology, so that sound images are always localised at their corresponding visual positions while the user rotates the head. A direct binaural recording with a dummy head is not suitable for this purpose as the microphone has a fixed orientation.

To allow user-interaction of image positions according to the head orientation, individual sound sources or Ambisonically decoded signals need to be processed with Head-Related Impulse Responses, a set of filters that contain the aural fingerprints of different directions of arrival (DoA) of sound. As the shape of the external ear (i.e. pinna) is so complex that a sound arriving from every different direction is reflected and scattered by the pinna in a different way, giving rise to a unique frequency response the specific DoA. The frequency response of a HRIR is the so-called HRTF (Head-Related Transfer Function). While Interaural Time Difference (ITD) and Interaural Level Difference (ILD) are the main cues for horizontal auditory localisation, the HRTF plays a crucial role for accurately localising sounds arriving vertically or from the back.

Need for individual HRTF

Traditionally, HRIR datasets created using a dummy head, such as Knowles KEMAR (*Resolution* V15.4) or Neumann KU100, have been widely used for spatial audio research and binaural rendering tools. Artificial external ears used for dummy head microphones are modelled from many people's ear shapes and sizes, and therefore HRIRs obtained from such microphones are of human 'average'. They usually work reasonably well for binaurally panned sources in the horizontal plane, although they could often cause a front-back confusion especially in the median plane (e.g. 0° and 180° azimuth angles). However, the main problem with generalised HRTFs is poor accuracy in vertical localisation.

Auditory elevation perception relies on the positions and magnitudes of peaks and notches present at frequencies above around 5kHz, but the thing is that they vary considerably person-to-person because the ear size and the pinna shape are highly individual. This is one of the reasons why some people (including myself) really struggle to localise elevated sounds with generalised HRTFs. However, if HRTFs were properly measured with your own ears, there would be much less chance for the brain to be confused in resolving directional cues. In binaural rendering of 3D audio, using *individual* HRTFs could substantially improve front/back localisation accuracy as well as vertical one.

Measuring individual HRTF

However, what hinders individual HRTFs from being widely used is the tedious and complicated acquisition process as well as the lack of accessibility to the required facility. My research is concerned with spatial perception in 3D sound recording and reproduction, and I often need to measure several subjects' HRTFs in my lab for binaural listening tests.

Acquiring accurate HRIRs with a high directional resolution is always a delicate and time-consuming task. Not only that mounting and fixing a miniature microphone at the correct position at the ear canal entrance of each subject could be a tricky job, but also there are a series of post-processing tasks to be performed to obtain reliable HRIRs. As a subject, you would have to sit down still with your head position fixed for a long time while the measurement is taking place. Last time I got mine measured, it took nearly an hour! Most of all, you would need a special facility and expert skills in signal processing to be able to obtain a high-quality set of individual HRIRs. This is why they are still mainly used only in academic research.

Photogrammetry

For the above reasons, I was so delighted when Genelec announced Aural ID last year because it makes HRTF acquisition much easier and quicker — and most importantly — makes high-precision 3D audio experience more accessible for everyone. Rather than physically measuring HRIRs, Aural ID simply uses a photogrammetry technique to synthesise individual HRIRs based on the anthropometric data of your ears.

In other words, the way a sound wave from a specific DoA (Direction of Arrival) is reflected inside the pinna and by the shoulder can be predicted based on the head size, the shape of the pinna and the distance between ear and shoulder. Aural ID HRIRs are delivered in the SOFA (Spatially Oriented Format for Acoustics) format, which is an AES standard file format for storing and reading spatial audio impulse responses. All that is required from your end is to make a video recording of your head and torso using your smartphone and send it to the Genelec Aural ID team.

How does it sound then?

For this review, I subjectively compared my Aural ID HRIRs against a KEMAR dummy head HRIR dataset as well as my other set of individual HRIRs measured using the traditional method (I will call this 'physical' HRIRs for convenience). I tested them in several different scenarios.

Firstly, I wanted to examine horizontal and vertical localisation accuracies of the HRIRs using a single sound source binauralised for various target positions. I used a custom Max patch written using the APL's SOFA for Max objects (Figure x, available at https://doi.org/10.5281/zenodo.3268541) to binauralise a noise burst and some anechoic musical sources. This patch allowed me to easily move the binaural image to different target azimuth and elevation angles, and also quickly switch between the three different SOFA files.

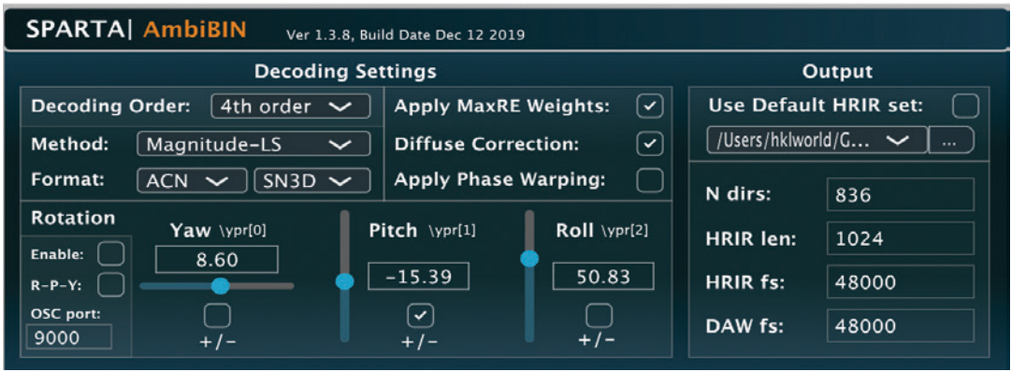
My first impression was that the horizontal imaging of Aural ID was impressively accurate — there was no perceivable spatial difference to the physical BRIRs, which I know work

accurately for my ears. But what struck me most was that there was no back to front confusion that I often experience with the KEMAR HRIRs. Head tracked binauralisation can usually help resolve this issue, but with Aural ID the front and back localisation was already clear with a static binauralisation.

After more careful listening, I noticed that this was actually related to externalisation (i.e. out-of-head localisation). The Aural ID was clearly better externalised than the KEMAR HRIRs and this made it easier for me to discern whether the sound is from the front or the back. It sounded to me that the KEMAR HRIRs had a bit of a proximity effect due to slightly excessive low frequency energy. My physical HRIRs also performed quite well in terms of localisation and externalisation in the horizontal plane, but I was impressed that computer-synthesised HRIRs could work as accurately as the physically measured ones.

Now, I tested elevation localisation at various azimuths, and this made me even more impressed as the sense of height was actually better with Aural ID than my physical HRIRs. For example, at 45° azimuth and 45° elevation, which is a typical front height loudspeaker position in 3D sound reproduction (e.g. 5.1.4), I could really hear the sound around the target position with Aural ID, whereas with the physical HRIRs it was more difficult to judge the exact position. With the KEMAR HRIRs, I always struggled to perceive elevation accurately as mentioned above. This could be of course due to the mismatch between KEMAR's and my HRTFs, but again externalisation seemed to be another possible reason.

Moving on, I tried binauralising some of my 3D microphone array recordings made in reverberant concert halls to examine the overall spatial quality of binauralised multichannel recordings. The recordings were a 360° choral performance at the Chapter House within York Minster (for 7.1.4) and an orchestra concert at the Victoria Hall in Geneva (for 5.1.4). They were recorded using my PCMA-3D microphone array, which consisted of 7 or 5 main and 4 height layer microphones. The height layer of the array is mainly to pick up reflections and reverberation from above, whereas the main layer is for source imaging and rear ambience.



/ Figure 2: SPARTA AmbiBIN plugin used for binaural decoding of higher-order Ambisonics
Courtesy of Leo McCormack, Aalto University

Aural ID vs. KEMAR

Comparing between Aural ID and the KEMAR HRIRs, it was once again obvious that Aural ID created a better sense of externalisation. For the choral recording, with the KEMAR, the singers in the front and back almost sounded as if they were from the front, and the sense of height was weak for the ambience from the height channels. With the Aural ID, on the other hand, I could separately hear all singers without any front-to-back or back-to-front confusion, and the vertical spread of the sound image was more apparent. For the orchestral recordings, it was pleasing to be able to hear reverberation coming from the back of the concert hall so clearly with Aural ID. Again, due to the inherent front-back confusion issue of generalised HRTFs, binaural recording made using a dummy head often sounds quite "flat" rather than "deep".

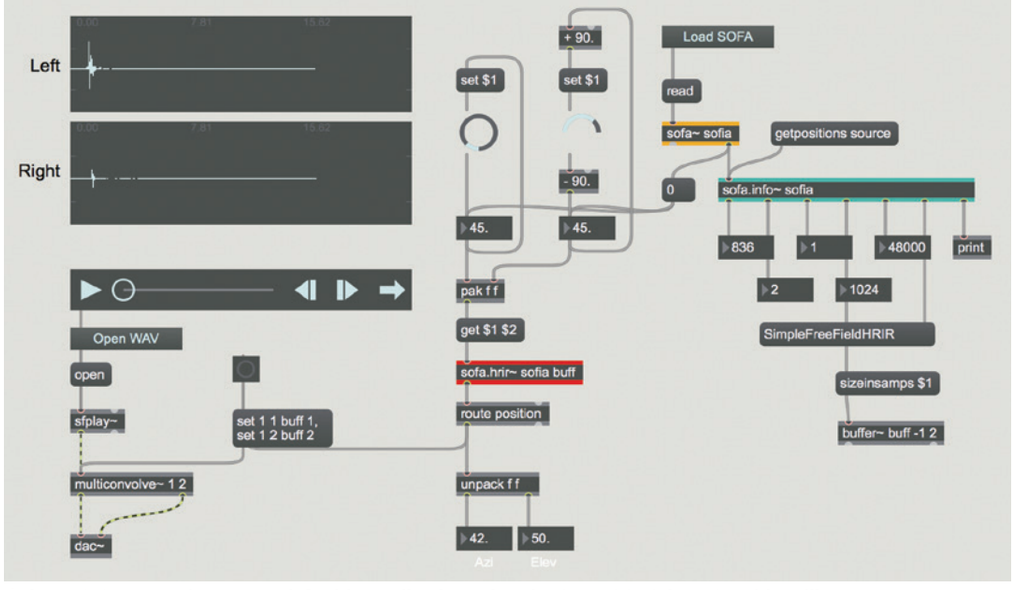
Additionally, I wanted to check how the HRIRs would perform with Ambisonic recordings made using an mhAcoustics Eigenmike spherical microphone array, which support Ambisonic rendering with the orders of 1 of 4 as well as beamforming. A number of 3D recordings made using the Eigenmike and various microphone arrays are freely available in the 3D-MARCo library (https://doi.org/10.5281/zenodo.3474285).

I tested the string quartet, piano trio and organ excerpts from the library. For binaural decoding, I used the Aalto University's SPARTA AmbiBIN plugins (Figure x) as it supports the SOFA format. The difference between the Aural ID and KEMAR HRIRs was very subtle for string quartet or piano trio recordings, especially at the 3rd and 4th orders. However, with a pipe organ recording, there was an apparently better sense of height with Aural ID whilst KEMAR gave a slightly more low-end energy. The organ, often called the king of instruments, is a perfect type of sound source that can demonstrate the benefit of 3D recording, and it is crucial to represent the physical height of the instrument in recording. This requires an effective rendering of vertical image spread by frequency distribution, and given my experience with Aural ID I believe individual HRTFs could really help achieve that goal in binaural reproduction.

To conclude

From these informal comparisons, I am convinced of the benefits that Aural ID HRTFs can provide in binaural 3D audio monitoring and listening. Cognitive load is substantially reduced especially when trying to localise sounds from elevated positions or the back. Externalisation and overall listening experience are also enhanced compared to generalised HRTFs. Individual HRTFs used to be mainly for researchers working on human auditory perception, but now it is more accessible to anyone who is keen on good 3D audio experience. I also think the quick and easy acquisition process of Aural ID can even make the researchers' jobs much easier too.

Dr Lee is director of the Applied Psychoacoustics Lab (APL), University of Huddersfield.



/ Figure 1: An Example Max patch for binauralisation using the APL's SOFA for MAX object library
Courtesy of Dr Dale Johnson, the University of Huddersfield