DEPTH PERCEPTION IN INTERACTIVE VIRTUAL ACOUSTIC ENVIRONMENTS USING HIGHER ORDER AMBISONIC SOUNDFIELDS

Gavin Kearney, Marcin Gorzel, Frank Boland and Henry Rice

School of Engineering
Trinity College Dublin,
Ireland
gavin.kearney@tcd.ie

ABSTRACT

In this paper, we present an investigation into the perception of source depth in interactive virtual auditory environments in the context of Higher Order Ambisonic (HOA) reproduction. In particular, we investigate the accuracy of soundfield reproduction over virtual loudspeakers (headphone reproduction) with increasing Ambisonic order. Performance of 1^{st}, 2^{nd} and 3^{rd} order Ambisonics in representing distance cues is assessed in subjective audio perception tests. Results demonstrate that first order soundfields are sufficient in representing distance cues for virtual loudspeaker reproduction.

1. INTRODUCTION

Recent advances in interactive entertainment technology have led to visual displays with a convincing perception of source depth, based not only on stereo vision techniques, but also on real time graphics rendering technology for correct motion parallax [1][2]. Typically, such presentations are accompanied by loudspeaker surround technology based on amplitude panning techniques and aimed at multiple listeners. However, in interactive virtual environments, headphone listening allows for greater control over personalized soundfield reproduction. One method of auditory spatialization is to incorporate Head Related Transfer Functions (HRTFs) into the headphone reproduction signals. HRTFs describe the interaction of a listeners head and pinnae on impinging source wavefronts. It has been shown that for effective externalization and localization to occur, head-tracking should be employed to control this spatialization process [3]. However, the switching of the directionally dependent HRTFs with head movement can lead to auditory artifacts caused by wave discontinuity in the convolved binaural signals [4]. A more flexible solution is to form 'virtual loudspeakers' from HRTFs, where the listener is placed at the centre of an imaginary loudspeaker array. Here, the loudspeaker feeds are changed relative to the head position and any technique for sound source spatialization over loudspeakers can be used. Many different spatialization systems have been proposed for such application in the literature, most notably Vector Based Amplitude Panning (VBAP) [5] and Wavefield Synthesis [6]. However, the Ambisonics system [7], which is based on the spherical harmonic decomposition of the soundfield, represents a practical and asymptotically holographic approach to spatialization. It is well known in Ambisonic loudspeaker reproduction, that as the order of sound field representation gets higher, the localization accuracy increases due to greater directional resolution.

However, in distance perception, the accuracy of localization is related not only to the direct to reverberant ratio, but also to the correct reproduction of the direction of early reflections from the perspective of the listening position as well as high frequency absorption cues. There are many unanswered questions of the capability of Ambisonic techniques in this regard and in particular to virtual loudspeaker reproduction. In this paper, we hypothesize that the correct perception of auditory source depth in Ambisonic soundfields improves with increasing Ambisonic order.

This paper is outlined as follows: We will begin by presenting a succinct review of the relevant psychoacoustical aspects of auditory localization and depth perception. We will then outline the incorporation of Ambisonic techniques to virtual loudspeaker reproduction and subsequent re-synthesis of measured 1^{st} order Ambisonic soundfields into higher orders. A case study investigating the perception of source depth at higher Ambisonic orders is then presented through subjective listening tests.

2. DISTANCE PERCEPTION

Although the human ability to perceive sources at different distances is not fully understood, there are several key factors which are known to contribute to distance perception. In reverberant rooms, the ratio of the direct to reverberant sound plays an extremely important role. For near sources, where the direct field energy is much greater than the reverberant field, the level approximately changes in accordance to the free-field conditions. However, for source-listener distances greater than the critical distance, the level of reverberation is in general independent of the source position due to the homogeneous level of the diffuse field. Thus, at distances greater than 1m the direct to reverberant ratio changes approximately 6dB per doubling of distance. Consequently as the source moves further into the reverberant field, the level of reflections will also affect the perceived source width. Furthermore, the closer the sound is to the listener, the greater the initial time delay gap between the direct sound and first reflections.

The perception of distance has been shown to be one that is not linearly proportional to the source distance. For example, both Nielson et al [8] and Gardner [9], have shown that the localization of speech signals is consistently underestimated in an anechoic environment. This underestimation has also been shown by other authors in the context of reverberant environments, both real and virtual. In [10], Bronkhorst et al. demonstrate that in a damped virtual environment, sources are consistently perceived to be closer than in a reverberant virtual environment, due to the direct to reverberant ratio. In their studies, the room simulation is conducted using simulated BRIRs created from the image source method [11]. They show how perceived distance increases rapidly with the number and amplitude of the reflections. In a similar study, Rychtarikova et al [12] investigated the difference in localization accuracy between real rooms and computationally derived BRIRs. Their findings show that at 1m, localization accuracy in both the virtual and real environments is in good agreement with the true source position. However, at 2.4m, the accuracy degrades, and high frequency localization errors were found in the virtual acoustic pertaining to the difference in HRTFs between the model and the subject. In the same vain, Chan et al [13] have shown that distance perception using recordings made from the in-ear microphones on individual subjects again lead to underestimation of the source in virtual reverberant environments, more so than with real sources. Waller [14] has identified that one of the key factors in distance perception is the importance of listener movement in the virtual space, which has not been considered in the previous studies. In the Ambisonics sense, it is therefore important that the soundfield transformations reflect well the movements of the listener.

3. AMBISONIC SPATIALIZATION

Ambisonics was originally developed by Gerzon, Barton and Fellgett [7] as unified system for the recording, reproduction and transmission of surround sound. The theory of Ambisonics is based on the decomposition of the the soundfield measured at single point in space into spherical harmonic functions defined as

$$\begin{aligned} Y^{\sigma}_{mn}(\Phi,\Theta) &= A_{mn}P_{mn}(\cos\Theta) \\ &\times \begin{cases} \cos m\Phi & \text{if } \sigma=+1 \\ \sin m\Phi & \text{if } \sigma=-1 \end{cases} \end{aligned} \tag{1}$$

where m is the order and n is the degree of the spherical harmonic and P_{mn} is the associated Legendre function. For each order m there are $(2m+1)$ spherical harmonics. In order for plane wave representation over a loudspeaker array we must ensure that

$$s\,Y^{\sigma}_{mn}(\Phi,\Theta) = \sum_{i=1}^{I} g_i\,Y^{\sigma}_{mn}(\phi_i,\theta_i) \tag{2}$$

where s is the pressure of the source signal from direction (Φ,Θ) and g_i is the i^{th} loudspeaker gain from direction (ϕ_i,θ_i). We can then express the left hand side of the equation in vector notation, giving the Ambisonic channels

$$\begin{aligned} \mathbf{B}_{\Phi\Theta} &= \mathbf{Y}_{\Phi\Theta}s & (3)\\ &= [Y^{1}_{0,0}(\Phi,\Theta), Y^{1}_{1,0}(\Phi,\Theta), ...Y^{\sigma}_{mm}(\Phi,\Theta)]^{T}s & (4) \end{aligned}$$

Equation 2 can then be rewritten as

$$\mathbf{B} = \mathbf{C}\cdot\mathbf{g} \tag{5}$$

where $\mathbf{C}$ are the encoding gains associated with the loudspeaker positions and $\mathbf{g}$ is the loudspeaker signal vector. In order to obtain $\mathbf{g}$, we require a decode matrix, $\mathbf{D}$, which is the inverse of $\mathbf{C}$. However, to invert $\mathbf{C}$ we need the matrix to be a square which is only possible when the number of Ambisonic channels is equal to the number of loudspeakers. When the number of loudspeaker channels is greater than the number of Ambisonic channels, which is usually the case, we then obtain the pseudo-inverse of $\mathbf{C}$ where

$$\mathbf{D} = \text{pinv}(\mathbf{C}) = \mathbf{C}^{T}(\mathbf{C}\mathbf{C}^{T})^{-1} \tag{6}$$

Since the soundfield is represented by a spherical coordinate system, soundfield transformation matrices can be used to rotate, tilt and tumble the soundfields. In this way, the ambisonic signals themselves can be controlled by the user, allowing for the virtual loudspeaker approach to be employed. The number of I virtual loudspeakers employed with the Ambisonics approach is dependent on the Ambisonic order m where

$$I \geq N = (m+1)^2 \tag{7}$$

4. VIRTUAL LOUDSPEAKER REPRODUCTION

In the 'virtual loudspeaker' approach, HRTFs are measured at the 'sweet-spot' (the limited region in the centre of a reproduction array where an adequate spatial impression is generally guaranteed) in a multi-loudspeaker reproduction setup, and the resultant binaural playback is formed from the convolution of the loudspeaker feeds with the virtual loudspeakers. For the left ear we have

$$L = \sum_{i=1}^{I} h_{Li} * q_i \tag{8}$$

where $*$ denotes convolution and h_{Li} is the left ear HRIR corresponding to the i^{th} virtual loudspeaker and q_i is the i^{th} loudspeaker feed. Similar relations apply for the right ear signal. This method was first introduced by McKeag and McGrath [15] and examples of its adoption can be found in [16] and [17]. This approach has major computational advantages, since a complex filter kernel is not required and head rotation can be simulated by changing the loudspeaker feeds p as opposed to the HRTFs. Most existing research uses a block frequency domain approach to this convolution. However, given that the virtual loudspeaker feeds are controlled via head-tracking in real-time, a time-domain filtering approach can also be utilized. For *short* filter lengths, obtaining the output in a point wise manner avoids the inherent latencies introduced by block convolution in the frequency domain. A strategy for significant reduction of the filter length without artifacts has been proposed in [18] .

5. HIGHER ORDER SYNTHESIS

In order to compare the depth perception of different orders of Ambisonic soundfields, it is desirable to take real world soundfield measurements. However, the formation of higher order spherical harmonic directional patterns is non-trivial, and currently only first order soundfield microphones are widely commercially available. Thus, in order for us to change 1st order Ambisonic impulse responses to 3rd order representations, we will employ a perceptual based approach which will allow us to to synthesize the same spatial impression as would be experienced with a higher order Ambisonic soundfield recording. For this we adopt the directional analysis method of Pulkki and Merimaa, found in [19]. Here the B-format signals are analyzed in terms of sound intensity and energy in order to derive time-frequency based direction of arrival and diffuseness. The instantaneous intensity vector is given from the pressure p and particle velocity $\mathbf{u}$ as

$$\mathbf{I}(t) = p(t)\,\mathbf{u}(t) \tag{9}$$

Since we are using 1st order Ambisonic impulse response measurements, the pressure can be approximated by

$$p(t) = w(t) \tag{10}$$

and the particle velocity by

$$\mathbf{u}(t) = \frac{1}{\sqrt{2}Z_0}\left(x(t)\mathbf{e}_x + y(t)\mathbf{e}_y + z(t)\mathbf{e}_z\right) \tag{11}$$

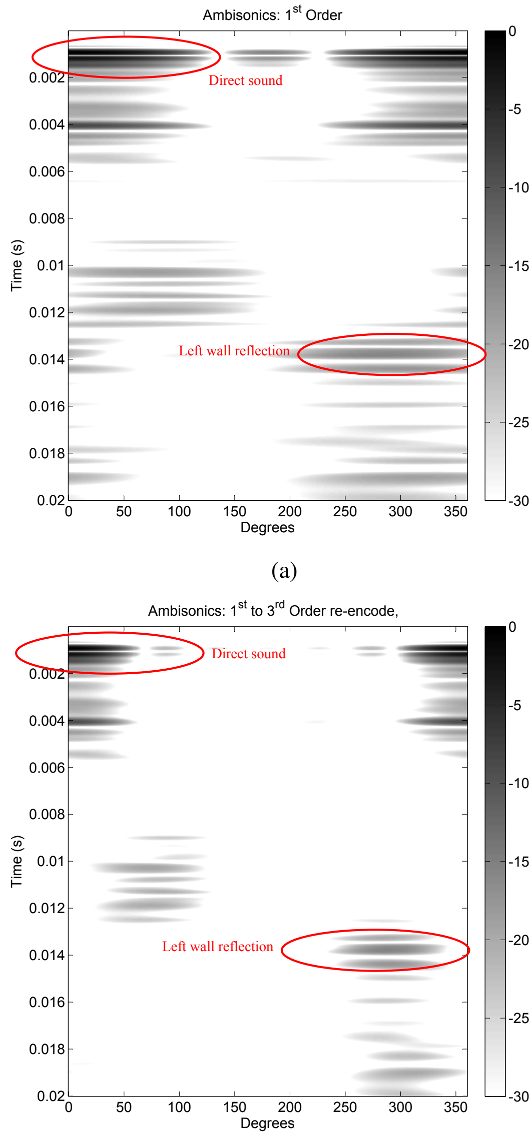
where $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ represent cartesian unit vectors and Z_0 is the characteristic acoustic impedance of air. The instantaneous intensity represents the direction of the energy transfer of the soundfield and the direction of arrival can be determined simply by the opposite direction of $\mathbf{I}$. For 1^{st} order Ambisonics, we can calculate the intensity for each coordinate axis, and in the frequency domain. Since a portion of the energy will also oscillate locally, a diffuseness estimate can be made which is given by the ratio of the magnitude of the intensity vector to the overall energy density given as

$$\psi = 1 - \frac{||\langle \mathbf{I} \rangle||}{c\langle E \rangle} \tag{12}$$

where $\langle \cdot \rangle$ denotes time averaging and $|| \cdot ||$ denotes the norm of the vector. The diffuseness estimate will yield a value of zero for incident plane waves from a particular direction, but will give a value of 1 where there is no net transport of acoustic energy, such as in the cases of reverberation or standing waves. Time averaging is used since it is difficult to determine an instantaneous measure of diffuseness.

The output of the analysis is then subject to smoothing based on the Equivalent Rectangular Bandwidth (ERB) scale, such that the resolution of the human auditory system is approximated. Since the frequency dependent direction of arrival of the non-diffuse portion of the soundfield can be determined, HOA reproduction can be achieved by re-encoding point like sources corresponding to the direction indicated in each temporal average and frequency band into a higher order spherical harmonic representation. The resultant Ambisonic signals are then weighted in each frequency band i according to $\sqrt{1-\psi_i}$. The diffuse field can be obtained by multiplying the first order signals by $\sqrt{\psi_i}$ and forming a first order decode. This is justified since it is only vital that the main directional information is re-encoded to higher order. Furthermore, if there exists a general directional distribution to the diffuse field, this will still be preserved in first order form.

Figure 1 shows an example of the first 20ms of a 1^{st} order impulse response taken in a reverberant hall. Here the source was located 3m from a Soundfield ST350 microphone, and the SRIR captured using the logarithmic sine-swept tone technique [20]. In these plots, particular attention is drawn to the direct sound (coming from directly in front of the microphone) and a left wall reflection at approximately 14ms. It can be seen that the directional resolution increases significantly with higher order Ambisonic representation. It should be noted, that since the A-format capsule on soundfield microphones only displays adequate directionality up to 10kHz [21]. Spatial aliasing is therefore an issue for high frequencies and as a result, the directional information above 10kHz cannot be relied upon.



(a)

(b)

Figure 1: Ambisonic soundfield from 1^{st} order measurement with a Soundfield ST350: (a) 1^{st} order representation, (b) 3^{rd} order re-encode.

6. LOCALIZATION OF TEST SOURCES

6.1. Test Phase 1: Depth perception of real sources in test environment

In order to effectively gauge the subjective performance of any spatialization system in a reverberant environment, it is first necessary to study the effect of room acoustics on localization accuracy. This can then be considered as the 'best case' scenario for any virtual audio tests performed regarding the same environment. In light of this, a series of experiments were set up in a small sized reverberant hall in Trinity College Dublin. The hall has a spatially averaged reverberation time of 0.95 seconds at 1kHz, and a critical distance of 1.4m.

The experimental setup used for tests is shown in Figure 2. Here, nine loudspeakers (Genelec 1029a) are spaced 1m apart from a reference 0m point at the listening position. In the design of this test, it was deemed necessary for the subject to clearly 'acoustically' see each loudspeaker at each position. However, unconsidered placement of the loudspeakers could potentially lead to horizontal localization cues that will bias the perception of source depth. Previous studies by the authors have revealed that there is an azimuthal localization blur of 0.25m and an elevation blur of 0.6m for a source located at 1m from the listener in this test environment [22]. Thus, the array was setup so as to not exceed these perceptual limitations, whilst giving the subject clear view of each loudspeaker. This is shown in Figure 2(b). The loudspeakers were calibrated to 80dBA at 1m from their on-axis tweeter position.

(a) (b)

Figure 2: Loudspeaker setup for depth perception tests: (a) Side perspective of staggered loudspeakers (b) Subject perspective.

The first set of test stimuli were presentations of anechoic music, extracted from the Denon anechoic orchestral database. The second set consisted of female speech samples of phonetically balanced phrases selected from the TIMIT Acoustic-Phonetic Continuous Speech Corpus database, and re-recorded at higher resolution (96kHz, 24-bit) [23]. The test subjects consisted primarily of music technology students under 35 years of age and of good hearing. 15 participants were used in total. Each participant was first subject to a training session prior to the tests, where they were presented with stimuli from each of the loudspeakers and were asked to become accustomed to the source types and acoustics of the room, and most significantly the perception of depth. During the course of this training a visual indicator in the form of an LED attached to top of each loudspeaker confirmed the active loudspeaker.

In the official test phase, each participant was presented with stimuli from pseudorandom loudspeaker positions and asked to identify the location of the sources via the test interface shown in Figure 3. This randomized method was used to negate any order effects during the tests. Once presented with stimulus the subject chose which loudspeaker they felt the sound originated from and visual confirmation of their choice was given by the LED on top of their chosen loudspeaker. The interface, audio engine and lighting control were all implemented in Pure Data [24].

6.2. Test Phase 2: Depth perception of virtual sources in test environment

In the second analysis phase subjects were asked to identify source depth using Ambisonic soundfields presented over headphones. The test stimuli again consisted of the same music and female speech samples, but convolved with Ambisonic impulse responses. Prior to this test phase, first order Ambisonic impulse response measurements were taken from the listener position of each loudspeaker using the logarithmic sine swept tone method [20]. From these measurements, 2nd and 3rd order impulse response sets were extracted using the directional analysis approach outlined in Section 5.

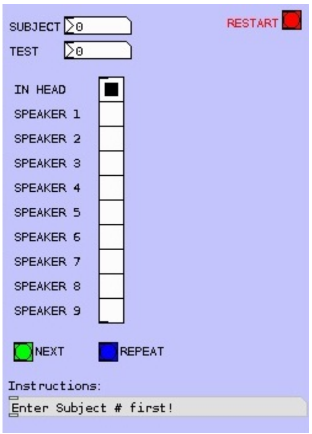


Figure 3: Test interface for depth perception tests.

24 virtual loudspeakers were implemented for the test, arranged in equispaced diametrically opposed pairs around the sphere of the listener. The headphones used were AKG-K601 open back headphones, which exhibit low levels of interaural magnitude and group delay distortion. Soundfield rotation, tilt and tumble control was implemented via an IntertiaCube head tracker system, resulting in stable virtual images with head movement. It is important to note, that non-individualized HRTFs were used in these tests. However, the resultant externalization was extremely effective, largely due to head-tracking, and no front-back confusion was reported. The HRTFs used were extracted from the IRCAM LISTEN database (subject 1021) and were diffuse field equalized [25].

The Ambisonic decodes were also psychoacoustically optimized. Shelf filters were implemented to satisfy the Gerzon localization criteria for maximized velocity decode at low frequencies and energy decode at higher frequencies [26]. This involved changing the ratio of the pressure to velocity components over the full spectrum. Whilst the crossover frequency for the high frequency boost in the pressure channel at first order is normally in the region of 400Hz for regular loudspeaker listening, here, we restore the crossover point to 700Hz, since the subject is always perfectly centre in the virtual loudspeaker array. Near field compensation filters were also employed to overcome any low frequency errors in the reproduced soundfields due to the close proximity of the virtual loudspeakers [27].

7. RESULTS

In this experiment, the perceived source depth from the 15 listeners for each stimuli was collected for 5 different positions (1m, 2m, 4m, 6m and 8m) for both the reference sources and for 1st, 2nd and 3rd Order Ambisonics . For each source type and for each position, we have computed the average distance μ (over the 15 subject answers) and the corresponding standard error $se(\mu)$. We then compute the difference in between the average distance for the real monophonic source presentations and the average distance for each Ambisonic presentation. For example, if we wish to consider the spatialization at the intended position of 2m from the listener for the music stimulus, for 1st

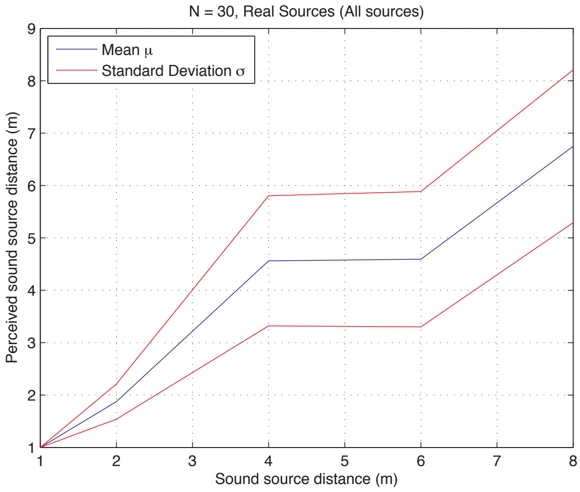


Figure 4: Depth perception of real sources in test environment.

order then we have

$$\begin{aligned} d \;=\; & \mu(\text{Sys} = \text{FOA}, \text{Sound} = \text{music}, \text{Pos} = 2\text{m}) \\ & -\mu(\text{Sys} = \text{`REF'}, \text{Sound} = \text{music}, \text{Pos} = 2\text{m}) \end{aligned} \quad (13)$$

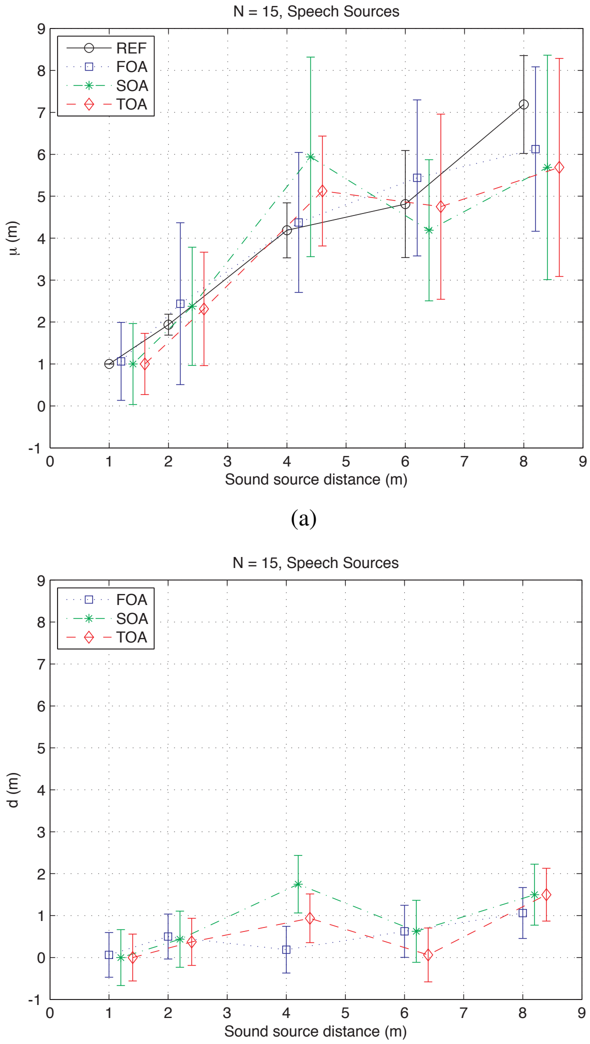
We also compute its standard error $se(d)$.

As expected, the perception of depth for the real sources was more accurate for near sources. Beyond 4m, distance perception was continuously underestimated which is congruent with the previous studies outlined in Section 2. This is clearly seen in the plot of the mean localization for all reference sources shown in Figure 4. Furthermore, the standard deviation of localization increases as the source moves further into the diffuse field. Figures 5(a) and 6(a) show the reference localization results for speech and music sources respectively. It can be seen that the speech sources are better localized than music.

The depth perception for the virtual speech sources is also shown in Figure 5 (a). The mean localization of the virtual sources follows the reference source localization well. The virtual sources deviate from the mean results in the same fashion as the reference sources, as localization becomes more difficult within the diffuse field. On the same scale, we see the error of the virtual speech sources from the reference sources in Figure 5 (b). It is noteworthy that the extent of the error is within 1.5m and the precision on the results indicate no significant statistical difference between 1^{st} , 2^{nd} and 3^{rd} order Ambisonics. Similarly for music sources, the general trend is that the Ambisonic virtual sources match the reference sources well. This is shown in Figure 6(a). Again, low errors of within 1.5m are exhibited, and overall each order performs well, despite the use of non-individualized HRTFs. Informal comments from subjects after the tests indicated that whilst the overall spatial impression was more defined with the higher order samples, it did not affect their judgement of the source distance. Moreover, the level of the direct sound to the diffuse field, as well as the stable virtual imaging was reported to give strong distance cues.

8. CONCLUSIONS

We have assessed through subjective analysis the perceived source depth in virtual Ambisonic soundfields in comparison to real world sources. Higher order soundfield synthesis was achieved using the directional analysis method of [19]. It was shown that Ambisonic reproduction matches the perceived real world source distances well at each order. No significant statistical difference was exhibited by increasing the Ambisonic order in this regard. It must be emphasized however, that this analysis applies to virtual loudspeaker decode only and that each decode has been psychoacoustically optimized. Further work will examine this topic for loudspeaker reproduction for both centre and off-center listening as well as investigating HOA synthesis in comparison to real world HOA measurements.



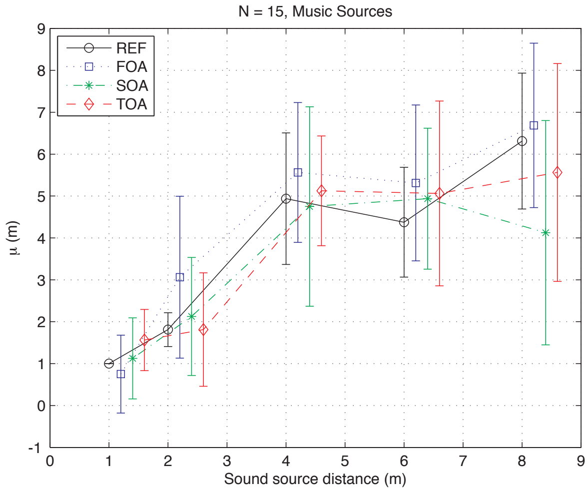
(a)

(b)

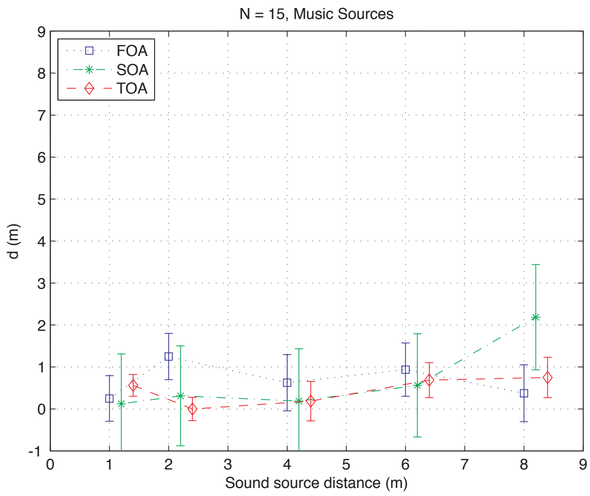
Figure 5: Depth perception of virtual speech sources in test environment. (a) Mean localization (b) d-error (95% confidence interval)

9. ACKNOWLEDGMENTS

The authors gratefully acknowledge the participation of the test subjects for both their time and constructive comments. This research is supported by Science Foundation Ireland.



(a)



(b)

Figure 6: Depth perception of virtual music sources in test environment. (a) Mean localization (b) d-error (95% confidence interval)

10. REFERENCES

[1] L. Fauster, "Stereoscopic techniques in computer graphics", 2007, thesis, TU Wien.

[2] J. Lee, "Head tracking for desktop VR displays using the Wii remote", 2007, http://www.cs.cmu.edu/~johnny/projects/wii/.

[3] D. R. Begault, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual sound source", *Journal of the Audio Engineering Society*, vol. 49, no. 10, 2001.

[4] M. Otani and T. Hirahara, "Auditory artifacts due to switching head-related transfer functions of a dynamic virtual auditory display", *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E91-A, no. 6, pp. 1320–1328, 2008.

[5] V. Pulkki, "Virtual sound source positioning using Vector Base Amplitude Panning", *Journal of the Audio Engineering Society*, vol. 45, pp. 456–466, 1997.

[6] A. J. Berkhout, "A Holographic Approach to Acoustic Control", *Journal of the Audio Engineering Society*, vol. 36, pp. 977–995, 1988.

[7] M. A. Gerzon, "Periphony: With-height sound reproduction", *Journal of the Audio Engineering Society*, vol. 21, pp. 2–10, 1973.

[8] S. H. Nielsen, "Auditory distance perception in different rooms", *Journal of the Audio Engineering Society*, vol. 41, no. 10, pp. 755–770, 1993.

[9] M. B. Gardner, "Distance estimation of 0[degree] or apparent 0[degree]-oriented speech signals in anechoic space", *The Journal of the Acoustical Society of America*, vol. 45, no. 1, pp. 47–53, 1969.

[10] A. W. Bronkhorst and T. Houtgast, "Auditory distance perception in rooms", *Nature*, vol. 397, pp. 517–520, February 1999.

[11] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics", *JASA*, vol. 65, pp. 943–950, 1979.

[12] M. Rychtarikova, T. V. d. Bogaert, G. Vermeir, and J. Wouters, "Binaural sound source localization in real and virtual rooms", *Journal of the Audio Engineering Society*, vol. 57, no. 4, pp. 205–220, 2009.

[13] J. S. Chan, C. Maguinness, D. Lisiecka, C. Ennis, M. Larkin, C. O'Sullivan, and F. Newell, "Comparing audiovisual distance perception in various real and virtual environments", in *32nd European Conference on Visual Perception*, August 2009.

[14] D. Waller, "Factors affecting the perception of interobject distances in virtual environments", *Presence: Teleoper. Virtual Environ.*, vol. 8, no. 6, pp. 657–670, 1999.

[15] A. McKeag and D. McGrath, "Sound field format to binaural decoder with head-tracking", in *6th Australian Regional Convention of the AES*, 1996.

[16] M. Noisternig, A. Sontacchi, T. Musil, and R. Holdrich, "A 3D Ambisonic based binaural sound reproduction system", in *24th International Conference of the Audio Engineering Society*, June 2002.

[17] B.-I. Dalenbäck and M. Strömberg, "Real time walkthrough auralization - the first year", in *Proceedings - Institute of Acoustics, vol. 28, Pt.2*, 2006.

[18] C. Masterson, S. Adams, G. Kearney, and F. Boland, "A method for head related impulse response simplification", in *Proc. of the 17th European Signal Processing Conference (EUSIPCO)*, August 2009.

[19] J. Merimaa and V. Pulkki, "Spatial impulse response rendering I: Analysis and synthesis", *Journal of the Audio Engineering Society*, vol. 53, 2005.

[20] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique", in *108th Convention of the Audio Engineering Society*, 2000.

[21] M. Gerzon, "The design of precisely coincident microphone arrays for stereo and surround sound", in *50th Convention of the Audio Engineering Society*, 1975.

[22] M. Gorzel, "Investigations into OPSI with height", M.Phil Masters Thesis, Music and Media Technology, Trinity College Dublin, June 2009.

[23] W. M. Fisher, G. R. Doddington, and K. M. Goudie-Marshall, "The darpa speech recognition research database: Specifications and status", in *Proceedings of DARPA Workshop on Speech Recognition*, 1986.

[24] M. Puckette, "Pure data," 2009, http://puredata.info/.

[25] IRCAM, "LISTEN HRTF database", 2003, http://recherche.ircam.fr/equipes/salles/listen/index.html.

[26] M. A. Gerzon and G. J. Barton, "Ambisonic decoders for HDTV", in *92nd Convention of the Audio Engineering Society*, March 1992.

[27] J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new Ambisonic format", in *23rd International Conference of the Audio Engineering Society*, 2003.