24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

ICSV24
23-27JULY 2017
LONDON CALLING

HEAD-TRACKED AURALISATIONS FOR A DYNAMIC AUDIO EXPERIENCE IN VIRTUAL REALITY SCENERIES

Eric Ballestero
London South Bank University, Faculty of Engineering, Science & Built Environment, London, UK
email: ballese2@lsbu.ac.uk

Philip Robinson
Independent Researcher, Seattle, WA, USA

Stephen Dance
London South Bank University, Faculty of Engineering, Science & Built Environment, London, UK

This paper aims to take advantage of the new cutting-edge virtual reality technologies – such as head-mounted displays for virtual reality and ambisonics – in order to recreate 3D immersive environments; both aural and visual. The work presented here is believed to encourage investigations into buildings yet to be, or those lost to civilisation. Through a combination of acoustic computer modelling, network protocol, game design and signal processing, this paper proposes a method for bridging acoustic simulations and interactive technologies, i.e. fostering a dynamic acoustic experience for virtual scenes via VR-oriented auralisations.

Keywords: Auralisations | Ambisonics | 3D head-tracking | Max/MSP | Oculus Rift

1. Introduction

The ability to create computer models, based on real or imaginary environments, has been evolving at an extraordinarily fast pace for the past decades. Computer Aided Design (CAD) software and new user-machine interfaces have played a major role in developing easier and faster means to build virtual environments with increasing realism. Alongside the latter, computer processing capabilities for undertaking acoustic calculations have been rapidly expanding, allowing faster and more efficient deterministic calculations of the emulated physics' behaviour inside computer models; be it with either Geometrical Acoustics (GA) or Numerical Acoustics (NA) approaches.

Even if ground-breaking improvements are being made in wave-based computations (FEM, BEM, FDTD) allowing quicker and more efficient simulations [1] [2], the present context in acoustic computer modelling is still widely under the influence of the geometrical – optic based – approximation of sound propagation.

The GA method is indeed used in most companies as a means to calculate and predict the acoustics of many spaces; already existing or yet to be. Despite the theoretical and practical limitations induced by such approach of sound behaviour, GA simulations – if correctly built [3] – can still provide sufficient acoustic data for approximating real world scenarios. Nowadays, the technique being commonly used to prospectively share the hypothetical auditory sensation of some acoustic space is called auralisation, i.e. the process of rendering audio data through binaural synthesis by digital means to achieve a virtual reconstruction of the sound field at a given position.

The auralisation process can be achieved in many ways, either manually through home-made signal processing or, more commonly, by using built-in functions already integrated in most acoustic computer modelling software (eg. CATT-Acoustic, ODEON). However, one of the underlying restrictions with respect to most of these applications is the lack of dynamism they provide (eg. head movement tracking). It is well known that one of the ways we naturally judge the acoustic quality of a space is by slightly moving our head around, recording the relative changes in intensity and time of arrival of sound between our two ears [4]. As the major aim for auralising a space is to be immersively propelled into a virtual environment; mimicking reality at best so our senses can be fooled; it seems therefore logical and natural to account for such behaviour in our auralisation processes.

2. A Dynamic Audio Experience

This paper highlights the need for a dynamic audio experience to support the immersion provided by virtual environments; this method being already integrated for visual purposes (eg. virtual reality (VR) headsets). To fulfil such need, a head-tracked auralisation system was created during an MSc project framework, using virtual game features along with GA predictions and audio signal processing. Throughout this process, a virtual 3D sound field can be recreated by decoding a pre-calculated B-format impulse response into an ambisonic sound reproduction configuration – for a given position inside the model. The latter can then be virtually synthesised for a binaural listening experience through the use of generic HRTFs.

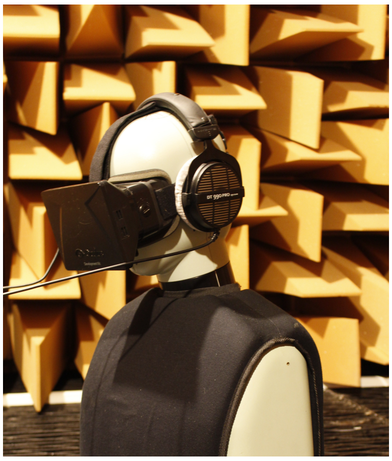


Figure 1: Dummy head equipped with a VR headset display and headphones

Physically speaking, rotational head-tracking of the subject is being supported by the gyroscopic sensor mounted inside a VR headset, presently an Oculus Rift DK1, where the visual information and rotational data can be set in any game engine; Unity3D in our case. Whilst visual information is being directly rotated in function of the user's head movements, gyroscopic data of the user's head is to be sent indirectly via UDP communication to a signal processing software (Max/MSP), which is used to rotate the B-format representation of the recorded sound field in function of such input data.

At the end of this procedure, the listener is being given a three dimensional representation of a sound field with the ability to rotate his head around multiple axes, respectively changing the visual display as well as the binaural information.

These improved features for standard auralisations – first person visual and aural experiences of a space with physical feedback – could be strongly used in prospective architectural design, being a cheap alternative for full sound surround rooms, or a subject to more detailed sound design investigations in the video game industry. With the recent rise of VR technologies, static auralisations will progressively become obsolete, henceforth the need to upgrade this audio technique to a new level.

3. GA computer modelling

In order to implement the aforementioned procedure, it is first required to record the B-format impulse response at a particular position inside a virtual environment. Commercial computer modelling software such as CATT-Acoustic or ODEON possess both user-friendly built-in functions allowing this kind of sound field recording.

To illustrate the underlying design procedure for dynamic auralisations, any kind of virtual space can be modelled. For this paper, an example of classical acoustics will be taken, i.e. the Roman theatre of Arles, as used for the MSc project – the original theatre being in ruins for several centuries.

The reason for choosing such subject was to provide a new type of virtual immersion for archaeological sites, an alternative to visual-only reconstructions of old monuments. Roman theatres supposedly being venues where acoustics were important, this made it an interesting investigation subject for implementing a dynamic audio approach.

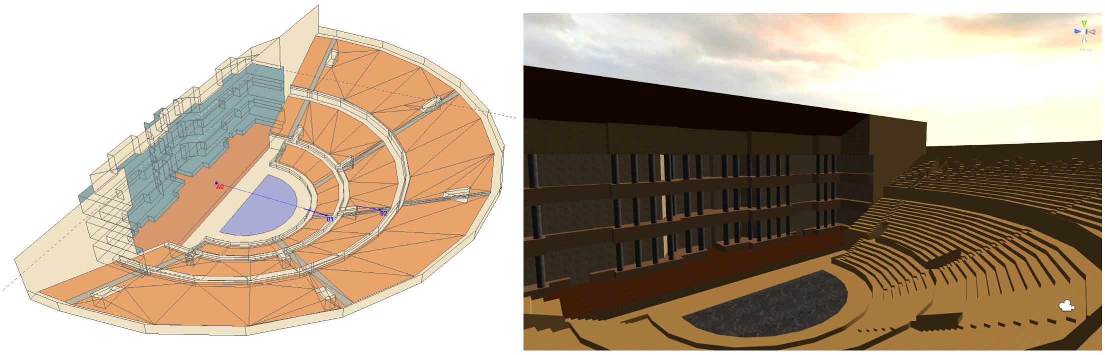


Figure 2: Virtual reconstruction of the Roman theatre of Arles. (Left): CATT-Acoustic computer model (empty audience); (Right): Visual rendering of the theatre in Unity3D.

The computer model of the Roman theatre was created through SketchUp and imported into CATT-Acoustic for GA calculations. Geometric meshes and visualisations are shown in Figure 2.

The Roman theatre was optimally built for a GA approach of sound behaviour by following various guidances on either general computer modelling using CATT-Acoustic [3], or by particular case studies on acoustic computer modelling of other Greek and/or Roman theatres [5] that were conducted during the ERATO project.

Based on this guidances, the smallest surface of the model was chosen to be of 1 m, narrowing the uncertainty of predictions from 1 kHz and above ($\lambda \gg d$, d being the smallest represented surface); frequencies lower than 1 kHz might still give plausible results despite the absence of wave-based phenomena, but with higher uncertainties. Acoustic absorption coefficients for stone, the main material in contact with air when the theatre is empty, have been taken from measurements made by the ODEON team in the Roman theatre of Aspendos, Turkey, during the ERATO project. Appropriate scattering coefficients of non-smooth surfaces were estimated in function of the surface irregularities. In order to stick to a simple model configuration, a single monopole sound source was incorporated in the middle of the stage, while two receivers were placed in the central axis of the theatre at the rear of the first and second tiers of the audience (*ima cavea* and *media cavea*). B-format impulse responses can therefore be recorded at both receiver locations once GA predictions are successfully run through.

This model configuration is believed to reduce the amount of uncertainty when running the acoustic calculations for a frequency range of 500 Hz - 16 kHz; the overall high reflection coefficients of the theatre materials being more suitable for a deterministic tracing, and the large width of the space ($\approx$ 100 m) reducing any modal behaviour, thus increasing the diffuseness required for stochastic predictions – in the limits of an open-air model.

4. Unity 3D & Oculus Rift features

Unity3D is a game engine used for many video game applications, featuring a wide panel of interactive tools. Mostly programable, Unity3D works under a C# language environment. Integration of Oculus Rift features within Unity3D is made easy by the use of Oculus/Unity integration packages and Software Development Toolkits (SDK). The integration package provides all the required tools in order to create a virtual stereoscopic camera following the angular rotations of the Oculus Rift VR headset. The hardware and software connection chain is illustrated in Figure 3.

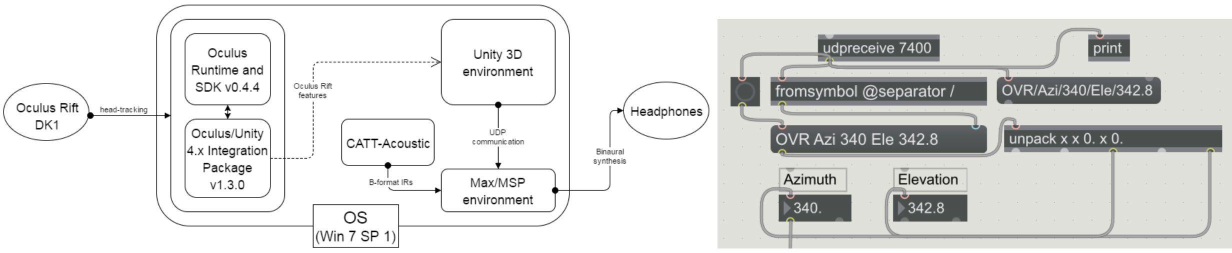


Figure 3: (Left): Hardware and software connection chain. (Right): gyroscopic data in Max/MSP

Thanks to the widely programable environment provided by Unity3D, it is possible to write a C# script for taking in real-time the gyroscopic angular values (Azimuth and Elevation) of the VR headset and send them via UDP communication to any broadcasting port inside a local or external network. Such feature is achieved with the help of three main scripts; one for the Input/Output (I/O) connection protocol, another used for building a library of OSC (Open Sound Control) functions in order to encode the string array of angular values, and last but not least, a third script in charge in calling all I/O settings, Euler angular values of the VR headset for Azimuth and Elevation changes, and finally encoding the information into an OSC message ready to be sent to a local broadcasting port via UDP communication.

At the end of the communication chain, UDP data is currently received through the signal processing software Max/MSP and unpacked in order to isolate each rotational value in separate numerical floating boxes, as illustrated in Figure 3. This real-time information of head-related rotational values represents the corner stone from which ambisonic sound fields can be thereafter rotated in function the user's head movements.

5. Ambisonics and Binaural Synthesis

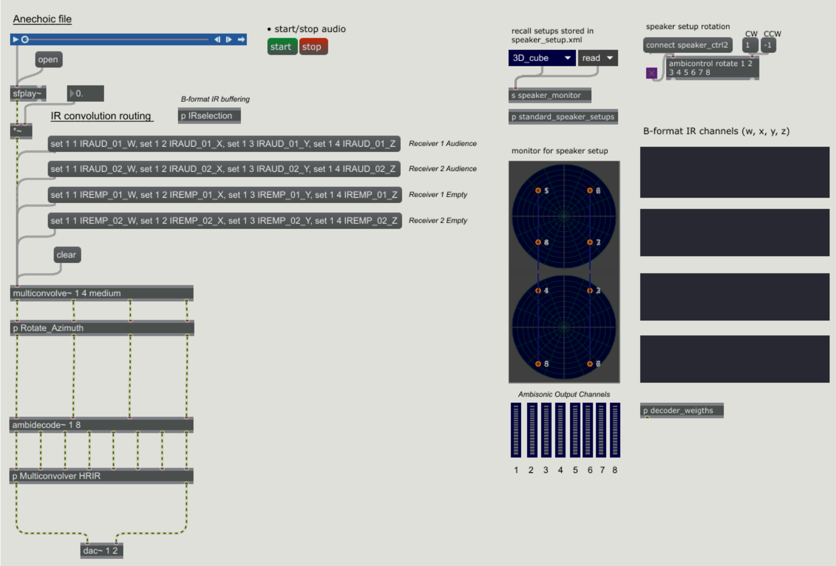

Anechoic file
open
sfplay~
0.
start/stop audio
start
stop
B-format IR buffering
IR convolution routing
p IRselection
*~
set 1 1 IRAUD_01_W, set 1 2 IRAUD_01_X, set 1 3 IRAUD_01_Y, set 1 4 IRAUD_01_Z
Receiver 1 Audience
set 1 1 IRAUD_02_W, set 1 2 IRAUD_02_X, set 1 3 IRAUD_02_Y, set 1 4 IRAUD_02_Z
Receiver 2 Audience
set 1 1 IREMP_01_W, set 1 2 IREMP_01_X, set 1 3 IREMP_01_Y, set 1 4 IREMP_01_Z
Receiver 1 Empty
set 1 1 IREMP_02_W, set 1 2 IREMP_02_X, set 1 3 IREMP_02_Y, set 1 4 IREMP_02_Z
Receiver 2 Empty
clear
multiconvolve~ 1 4 medium
p Rotate_Azimuth
ambidecode~ 1 8
p Multiconvolver HRIR
dac~ 1 2
recall setups stored in speaker_setup.xml
3D_cube
read
s speaker_monitor
p standard_speaker_setups
monitor for speaker setup
5
8
8
3
4
2
8
6
Ambisonic Output Channels
1 2 3 4 5 6 7 8
speaker setup rotation
CW
CCW
connect speaker_ctrl2
1
-1
ambicontrol rotate 1 2 3 4 5 6 7 8
B-format IR channels (w, x, y, z)
p decoder_weigths

Figure 4: Max/MSP main patch

Figure 4 presents the main Max/MSP audio signal processing patch used for this project. The four channels (W, X, Y and Z) of the first order B-format IRs obtained in CATT-Acoustic were separated into four standalone audio files; providing an easier audio buffering in order to recall the appropriate B-format channel.

Once the buffering of the B-format IRs is completed, the latter are to be sent to a multichannel convolution tool, created by the Music Department of the University of Huddersfield (`http://www.thehiss.org`), which will separately convolve each B-format channel with an anechoic audio signal. This operation results in four B-format channels conveying audio information along with the acoustic signature of the model for a given listening position.

The step responsible for rotating the sound field in function the user's head movements is set right before decoding the B-format channels into D-format loudspeaker feeds. In Figure 4, this can be seen by the integration of a sub-patch called '`p_Rotate_Azimuth`' – Azimuth dynamism being mostly of interest in this paper. This sub-patch presently records the incoming gyroscopic values from Unity3D and applies an input/output rotational matrix to the B-format signal feeds. Such rotational matrix is illustrated in Table 1. A similar matrix also exists for Elevation movement.

Table 1: Z axis rotational matrix (Inputs/Outputs) for first order ambisonics, where (a) is the incident angle of sound.

	W In	X In	Y In	Z In
W Out	1	0	0	0
X Out	0	$cos(a)$	$-sin(a)$	0
Y Out	0	$sin(a)$	$cos(a)$	0
Z out	0	0	0	1

The decoding step of B-format sound field signals into D-format loudspeaker feeds was made possible thanks to the ICST team from Zurich University of the Arts, who coded multiple tools for ambisonic patching into a Max/MSP environment [6]. This step can be recognised in Figure 4, where the four B-format signals entering the ‘`Ambidecode∼ 1-8`’ filter are being converted into eight loudspeaker feeds. These signals are connected to eight virtual loudspeakers in a full sphere periphonic configuration – a 3D cube around the listener with a loudspeaker in each corner. The listening position is thus set equidistantly from all loudspeakers.

The ultimate step required for auralising the reproduced sound field is to proceed to a binaural synthesis of the sounds generated by every loudspeaker. The use of Head-Related Transfer Functions (HRTFs) is therefore an imperative. For this project, HRTF data of a ‘standard human subject’ was downloaded from IRCAM’s database (`http://recherche.ircam.fr/equipes/salles/listen/system_protocol.html`). Then, a selection of eight HRIRs was made so each HRIR could match every loudspeaker’s angular position. Each loudspeaker feed is henceforth convolved with the left and right HRIRs of the corresponding angular incidence (eg. the [45°, 45°] speaker must be convolved with the HRIRs of the same angular position). This results in the summation of all signals arriving at the left and right ears, thus achieving a binaural synthesis of the reproduced sound field. The related Max/MSP sub-patching responsible for such binaural synthesis is illustrated in Figure 5 below.

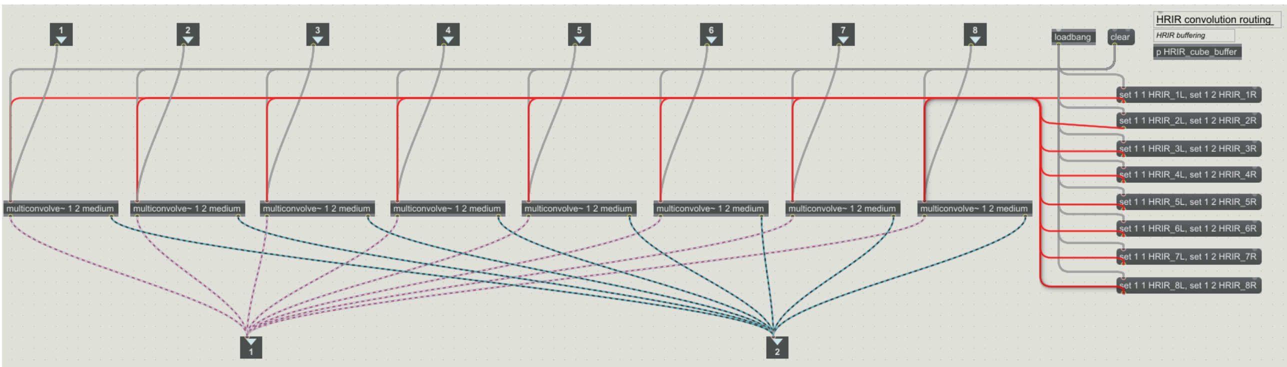

1
2
3
4
5
6
7
8
loadbang
clear
HRIR convolution routing
HRIR buffering
p HRIR_cube_buffer
set 1 1 HRIR_1L, set 1 2 HRIR_1R
set 1 1 HRIR_2L, set 1 2 HRIR_2R
set 1 1 HRIR_3L, set 1 2 HRIR_3R
set 1 1 HRIR_4L, set 1 2 HRIR_4R
set 1 1 HRIR_5L, set 1 2 HRIR_5R
set 1 1 HRIR_6L, set 1 2 HRIR_6R
set 1 1 HRIR_7L, set 1 2 HRIR_7R
set 1 1 HRIR_8L, set 1 2 HRIR_8R
multiconvolve~ 1 2 medium
multiconvolve~ 1 2 medium
multiconvolve~ 1 2 medium
multiconvolve~ 1 2 medium
multiconvolve~ 1 2 medium
multiconvolve~ 1 2 medium
multiconvolve~ 1 2 medium
multiconvolve~ 1 2 medium
1
2

Figure 5: Max/MSP HRIR (Head-Related Impulse Response) patching used for binaural synthesis

6. Measurements and Analysis

In order to verify the dynamic change of localisation cues in the Azimuth plane, a binaural measurement was set up so to collect the Interaural Level Differences (ILDs) and Interaural Time Differences (ITDs) for specific head rotations. ILD measurements were conducted using the audio signal acquisition software ARTA, whereas ITDs were obtained by recording in Max/MSP the audio files being played. Details concerning the measurement chain and the equipment used are illustrated in Figure 6. Measurement results for ILDs and ITDs are shown in Figure 7 (next page).

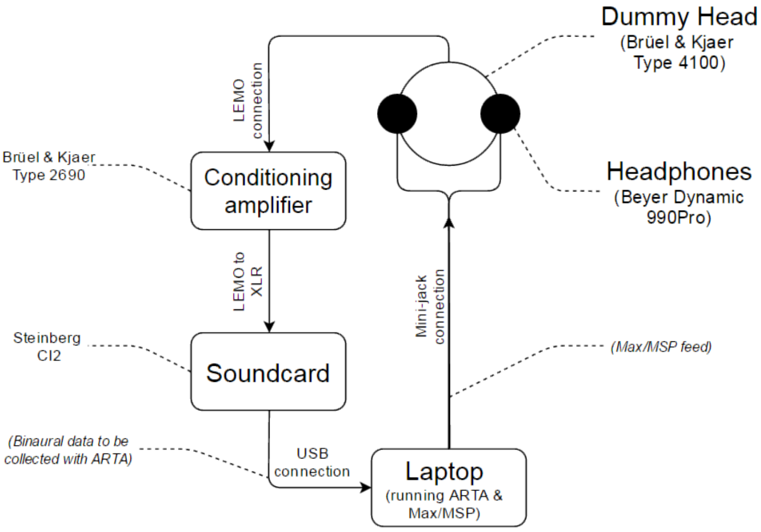


Figure 6: Binaural measurement flow chart

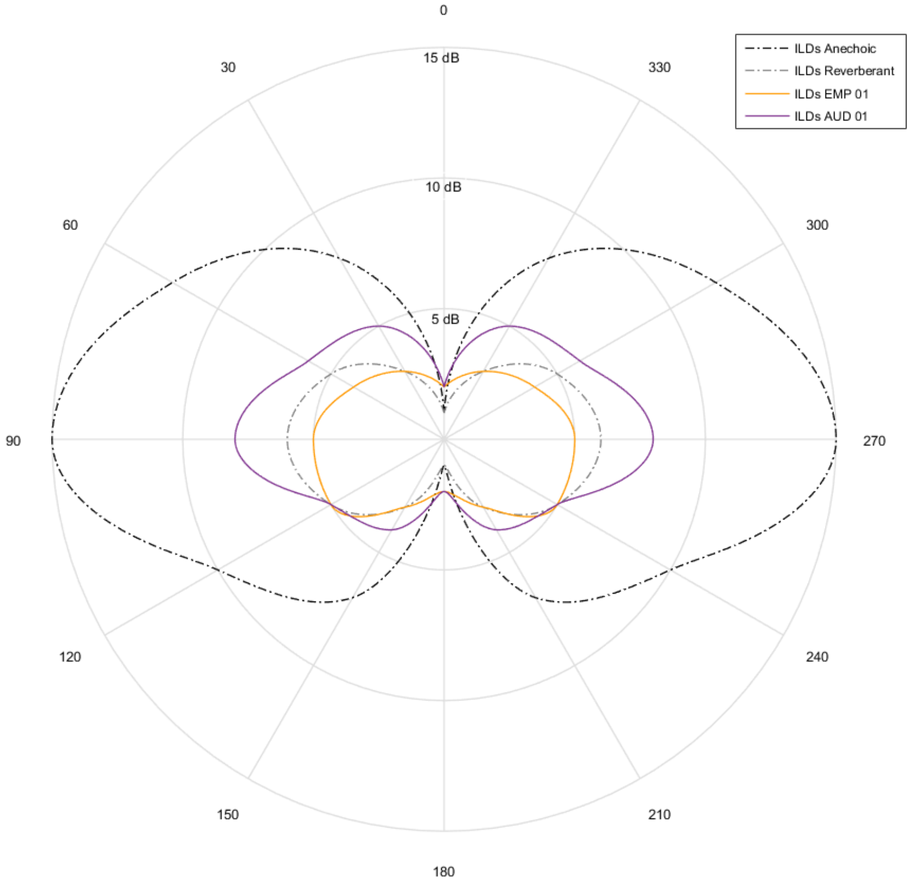
On the one hand, ITDs seemed to be concordant to values reported in literature, independently of the receiver location and model used (*with audience* or *empty*); an expected behaviour as ITDs cannot be lowered or increased, i.e. being affected by the model's acoustic characteristics. On the other hand, ILDs show a more dynamic behaviour dependently of which model or position is being played. It demonstrates that the built-up of sound within the model reduces the level differences happening between the two ears – hence giving more level dynamism in 'dead' environments that in 'live' ones. This particular aspect plays a major role in determining the sensation of reverberance in the space.

Generally, measurements allowed to verify the presence of HRTF characteristics within the reproduced audio. The dynamic tracking of head movements therefore enables the recreation of a binaural and dynamic listening environment, resulting in a better spatial resolution for the listener, as well as giving physical feedback for more immersion.

7. Conclusion

Through the use of various tools, each one related to a specific sector of activity (eg. video game engines, virtual reality headsets, computational acoustics and signal processing), it has been possible to build a dynamic auralisation process which accounts for head movements and thus reproduces the binaural changes usually experienced by humans, hence leading to a natural aural approach in virtual reality sceneries.

Considering further improvements and development in VR design – mostly focusing in virtual acoustic features – the latter technique would allow particular connections to be created between prospective acoustic design and other fields, such as game design, architecture or even archaeology.



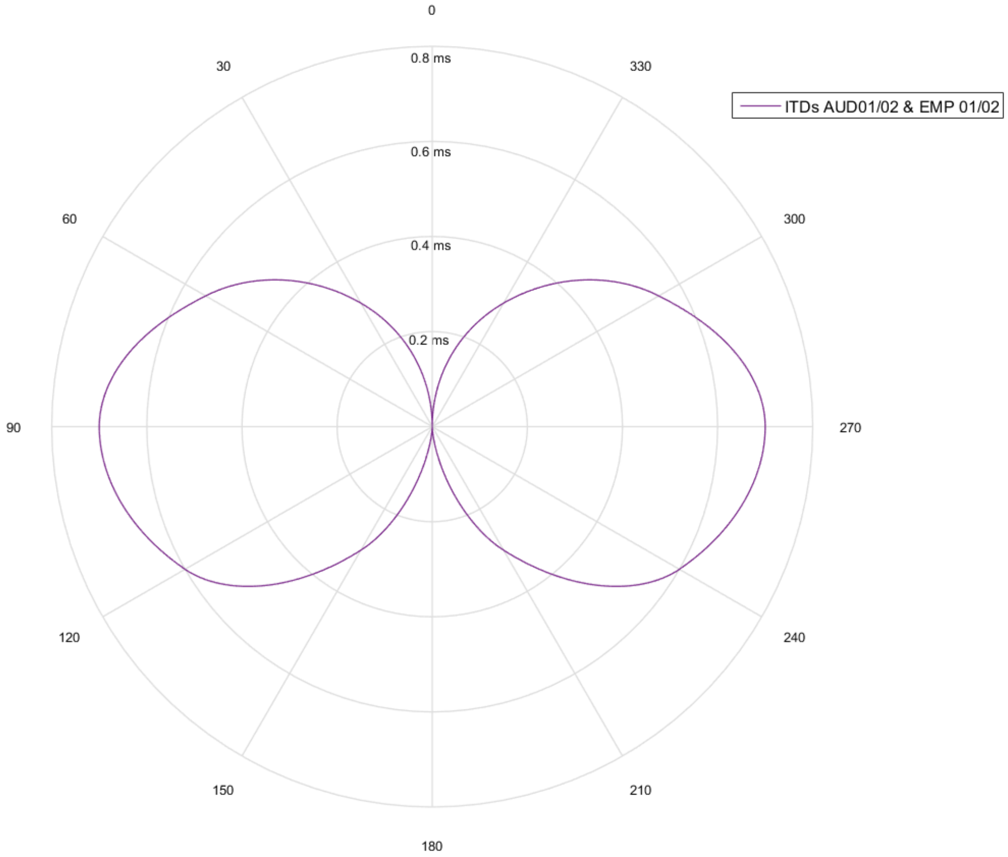


Figure 7: (Top): ILDs for the closest receiver in both '*with audience*' or '*empty*' model configurations. Anechoic and reverberant curves were obtained by measuring the ILDs in both environments with a speaker rotating around the dummy head – this gives an end-to-end range of plausible values. (Down): ITDs for every receiver location and model configuration.

REFERENCES

1. R. Mehra, N. Raghuvanshi, L. Savioja, M. C. Lin, D. Manocha, *An Efficient GPU-based Time Domain Solver for the Acoustic Wave Equation*, Applied Acoustics 73, 83-94, (2012).

2. N. Raghuvanshi, R. Narain, M. C. Lin, *Efficient and Accurate Sound Propagation Using Adaptive Rectangular Decomposition*, IEEE Transcriptions on Visual Computer Graphics, 2009;15(5):789-801, (2009).

3. B. -I. Dalenbäck, *Engineering Principles and Techniques in Room Acoustics Prediction*, BNAM, (2010).

4. M. Vorländer, *Auralization - Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, RWTH Aachen, First edition, (2008).

5. M. Lisa, J. H. Rindel, A. C. Gade and C. Lynge, *Roman Theatre Acoustics; Comparison of Acoustic Measurement and Simulation Results from the Aspendos Theatre, Turkey*, (2004).

6. J. C. Schacher, P. Kocher , *Ambisonics Spatialization Tools for Max/MSP*, ICST Institute for Computer Music and Sound Technology, Zurich School of Music, Drama and Dance, (2003).