



HAL
open science

THIS PRESENTATION IS CANCELLED (PAPER IS AVAILABLE). On loudspeaker rendering of auditory distance in higher order Ambisonics

Gavin Kearney, Marcin Gorzel, Frank Boland

► **To cite this version:**

Gavin Kearney, Marcin Gorzel, Frank Boland. THIS PRESENTATION IS CANCELLED (PAPER IS AVAILABLE). On loudspeaker rendering of auditory distance in higher order Ambisonics. Acoustics 2012, Apr 2012, Nantes, France. hal-00811349

HAL Id: hal-00811349

<https://hal.science/hal-00811349>

Submitted on 23 Apr 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



ACOUSTICS 2012

THIS PRESENTATION IS CANCELLED (PAPER IS AVAILABLE). On loudspeaker rendering of auditory distance in higher order Ambisonics

G. Kearney^a, M. Gorzel^b and F. Boland^b

^aUniversity of York, Department of Theatre, Film and Television, University of York, Baird Lane Heslington (East), YO10 5GB York, UK

^bDepartment of Electronic and Electrical Engineering [Dublin], Trinity College Dublin Dept. of Electronic and Electrical Engineering Printing House Dublin 2
gavin.kearney@york.ac.uk

In this paper we investigate the perceived distance of sound sources rendered over loudspeaker arrays. We consider the perception of sound source distance with sources rendered at the loudspeakers as well as beyond the loudspeaker array radius. In particular, we explore this perception with Ambisonic soundfields up to 3rd Order. Sources are rendered for distances ranging from 2m to 8m and soundfields are presented over a 16 channel spherical loudspeaker array. Subjects are asked to gauge the source distance of female speech and pink noise bursts using Ambisonics reproduction. Test stimuli are created from measurements of 1st order spatial impulse responses in a reverberant room and are encoded to higher order spherical harmonic representations using the Directional Audio Coding methodology. Results demonstrate that the perception of source distance is largely independent of the Ambisonic order and more-so dependent on the monaural cues of level difference and direct-to-reverberant ratio.

1 Introduction

Immersive interactive media systems such as games consoles now accommodate stereoscopic 3-D graphics to enhance the visuals and create the perception of visual sources at different distances. Typically, such entertainment systems are supported by horizontal based 2-channel, 5.1 or 7.1 surround sound systems based on amplitude panning techniques. Recent commercial attempts also include so-called ‘with-height’ surround sound, such as Dolby Pro Logic IIz [1], DTS 7.1 (with-height layout) [2] and Auro 3-D [3] to synthesize a more immersive acoustic around the listener, whilst offering a degree of backwards compatibility with stereo and 5.1 surround. However, spherical loudspeaker arrays, where the loudspeakers are evenly distributed around the central listening position, have the ability to render the physically correct velocity and pressure components of the soundfield at the sweet-spot, yielding full periphonic 3-D surround sound. This represents the Ambisonic approach, first introduced by Gerzon in the 1970’s [4].

Little has been investigated in the literature on the perception of distance using spherical loudspeaker arrays, in particular when the area of effective soundfield reproduction increases due to greater angular discrimination caused by increasing the order of reproduction (i.e. increasing the number of spherical harmonic basis functions used to describe the soundfield). Whilst raising the Ambisonic order has been shown to increase directional localisation accuracy, thereby making virtual sources more ‘point-like’ [5], it is debatable whether this also has an effect on the perception of distance, in particular in the presence of strong monaural distance cues such as level changes and direct-to-reverberant ratio. In this paper, we investigate the perception of source distance using Ambisonic soundfields of 1st and 3rd Order over a spherical loudspeaker array. The paper is organised as follows: First, we present a succinct review of auditory distance perception, as well as previous studies implemented for synthesized soundfields. We will then outline an experiment conducted over a 16 channel spherical loudspeaker array which investigates the perception of auditory distance with increasing Ambisonic order. Finally, the results of the experiment are presented and conclusions pertaining to the effectiveness of Higher Order Ambisonics (HOA) at delivering distance cues are drawn.

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. Firstly, changes in distance lead to changes in the monaural transfer function (the sound pressure at one ear) due to source intensity decaying in accordance with the inverse square law. For large distances and high sound pressure levels, the propagation speed of a sound wave in a medium ceases to be constant with frequency, which may lead to distortion of the waveform [6]. Furthermore, sound waves traveling a substantial distance also undergo a process of energy absorption by water molecules in the atmosphere. This is more apparent for high-frequency energy of the wave and leads to spectral changes (low-pass filtering) of the sound being heard.

However, absolute monaural cues will only be meaningful if we have some prior knowledge of the source level, i.e., how familiar we are with the source. In other words, a form of *semiosis* occurs, where the perception of localization is based on anticipation and experience [7]. For example, for normal level speech (approximately 60dB at 1m), we expect nearer sources to be loud, and quieter sources further away. However, this is more difficult to assess for synthetic sounds or sounds that we are unfamiliar with.

In reverberant rooms, the ratio of the direct to reverberant sound plays an extremely important role in distance perception. For near sources, where the direct field energy is much greater than the reverberant field, the sound pressure 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 and the direct to reverberant ratio changes approximately 6dB per doubling of distance from the source. The directions of arrival of the early reflections are another parameter, which change according to the source-listener position and can be regarded as an important factor in creating environmental depth. Whether it is useful to the listeners in determining the distance to the sound source in the presence of other cues like sound intensity, direct to reverberant energy ratio or the arrival pattern of delays, remains an open question that needs to be addressed. Ambisonics allows for enhanced directional reproduction of deterministic components of a sound field by increasing the order of spherical harmonic decomposition. However, better directional localization can be achieved without affecting other important cues for distance estimation like overall sound intensity or direct to reverberant energy ratio. Thus it can constitute an ideal framework for testing whether less apparent properties of a sound field can influence the perception of distance.

2.1 Former Psychoacoustical Studies on Distance Perception

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 Binaural Room Impulse Responses (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 vein, 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 distance in virtual reverberant environments, more so than with real sources. The authors have also investigated the perception of source distance using binaural synthesis [14]. Simulation of a virtual loudspeaker array was achieved through HRTF measurements of test subjects and Ambisonic soundfields of different orders with source distances ranging between 2m and 8m were investigated. It was found that the order of Ambisonics does not contribute significantly to the perception of environmental depth. Indeed, the current study represents a real-loudspeaker version of this work.

3 Ambisonic Spatialization

Ambisonics was originally developed by Gerzon, Barton and Fellgett [4] as a unified system for the recording, reproduction and transmission of surround sound. The theory of Ambisonics is based on the decomposition of the sound field measured at a single point in space into spherical harmonic functions defined as

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

where m is the order and n is the degree of the spherical harmonic and P_{mn} is the fully normalized (N3D) associated Legendre function. The coordinate system used comprises x , y and z axes pointing to the front, left and up respectively, Φ is the azimuthal angle with the clockwise rotation and Θ is the elevation angle from the x - y plane. For each order m there are $(2m + 1)$ spherical harmonics. In order for plane wave representation over a loudspeaker array we must ensure

that

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

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 Eq. 2 in vector notation, giving the Ambisonic channels

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

Eq. 2 can then be rewritten as

$$\mathbf{B} = \mathbf{C} \cdot \mathbf{g}, \quad (4)$$

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

$$\mathbf{D} = \text{pinv}(\mathbf{C}) = \mathbf{C}^T (\mathbf{C}\mathbf{C}^T)^{-1}. \quad (5)$$

Since the sound field is represented by a spherical coordinate system, sound field transformation matrices can be used to rotate, tilt and tumble the sound fields. In this way, the Ambisonic signals themselves can be controlled by the user, allowing for the virtual loudspeaker approach to be employed. For 3-D reproduction, the number of I virtual loudspeakers employed with the Ambisonics approach is dependent on the Ambisonic order m , where

$$I \geq N = (m + 1)^2. \quad (6)$$

4 Higher Order Synthesis

In order to compare the distance perception of different orders of Ambisonic sound fields, it is desirable to take real world sound field measurements. However, the formation of higher order spherical harmonic directional patterns is non-trivial. Thus, in order for us to change FOA impulse responses to HOA representations, we will employ a perceptual based approach which will allow us to synthesize the increased directional resolution that would be achieved with a HOA sound field recording. For this we adopt the directional analysis method of Pulkki and Merimaa, found in [15]. 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 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 sound field 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. However, it is only vital to re-encode non-diffuse components to higher order

and the diffuse components can be rendered using a first order decode. This is justified since source localisation is dependent on the direction of arrival of the direct sound and early reflections and not on late room reverberation [16]. Thus, from the perceptual point of view, it is questionable whether there is a need to preserve the full directional accuracy of the reverberant field. Furthermore, if there exists a general directional distribution to the diffuse field, this will still be preserved in first order form.

5 Test Methodology

Different protocols have been used in literature for subjective assessment of distance perception, most notably a verbal report [17, 18], direct or indirect blind walking [19, 20] or imagined timed walking [20]. All of these methods have proved to provide reliable and comparable results for both, auditory and visual stimuli, with direct blind walking exhibiting the least between-subject variability [19, 20]. In former work [21], authors of this paper developed a method where subjects indicated the perceived distance of real and virtual sound sources by selecting one of several physical loudspeakers lined up (and slightly offset in order to provide ‘acoustic transparency’) in front of their eyes. However, for the present study, in order to completely eliminate any possible anchors as well as visual cues, it was decided to utilize the method of direct blind walking. Of the main concerns in the experiment was a direct comparison of distance perception of real sound sources versus virtual sound sources presented over the loudspeaker array. Due to different apparatus requirements, the experiment had to be conducted in two separate phases.

5.1 Test Phase 1

A series of subjective listening tests was conducted in the Large Rehearsal Room in the Department of Theatre, Film and Television in the University of York. The room dimensions were 12 x 9 x 3.5[m] and the spatially averaged T_{60} at 1kHz was 0.26s. A low T_{60} was desired for this study, so the walls were covered with thick, heavy curtains, as shown in Fig. 1. Since the up-mix from 1st to 3rd order Ambisonics concerned only the deterministic part of the measured SRIRs, it was assumed that no advantage would be gained from using a more reverberant space.

A professional camera dolly track was set up roughly in the direction of the diagonal of the room. It not only allowed for testing distances of the real loudspeaker up to 8m but its non-symmetrical position also assured that early reflections of the same order from different surfaces did not easily coincide at the subjects ears, but instead arrived at different times. A single full-range loudspeaker (Genelec 8050A) was mounted on a camera dolly which enabled it to be noiselessly translated by the experiment assistant to different locations. The guiding rope was hung along the dolly track which was intended to help and guide the participants when walking toward the sound source. Since it was not possible to walk exactly on the dolly track, it was decided that the walking path would be directly next to it, as shown in Fig. 1. The only weakness of this solution was that the sound source horizontal angle varied from 14.04 degrees at the closest distance (2m) to 3.58 degrees at the



Figure 1: Participant performing a trial during Phase 1 of the experiment. Note that the headphones on the subject were part of previous study on binaural distance perception implemented by the authors [14].

furthest distance (8m). However, this did not have any effect on the distance judgments for two reasons: Firstly, the subjects were allowed (or even encouraged) to rotate their head in order to fully utilize the available Interaural Time Difference (ITD) and Interaural Level Difference (ILD) cues. Secondly, the initial head orientation was not in any way fixed. This, combined with the fact that there were no clear cues to the subject’s initial orientation in the room at the origin, made this small initial angular offset unimportant. Furthermore, none of the participants reported any bias in their assessment based on the horizontal offset of the sound source. Seven participants aged 24-58 took part in this phase of the experiment. All subjects were of good hearing and were either music technology students or practitioners actively involved in audio research or production.

The stimuli used in the experiment were pink noise bursts and phonetically balanced phrases selected from the TIMIT Acoustic-Phonetic Continuous Speech Corpus database and recorded by a female reader [22]. A sampling rate of 44.1kHz and 16-bit resolution was used. These two sample types were selected in order to represent both unfamiliar and familiar sound sources. They were presented to the subjects in a pseudo-randomized manner to avoid any ordering effects.

5.2 Test Phase 2

For loudspeaker reproduction, prior to the test phase, FOA impulse response measurements were taken from the listener position of each loudspeaker distance using the exponentially swept-sine tone technique [23]. From these measurements, 2nd and 3rd order impulse response sets were extracted using the directional analysis approach outlined in Section 4.

The only psychoacoustical optimization applied to the Ambisonics decodes was shelf filtering and was intended to satisfy Gerzon’s localization criteria for maximized velocity



Figure 2: Participant performing a trial during Phase 2 of the experiment

decode at low frequencies and energy decode at higher frequencies [24].

For the trials a purpose built, 16 channel loudspeaker array consisting of Genelec 8050A loudspeakers, arranged in a sphere was utilized. The array is housed in the Audio Lab at the Department of Electronics at the University of York. Acoustic absorbers and damping curtains were utilised to minimise the effect of early reflections in the listening space. Again, a guiding rope was erected to help the blindfolded participants walk towards the source. 10 participants undertook this phase of the experiment. The array radius is at 1.9m, with loudspeakers arranged in 3 tiers of 4 (bottom), 8 (middle) and 4 (top) loudspeakers respectively. 1st Order soundfields were reproduced over the inner cube of the array (bottom 4 and top 4 loudspeakers) and 3rd Order soundfields were rendered over the full array. Whilst the 16 loudspeaker arrangement is not fully optimal for 3rd order reproduction (due to the fact that the not all inter-loudspeaker spacing is equidistant), the arrangement offers a good compromise between 1st and 3rd Order reproduction for comparison, as well as ease of calibration.

5.3 Procedure

In both experiments, subjects entered the test environment blindfolded and without any prior expectation regarding the room dimensions, its acoustic properties or the test apparatus. They were guided by the experimenter to the reference point (the ‘origin’). After a short explanation of the experiment objectives, a training session began with a short (3-5min) walking-only trial until participants felt comfortable with walking blindfolded and using a guide rope. Next, they performed 4-6 training trials in which the same test stimuli to be used in the experiment (speech and pink noise) were played by the loudspeaker at randomly chosen distances. No feedback was given and no results were recorded after each test trial. The end of the training session was clearly announced and after a 1 minute interval, the first phase of the test began.

In test phase I, participants were asked to listen to static sound sources at randomly chosen points, focusing on the perceived distance. They could listen to any audio sample as many times as they wished. After the playback had stopped, they were asked to walk guided by the rope to the point

where they thought the sound originated from. The distance walked was subsequently recorded by the assistant using a laser measuring tool, after which the participant walked backwards to the origin. In the meantime, the loudspeaker was noiselessly translated to its new position and the test proceeded. Similar to the training session, no feedback was given at any stage.

Participants had to indicate the perceived distance for sound sources randomly located at 2m, 4m, 6m or 8m. Taking into account that both speech and pink noise bursts samples were used (in a pseudo-random order), the number of trials in the first phase added up to 8. Each subject performed all the trials only once.

During test phase II, subjects were also asked to identify the sound source distance, but this time using Ambisonic sound fields presented over the loudspeaker array. Other than the fact that loudspeakers were used, and the reproduction environment was now a dedicated listening room, the test protocol remained the same as in phase I. However, due to the fact that there were two playback configurations to be tested (1st and 3rd order Ambisonics), participants had to perform 16 trials instead of 8. Instead of separate phases for each Ambisonic order, all samples were randomly presented to the subject within the same test phase. Again, subjects performed all the trials only once and no feedback was given at any stage.

6 Results

The perceived sound source distance (indicated by the distance walked) was collected for each subject for 4 presentation points (2m, 4m, 6m and 8m), two stimuli (female speech and pink noise bursts) and three playback options (real sources, First Order Ambisonics (FOA) and Third Order Ambisonics (TOA)). We computed the mean values of walked distances μ for each test condition along with the corresponding standard errors $se(\mu)$. The results are presented separately for each stimulus type within 95% Confidence Intervals.

As expected, the perception of distance was more accurate for near sources. Beyond 4m, distance perception was continuously underestimated in all cases, which is congruent with the previous studies outlined in Section 2. Furthermore, the standard deviation of localization generally increases as the source moves further into the diffuse field.

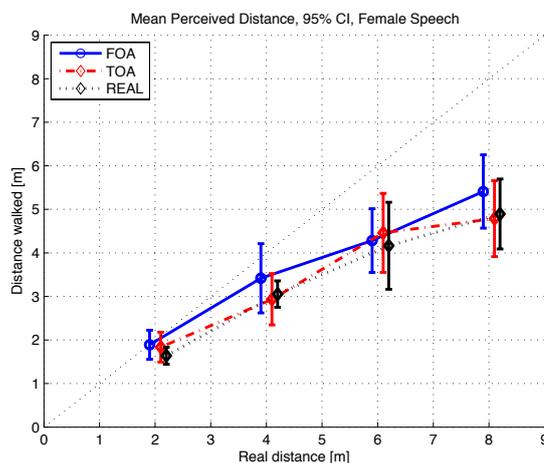


Figure 3: Mean distance perception: female speech.

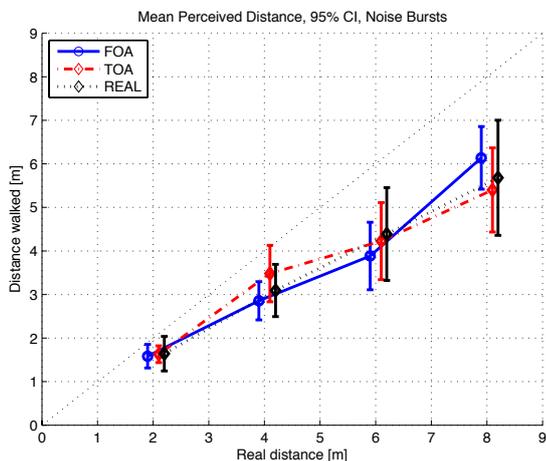


Figure 4: Mean distance perception: pink noise.

More importantly there exists no major significant difference in the mean perceived distance between the Ambisonics renderings and the real sources. This is true for both first and higher order Ambisonics, as well as across both source types.

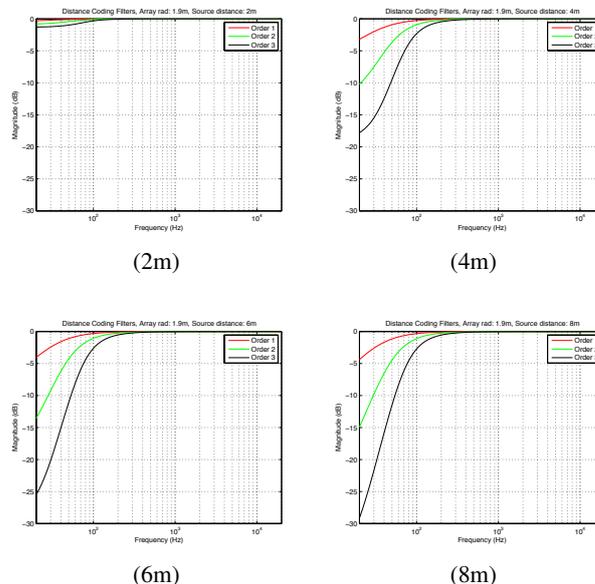
7 Discussion

The results presented for real sources corroborate the classic underestimation of source distance, as reported in the literature. These results were used as a basis with which to measure the ability of Ambisonic sound fields of different orders to present sources at different distances. It was expected that a further underestimation of the source distance would ensue with the virtual source rendering, as reported in [13]. However, this was not the case, even for first order presentations, and the apparent distances of the virtual sources matched the real source distances well.

Moreover the presented study demonstrates that the enhanced directional accuracy gained by presenting sound sources in HOA over loudspeakers does not yield a significant improvement in the perception of the source distance. What is noteworthy is that for low and high orders, there is no significant difference in the perception of the source location when compared to real-world sources. We therefore conclude that sound field directionality for distance perception is sufficient with 1st order playback.

It is important to note that the whilst the perception of distance can be sufficiently achieved with first order systems, the perception of naturalness and directional localisation are enhanced using higher orders. Informal discussions with the participants after the trials demonstrated a preference to 3rd order playback for the aforementioned reasons.

One should also note that distance compensation filtering (due to near field effects) was not implemented in this study. This is because the combination of the array radius (2m) and the source distances (<2m) leads to near field effects only prominent below 100Hz. Appropriate distance coding filters are shown in Figure 5 up to 3rd Order. For the female speech test stimuli, such filtering was unnecessary, since the first formant frequencies do not go down below 180Hz. Furthermore, such filtering would only be relevant for the direct sound component, whose distance is known and not the early reflections. For these reasons pink noise delivery was also bandlimited to 100Hz.

Figure 5: Distance coding filters required for direct sound compensation up to 3rd Order Ambisonics

Finally, there was no significant difference in the results presented for different sources, although the somewhat larger variance in the results for pink noise suggest that the familiarity of the source does indeed play a role in the perception of source distance, as mentioned in Section 2.1. Future studies will investigate the use of these monaural cues further, and will utilize 0th order sound field rendering, since it will remove the influence of any directional information. Considering the aforementioned study of Bronkhorst et al. [10], where the accuracy of distance perception increases with the number of reflections, our findings demonstrate that the net effect of the monaural cues of direct to reverberant ratio, level difference and time of arrival of early reflections are of greater importance in distance perception for loudspeaker rendering than Ambisonic directional accuracy beyond 1st order.

8 Conclusions

We have assessed through subjective analysis the perceived source distance in rendered Ambisonic sound fields in comparison to real world sources. The hypothesis tested was that enhanced directional accuracy of the deterministic part of the sound field may lead to better reconstruction of environmental depth and thus improve the perception of sound source distance. However, it was shown that Ambisonic reproduction matches the perceived real world source distances well even at 1st order and no improvement in this regard was observed when increasing the order. It must be emphasized though, that this analysis applies to Ambisonic decodes with higher order synthesis achieved using the directional analysis method of [15]. Further work will investigate the effectiveness of HOA synthesis in comparison to real world HOA measurements as well as examine this topic for off-centre listening positions.

Acknowledgments

The authors gratefully acknowledge the participation of the test subjects for both their time and constructive comments, as well as the technical support staff at the Department of Theatre, Film and Television and the Department of Electronics at the University of York for their assistance in the experimental setups. This research is supported by Science Foundation Ireland.

References

- [1] N. Tsingos, C. Chabanne, C. Robinson, and M. McCallus, "Surround Sound with Height in Games Using Dolby Pro Logic IIz," in *Audio Engineering Society Conference: 41st International Conference: Audio for Games*, 2011.
- [2] DTS, "DTS-HD high definition audio," <http://www.dts.com/dts-hd/>, 2007.
- [3] W. Van Daele, Bert; Van Baelen, "Auro 3-D Professional Workflow White Paper Rev 0.6," <http://www.auro-3d.com/wp-content/uploads/2011/06/Professional-Workflow-White-Paper-v0-6-20120228.pdf>, 2012.
- [4] M. A. Gerzon, "Periphony: With-height sound reproduction," *J. Audio Eng. Soc.*, vol. 21, pp. 2–10, 1973.
- [5] G. Kearney, "Auditory scene synthesis using virtual acoustic recording and reproduction," PhD Thesis, Trinity College Dublin, 2010.
- [6] E. Czerwinski, A. Voishvillo, S. Alexandrov, and A. Terekhov, "Propagation distortion in sound systems: Can we avoid it?" *J. Audio Eng. Soc.*, vol. 48, no. 1/2, pp. 30–48, 2000.
- [7] J. Blauert, *Communication Acoustics*. Springer, 2008.
- [8] S. H. Nielsen, "Auditory distance perception in different rooms," *J. Audio Eng. Soc.*, vol. 41, no. 10, pp. 755–770, 1993.
- [9] M. B. Gardner, "Distance estimation of 0° or apparent 0° oriented speech signals in anechoic space," *J. Acoust. Soc. Am.*, 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, 1999.
- [11] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, 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," *J. Audio Eng. Soc.*, 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 *Proc. of the 32nd Euro. Conf. on Vis. Percep.*, Regensburg, Germany, 2009.
- [14] G. Kearney, M. Gorzel, F. Boland, and H. Rice, "Distance perception in interactive virtual acoustic environments using first and higher order ambisonic soundfields," *Acta Acustica United with Acustica*, vol. 98, pp. 61–71, 2012.
- [15] J. Merimaa and V. Pulkki, "Spatial impulse response rendering i: Analysis and synthesis," *J. Audio Eng. Soc.*, vol. 53, 2005.
- [16] W. M. Hartmann, "Localization of sound in rooms," *J. Acoust. Soc. Am.*, vol. 74, no. 5, pp. 1380–1391, 1983.
- [17] C. Guastavino and B. F. G. Katz, "Perceptual evaluation of multi-dimensional spatial audio reproduction," *J. Acoust. Soc. Am.*, vol. 116, no. 2, pp. 1105–1115, 2004.
- [18] P. Zahorik, "Assessing auditory distance perception using virtual acoustics," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1832–1846, 2002.
- [19] J. M. Loomis, R. L. Klatzky, J. W. Philbeck, and R. G. Golledge, "Assessing auditory distance perception using perceptually directed action," *Perception And Psychophysics*, vol. 60, no. 6, pp. 966–980, 1998.
- [20] T. Y. Grechkin, T. D. Nguyen, J. M. Plumert, J. F. Cremer, and J. K. Kearney, "How does presentation method and measurement protocol affect distance estimation in real and virtual environments?" *ACM Trans. Appl. Percept.*, vol. 7, pp. 26:1–26:18, 2010.
- [21] G. Kearney, M. Gorzel, H. Rice, and F. Boland, "Depth perception in interactive virtual acoustic environments using higher order ambisonic soundfields," in *Proc. of the 2nd Int. Ambisonics Symp.*, Paris, France, 2010.
- [22] W. M. Fisher, G. R. Doddington, and K. M. Goudie-Marshall, "The darpa speech recognition research database: Specifications and status," in *Proc. of the DARPA Workshop on Speech Recognition*, 1986.
- [23] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. of the 108th Conv. of the Audio Eng. Soc.*, Paris, France, 2000.
- [24] M. A. Gerzon and G. J. Barton, "Ambisonic decoders for HDTV," in *Proc. of the 92nd Conv. of the Audio Eng. Soc.*, Vienna, Austria, 1992.