UNIVERSITY OF DERBY, DEPARTMENT OF ELECTRONICS, MATHEMATICS & COMPUTING

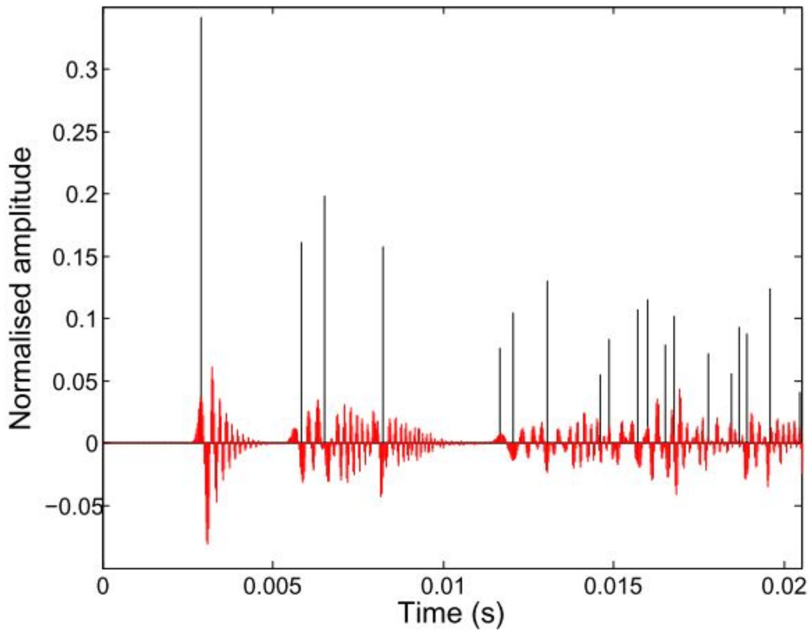


Figure 0.1: An impulse response with first reflections [17]

An Experimental Method to Quantify the Impact of Reverb Processing Techniques on Immersive Spatial Audio for Virtual Reality

Simon Durbridge

November 24, 2016

Contents

List of Figures

1 INTRODUCTION

Increases in available computing power, and great strides in research and development have brought a new surge of interest to virtual reality (VR) applications. With the introduction of improved VR systems, such as Google Jump, Oculus, Vive and facebook360, using both high end computers, and mobile phone technology to provide immersive visualisation; Further steps are being taken to provide users with immersive sound environments [18]. These environments may be created in an attempt to emulate real places, or to characterise fictional places.

A significant part of how humans identify with their surroundings, includes the perception of the reverberant characteristics of the surrounding area [20]. Cues such as the timing and strength of early reflections help humans to localize sound sources, as well as how far from the nearest boundaries the perceivers are. Some VR application development platforms provide a simplified model for how reverberation behaves in an audio system, lending creedence to the topics as discussed by Begault [2], Rumsey [20], Blauert [3] and Wiggins [21].and may be an oversimplification when attempting to create an immersive audio experience that is tending towards the suspension of disbelief[1].

The aim of this report is to introduce a testing method, to allow for the evaluation of different reverb simulation methods with respect to immersive VR audio. Initially, a basic description of sound localization theory is given. To provide context for spatial audio for VR systems, ambisonics is briefly described in relation to binaural decomposition for VR as this is the currently the more popular format. Following this, some theory behind reverberation and perception will be discussed. Finally, a testing framework will be proposed, in which subjects will evaluate different VR environments and reverb algorithms.

[1]This report is not intended to explore or discuss the artistic concept of the suspension of disbelief. For a discussion on this, please refer to Holland [11]

2 Ambisonic Spatial Audio for Virtual Reality

2.0.1 Spatial Perception

Though research into spatial audio for VR may have flourished in recent past, the foundation of localization theory remains based on the concepts explored by Rayleigh in 1907 [3]. Continued research into the field as matured and evolved understanding of these concepts, and for an in-depth review of auditory localization concepts, please review Blauert [3]. Lateralization is the term given to he human capacity to binaurally (with two ears) localize sounds laterally in the plane of the ears, and is determined by interaural time differences (ITD), phase differences, interaural level differences(ILD), and the physical listening apparatus itself (pinnae, head shape, torso etc).

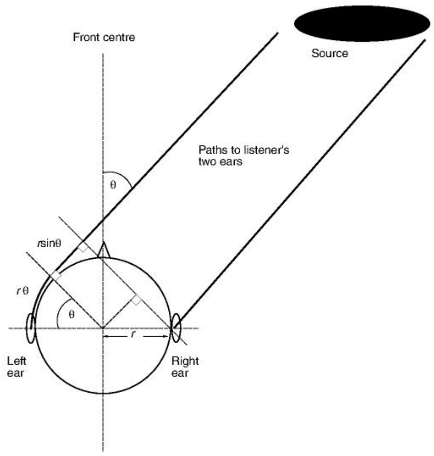


Figure 2.1: A basic model diagram of the ITD from a source to a listener given by $\frac{r(\theta+sin\theta)}{c}$ [20]

The ITD is the time difference between a sound arriving at each ear (not to be confused with phase), and is an important cue for localizing at frequencies below the wavelength relative to the size of a listeners head. The phase differences relates tothe asynchronous behaviour of sound diffracting around the head also provides temporally motivated localization cues [1]. At wavelengths larger than at least half the circumference of the head, sound waves can defract around the head. Above this frequency humans become more dependent of ILDs and phase differences.

The ILD is the level difference between ears of a sound, and is more critical to localization of shorter wavelengths where head and torso shadowing is dominant. The pinnae and torso have a distinct filtering affect that assists humans in localizing sources in the median plane [3] [2] [21], another benefit of this filtering is that humans are able to some degree overcome the 'cone of confusion' that would occur when a sound source is in a position that would otherwise produce the same ITD and ILD as another position (illustrated below). Another key cue is that humans have a tendency to move their heads, using the changes in what is heard to refine localization [3].

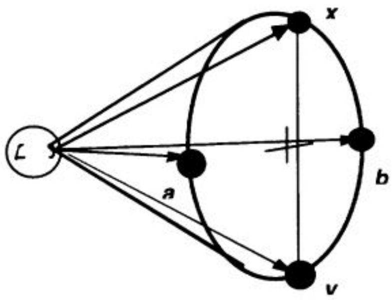


Figure 2.2: A diagram example of the cone of confusion. If a source were in location a, it would have the same ITD and ILD and thus would be ambiguous to position b. Positions x and y share a similar relationship in the vertical plane as oposed to the lateral plane of a and b [2]

The total system of localization effects from source to the entrance of the receivers ear canal(s) can be lumped into a head-related transfer function (HRTF), and can be translated into a head-related impulse response (HRIR). Upon synthesizing an HRIR filter, it is possible to reproduce audio with binaural cues over headphones or an appropriate speaker system, that gives listeners the impression of virtual source localization (and ideally externalization).

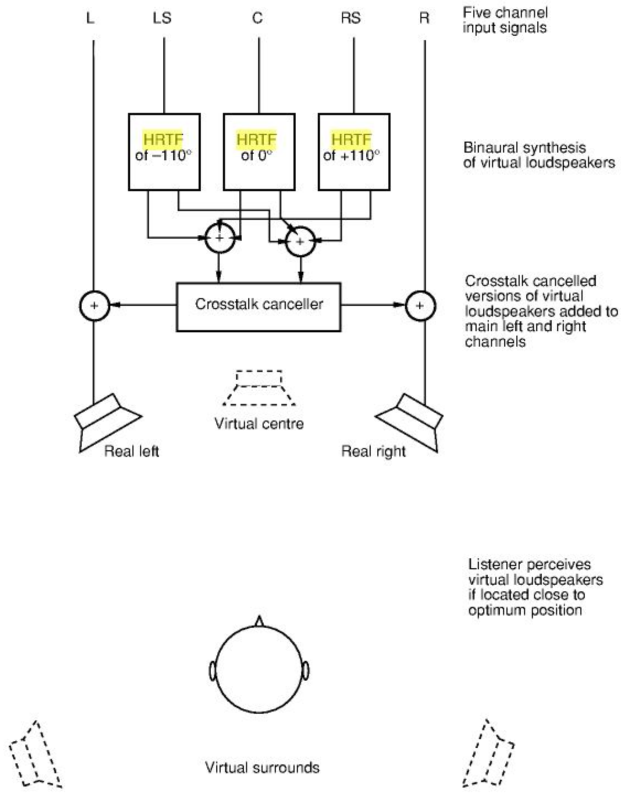


Figure 2.3: A diagram of a loudspeaker system set up for virtualization [20]

Auditory Scene Analysis(ASA) is the term given to the brains process of group and compartmentalizing sounds into components of the auditory 'scene' [20]. Through a complex process of continuous analysis and adaptation, the brain is able to accurately differentiate sounds into specific details such as how large a room is, what the wall materials are or which direction a complex source such as an oboe is facing. Bregman [4] suggests that a key part of the ASA process is learned via auditory steaming, which is essentially spectrogram analysis, comparing time and frequency based patterns to those already known, and associating those patterns as another object in the scene.

2.0.2 Ambisonics for virtual reality

Many of the recent VR systems consist of a viewing headset that may be powered by a mobile phone, a games console or a computer. The headsets viewing system provides a stereoscopic vision of a 3d environment that is linked to a head tracking interface, allowing the user to explore the 'immersive' visual environment in 3d. For many of the new virtual reality platforms, ambisonics has become the signal format of choice. This choice may be due to the flexability of a system that can encode and decode arbitrary numbers of input and output channels, while maintaining a high degree of spatial data. Another benefit may be the compatability of ambisonics with concepts such as object based audio CITE. For a review of various 'spatial' audio system formats, please refer to [21]. Many of the newer VR platforms rely on headphones as the preferred audio system format. This may be in part due to the difficulty in producing HRTFs for multiple listeners over loudspeaker, though research in this area is continually improving [9]. Another benefit of the use of headphones is that head tracking can be applied to HRTFs to improve the spatial audio quality [12] in a relatively discrete package.

Ambisonics in a simplistic description, is the spherical harmonic decomposition of a sound field into constituent components [20]. That is, by the combination and manipulation of sets of signals from coincident receivers with figure-of-eight polar responses with an omnidirectional receiver, it is possible to encode a 3 dimensional spatial sound field into an n channel signal format. For a discussion on different formats such as B, C, D, A and UHJ, please refer to Rumsey [20]. Perhaps the greatest benefit of the ambisonic system is that the encoding and decoding channel counts are not mutually exclusive i.e. it is possible to record a sound field in 3rd order using an appropriate sound-field microphone, and decode that signal for two channel headphone playback with the appropriate HRTFs [13] as could be used in a VR system [5]. This would allow for multiple virtual sound sources to be used to render a 3d sound field for an individual listener, and could be coupled with head tracking to create an auditory scene that changes as a listeners moves their head.

3 Reverberation & Perception

3.1 Reverberation

The reverberant sound field is the steady-state of diffusely scattered sound energy in a space due to the reflection of that energy from boundaries to a high order. Specifically, the amplitude of these reflections are such as to balance in amplitude with the steady state (source - decay) of the acoustic system, at or beyond the critical distance from a source (a classic analogy is given by Everest) [8].

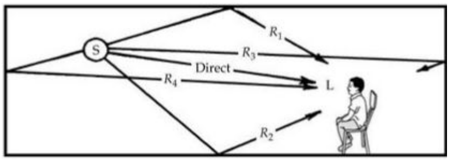
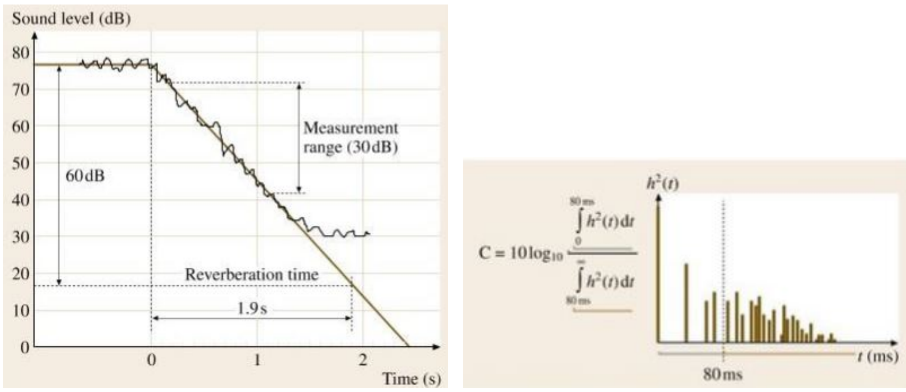


Figure 3.1: A conceptualisation of direct and reflected sound [8]

That is in contrast with low order reflections that may occur before the sound contributes to the reverberant field (early reflections), or may occur later and above the steady state amplitude (echos) [8]. Early and strong reflections are of significant interest in this study, due to the cues humans receive from perception of them. A reverberant sound field is often quantified by the decay time from steady state, to an amplitude of the steady state level $-60dB$. This is commonly described as the RT_{60} and was proposed by WC Sabine in 1900. For a discussion of reverb time calculation, please see Everest [8] page 153. The use of RT_{60} as the preferred metric of decay time is valid, assuming that the acoustics system is linear and time-invariant. A more comprehensive description of reverberation and overview of the associated parameters is given by Rossing [19].



(a) An example graph of reverberation decay time RT_{60}

(b) An example of the calculation of Clatiry score

Figure 3.2: Graphics pertaining to quantification of reverberant sound [19]

Further to this, the ratio of direct sound amplitude from a source to the amplitude of the reverberant field at any place in the sound field is also of interest [2] and is described as the **D**irect to **R**everberant (D/R) enegry ratio. This can be quantified in terms of amplitude at a point in the sound field using the modified Hopkins-Stryker equation [7]:

$$L_T = L_W + 10log\left(\frac{QMe}{4\pi D_x^2} + \frac{4N}{S\bar{a}Ma}\right) + K$$

where:

1. L_T = total sound pressure
2. L_W = sound power level of source
3. $\frac{QMe}{4\pi D_x^2}$ is a description of the sound source direct radiation properties
4. $\frac{4N}{S\bar{a}Ma}$ is a description of the reverberant field properties
5. $K = \frac{\rho c}{400}$ relating to the transfer of sound through air

This equation is particularly useful, as it can be rearranged to calculate important parameters such as the amplitude of the direct and reverberant sound fields, but also presents the intrinsic relationship between source parameters, receiver location and sound field geometry(including sound absorption properties etc.). The D/R ratio is intrinsically linked in this equation to the **critical distance**(D_c), the distance from sound source at which the reverberant sound field is equal in amplitude to direct sound energy from the source. At distances beyond D_c the reverberant sound field dominates the direct sound in amplitude. A third parameter of interest is the clarity score, that is determined by the ratio between direct sound and an integration of the sound received directly from the source for some arbitrary [2] time (commonly 80ms). Another parameter of note in calculation of the reverberant nature of a geometry is the mean free path (**MFP**), that is the average distance between reflections in space [7]:

$$\mathbf{MFP} = 4\frac{Volume}{Surfacearea}$$

As a wave must propagate further between reflective surfaces, the amplitude of higher order reflections is increasingly diminished. This is compounded by the absorption characteristics of the boundaries of the geometry. Thus, a higher mean free path may result in a reduction in the amplitude of a reverberant tail, compared to a much smaller space with the similar absorption characteristics.

3.2 Perception of Reflections

There is a plethora of research and understanding of the objective quantification of reverberation [19], less so about the subjective effects [14] beyond the seminal studies of early reflections and level threshold shifts by early pioneers (even before Haas) [10]. It may be argued that many facets of human auditory perception relate in some way to the perception of reverberation, the concepts of interest in this paper are those linked with localisation and perception of the auditory scene. Within these concepts are Haas/Precedence effect, and auditory masking.

The Haas effect is described as the first wave-front perceived by a listener, determines the source localisation in an auditory scene. The following reflections received within a 0.7 to 35mS window are integrated with the initial sound, such that the sound is given the impression of increased loudness, an increase in source width and tonal shifts of the percieved sound [8]. Beyond 40mS, following strong reflections are heard as echoes. Begault [2] suggests that the precedence effect is the intrinsic mechanism of the human auditory system, that allows for the localization of sound sources on a reverberant sound-field. Begault also suggests that there is a link between inter-aural time difference(ITD) perception for lateralization, and changes in apparent sound source 'width', suggesting a 5us to 1.5ms ITD 'window' within which a sound sources lateral location is determined. Another ascpet of note, is that if a sound source is occluded and the reflected sound is louder than the direct sound, the precedence effect is diminished [21].

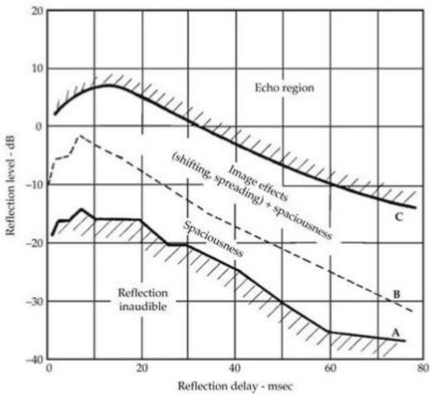


Figure 3.3: A plot of the window regions of precedence between reflection 'level' and delay [8]

Many of the texts referenced in this section of the report do not discuss auditory masking in great detail[2], though it may be of considerable importance when considering perception of noise-like sound. Auditory masking can be conceptualised as the time-envelope dependent blocking of new sounds being perceived, due to some other sound having already excited the inner ear. Specifically, sounds having already excited a portion of the bascilar membrane are thus stopping other sounds from being perceived as separate, that would otherwise excite the same portion of the bascilar membrane [8]. The basilar membrane is often regarded as to be discretized into critical bands or equivalent rectangular bandwidths (ERB), and these bandwidths are different depending on sound level as well as centre frequency. General works by Moore [16] should be reviewed for a more in-depth discussion of masking. The relevance of masking in this context, is that continuous noise is often used as a test signal in

[2]Masking is noted several times by Begault, but not discussed

masking experiments, and steady-state reverberation may be similar to noise and thus cause masking of reflections and general reverberation. As such, creation of realistic reverberation may be important for masking in a natural way as opposed to having a spectrally dense or high amplitude artificial reverb that may cause excessive masking.

Sigfried Linkwitz [15] also discussed the idea that human perception of sound in enclosed spaces is powerful enough that given an appropriate pair of sound sources, a person can subconsciously differentiate between reflected sounds and the stereo sound-stage created by the two sound sources. This is not disimilar to abstractions made in much of the literature such as Begault [2] or Wiggins [21], who describe typical scenarios pertaining to walking into large spaces and using cues from reflections to determine that the space they have entered is actually large. The relevant link between reverberation and associated perceptual faculties is with auditory scene analysis. Reverberation perception (particularly precedence) is intrinsically linked to humans capacity to analyse the auditory characteristics of their surroundings. This is compounded by Begault [2] who notes the significant breadth of reverberation research for concert hall acoustics, suggesting that the quality of reverberation may be an important part of audio for VR and multimedia.
The argument is that more realistic reverberation modelling may provide a more realistic scene for our brains to analyse. In fact it has been shown by Corey [6] that for a 5 channel system, adding simulated early reflections are enough to allow listeners to accurately localize sound sources with greater certainty. However, in a similar way some of the VR packages, Corey's early reflections were generated using an image source method. As note in acoustic modelling research such has those by Mourik and Murphy [17], geometric or ray based simulations do not produce accurate results at low frequencies. The image based method described by Begault [2], and used by Corey is also only usable for rectangular spaces and is limited further. Some reverb calculation methods are discussed in the next section.

3.3 Reverberation Algorithms

Derived Reverberation

Geometric Modelling

Combined Geometric and derived reverb.

Wave Modelling

4 Experiment Protocol

Users will be asked to evaluate 5 different instances of VR environments, and determine which audio is from the real environment.

Karjalainen *et al* [14] undertook a study into the perception of late reverberation, using a perceptual model to reinforce listening test data.

5 Lists

5.1 Example of list (3*itemize)

- First item in a list
  - First item in a list
    - First item in a list
    - Second item in a list
  - Second item in a list
- Second item in a list

5.2 Example of list (enumerate)

1. First item in a list
2. Second item in a list
3. Third item in a list

$$\begin{aligned}(x+y)^3 &= (x+y)^2(x+y)\\ &= (x^2+2xy+y^2)(x+y)\\ &= (x^3+2x^2y+xy^2)+(x^2y+2xy^2+y^3)\\ &= x^3+3x^2y+3xy^2+y^3\end{aligned} \tag{5.1}$$

$$A = \begin{bmatrix} A_{11} & A_{21} \\ A_{21} & A_{22} \end{bmatrix} \tag{5.2}$$

$RT_{60} = \frac{0.161A}{Sa}$

6 REFERENCES

REFERENCES

[1] Neil L. Aaronson and William M. Hartmann. Testing, correcting, and extending the Woodworth model for interaural time difference. *The Journal of the Acoustical Society of America*, 135(2):817–823, 2014.

[2] Durand R Begault. 3-d sound for virtual reality and multimedia. *Computer Music Journal*, 19(April):99, 1995.

[3] Jens Blauert. Spatial Hearing- The Psychophysics of Human of sound. *J. Acoust. Soc. Am.*, 77(1):334–335, 1997.

[4] Albert S Bregman. The Auditory Scene. *Auditory Scene Analysis: The perceptual organization of sound*, pages 1–45, 1994.

[5] T Collins. Binaural Ambisonic Decoding with Enhanced Lateral Localization. *134th AES Convention*, pages 1–10, 2013.

[6] Jason;Woszczyk Wieslaw Corey. Localization of Lateral Phantom Images in a 5-channel System with and without Simulated Early Reflections. *Aes*, pages 1–19, 2002.

[7] D. Davis and E. Patronis. *Sound System Engineering*. Elsevier Focal Press, 2006.

[8] F. Alton Everest and Neil A. Shaw. *Master Handbook of Acoustics, 5th Edition*, volume 110. focal press, 5th edition, 2009.

[9] Simon Galvez and Filippo Fazi. Listener Adaptive Filtering Strategies for Personal Audio Reproduction over Loudspeaker Arrays. *AES Conference on Sound Field Control*, 38:183–184, 2016.

[10] Helmut Haas. The Influence of a Single Echo on the Audibility of Speech. *J. Audio Eng. Soc*, 20(2):146–159, 1972.

[11] Norman N Holland. The Willing Suspension of Disbelief : A Neuro-Psychoanalytic View The Willing Suspension of Disbelief : A Neuro-Psychoanalytic View. *PsyArt : An Online Journal for the Psychological Study of the Arts*, pages 1–5, 2003.

[12] Kiyofumi Inanaga, Yuji Yamada, and Hiroshi Koizumi. Headphone System with Out-of-head Localisation applying Dynamic HRTF. *98th AES Convention*, 11:Convention Paper 4011, 1995.

[13] Jean-Marc Jot, Scott Wardle, and Véronique Larcher. Approaches to binaural synthesis. In *105th AES Conference*, volume 4861, 1998.

[14] M Karjalainen and H Jarvelainen. More about this reverberation science: Perceptually good late reverberation. *Audio Engineering Society Convention 111*, pages 1–8, 2001.

[15] Siegfried Linkwitz. THE MAGIC IN 2-CHANNEL SOUND REPRODUCTION - WHY IS IT SO RARELY HEARD ? *Reproduced Sound 2015*, pages 1–22, 2015.

[16] B C J Moore. Masking in the human auditory system. *Collected Papers on Digital Audio Bit-Rate Reduction*, pages 9–19, 1996.

[17] Jelle Van Mourik and Damian Murphy. Geometric and wave-based acoustic modelling using Blender. *AES 49th International Conference: Audio for Games*, pages 1–9, 2013.

[18] Oculus. Oculus connect: Introduction to audio in vr - youtube, October 2014. (Accessed on 11/21/2016).

[19] T. Rossing. *Springer Handbook of Acoustics*. Springer Handbook of Acoustics. Springer New York, 2007.

[20] F. Rumsey. *Spatial Audio*. Taylor & Francis, 2012.

[21] Bruce Wiggins. *An investigation into the real-time manipulation and control of three-dimensional sound fields*. PhD thesis, University of Derby, 2004.