MODELLING THE PERFORMANCE OF SPEAKER ARRAYS IN DOMESTIC LISTENING ENVIRONMENTS

M Dring University of Derby, Derby, UK
C Middlicott University of Derby, Derby, UK
B Wiggins University of Derby, Derby, UK
A Vilkaitis University of Derby, Derby, UK

1 INTRODUCTION

The evaluation of multi-speaker arrays is best conducted in ideal anechoic conditions as to minimize distortions introduced by room reflections and therefore better enable the observation of speaker interaction at a known listening position. However, the consumption of audio from loudspeakers is not likely to be in anechoic conditions and is therefore not identifying the 'real' experience. It is most common to enjoy music in the home, perhaps in a common living space. In such spaces the sound field is non-diffuse with a dominance of the direct and early reflected energy with different interactions between the loudspeaker and room above and below a frequency transition region around 300Hz[1].

Studies have shown that any number of factors from program material, number of loudspeakers, loudspeaker position and directivity, room layout and spread of absorption will instigate both objective and subjective changes[2,3]. Making changes to a single or multiple variables in a 'real-room' can have practical limitations, with the process being time-consuming and resource heavy. In both environments the accuracy between captured data sets when repeating experiments can also be questioned due to changes in the environment, variations in the equipment used and absolute positioning. Computer modelling and simulation techniques overcome these issues and enable quick changes in variables to be administered providing opportunity for greater data sets to be generated for analysis. In addition, subjective assessment can be administered by using the binaural rendering processes embedded within the software where previous research states such techniques offer less potential for visual bias[4], the ability to achieve fast, blind comparisons of multiple listening positions or loudspeaker systems[5], whilst achieving an excellent simulation of the true sound of that space[6].

This project (which is part of a larger work) aims to develop an approach in the utilisation of acoustic simulation software and binaural rendering to validate its application in future research when studying the behaviour of sound from loudspeakers in acoustically small spaces.

2 DESIGNING THE ENVIRONMENT

2.1 Modelling the Space

When considering the acoustic space to model, BS6840-13 1998 was used to set the design goals of a typical residential listening environment. By following this standard, the room dimensions (L7m, W5.3m, and H2.7m), reverberation time limits (see Figure 2), environmental conditions and background noise levels were adhered to. A single room was modelled at this stage as to limit the scope of the investigation in order to better analyse the impact of making changes to the speaker arrays. Variation in room layout, object placement and acoustic conditioning between residential dwellings can be implemented with relative ease in future work.

The room and objects were created using Google Sketchup, which enabled the creation of more complex shapes and placement of items (i.e. furniture) to be completed with greater precision. The

inclusion of furniture was justified to represent typical regional practices[7] which will subsequently generate variations in the reflected and scattered energy throughout the room.
A listening position at the centre of the room was chosen to allow the proposed speaker arrays to be positioned within the boundary limited space and equidistant from the focal point, eliminating variations in the direct sound arrival time. The expectation is therefore that the proximity of possible speaker positions and furniture will impact upon the early reflections which along with the direct sound are known to dominate in small rooms[3]. Figure 1 shows the final version of the model in Google Sketchup.

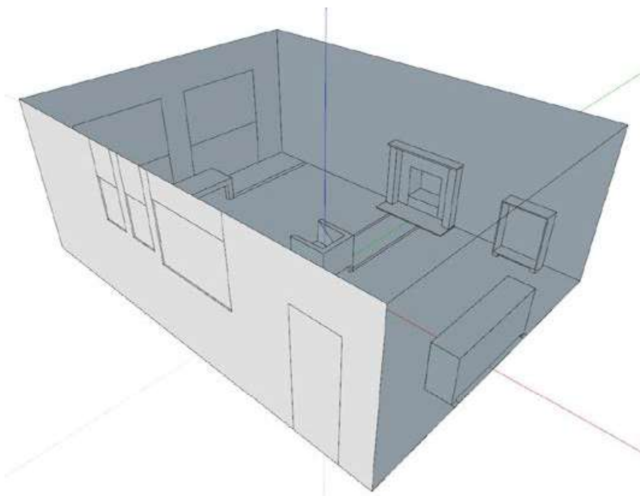

Figure 1 - Modelled domestic listening space

To conduct acoustic simulations, the model was imported into EASE™. The materials allocated to the faces were informed by the recommendations in BS6840-13 1998. Based on the materials selected the software calculated a reverberation time as shown in Figure 2, this is presented against the tolerance limits advised in the standard. For evaluation purposes our interest in reverberation time as an objective measurement is of less significance in this type of space, where it is considered less meaningful to characterise the room[8].

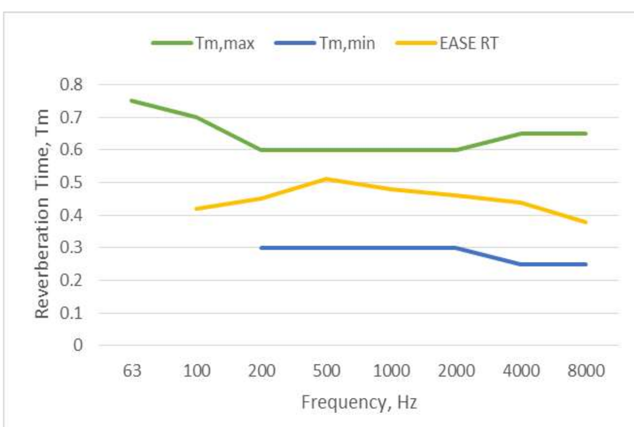

Figure 2 - Reverberation time generated in EASE™ against BS6840-13:1998 tolerance limits

2.2 The Loudspeaker Array

The interaction between the room and loudspeaker as a 'system' will change as a result of the type of loudspeaker and by extension, its performance characteristics, and also the position of the loudspeaker relative to the room[2,9].

Of interest to this study was the performance of horizontal Ambisonic loudspeaker arrays in the domestic environment, limiting the evaluation to 1st, 2nd and 3rd order regular shaped horizontal (2 dimensional) Ambisonic arrays within the modelled room. Following the theories relating accurate sound field reproduction and the number of speakers[10]; the 1st order arrangement will utilise 4 speakers, 2nd order 6 speakers and 3rd order 8 speakers. It is anticipated that the higher order array will better reproduce the sound field where improvements in interaural cues are observed[11], however, these studies document this behaviour in anechoic spaces.

Using the scene based Ambisonics system, a far field sound source from any direction can be constructed as a summation of plane waves from a number of set directions via the Spherical Harmonic Transform (spatially sampled, reproduced using loudspeakers) allowing for a certain order of Ambisonics to be achieved (where 4 speakers are needed for 1st order, 6 for 2nd and 8 for third order). As the speakers will not be in the far field, then the sound source will be reproduced as if at the same distance as the loudspeakers. The generation of Ambisonic decoders for regular arrays is well documented, with mode-matching, velocity vector optimised, decoders used in this work, calculated using the pseudo-inverse of the Spherical Harmonic Transform of the loudspeaker array positions. As only horizontal arrays are simulated, the system is simpler than the full 3D case, with the spherical harmonics from 1st to 3rd order shown below in Figure 3.

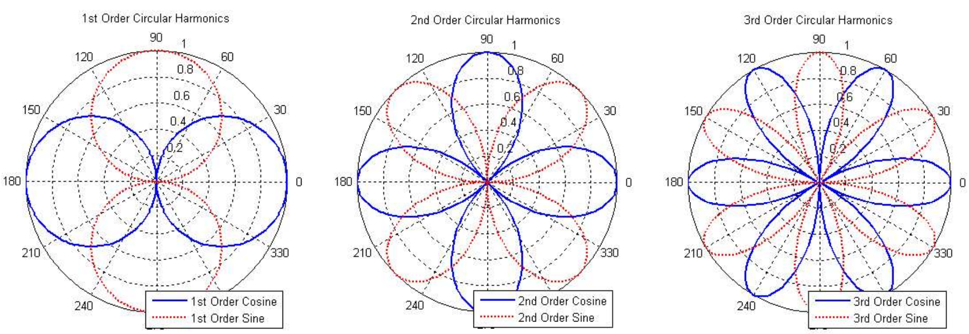

Figure 3 - 1st (left), 2nd (middle) and 3rd (right) order harmonics

Truncating the system to a specified order in the spherical harmonic domain, will affect the accuracy of the binaural cues reproduced by the system (whether auditioned over loudspeakers, or binaurally over headphones). It is shown that inter-aural level and time differences (ILD and ITD) will be correctly reproduced up to the spatial aliasing frequency which will increase with Ambisonic order[11]. For example, the error in dB between the ILD of a real and Ambisonically reproduced source can be seen for 1st, 2nd and 3rd order arrays (anechoic) in Figure 4 where white indicates no difference.

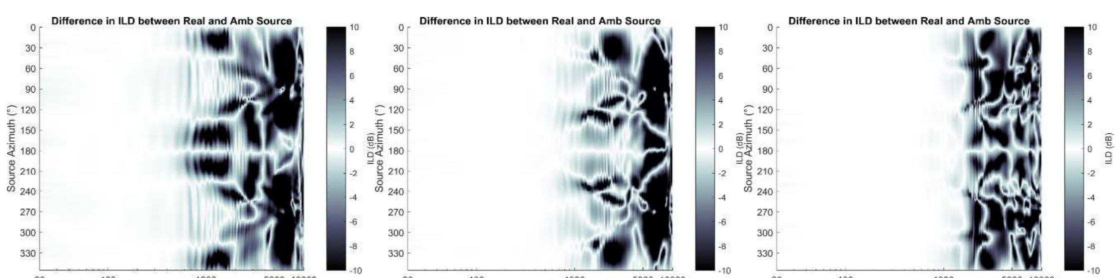

Figure 4 – ILD difference between real and Ambisonic source for 1st (left), 2nd (middle) and 3rd (right) order reproduction

2.2.1 Loudspeaker Selection

In order to more realistically simulate a real listening environment, it's also desirable to model the directional characteristics of real loudspeakers. EASE™ provides a database of loudspeakers for simulation purposes; this project used the JBL Control 25, justified on the basis that it is a compact loudspeaker most akin to typical loudspeakers found within domestic environments for audio playback. Figure 5 shows the loudspeaker directivity at 250Hz and 8kHz, indicating omnidirectional output up to 500Hz and a narrowing at higher frequencies to a cardioid pattern.

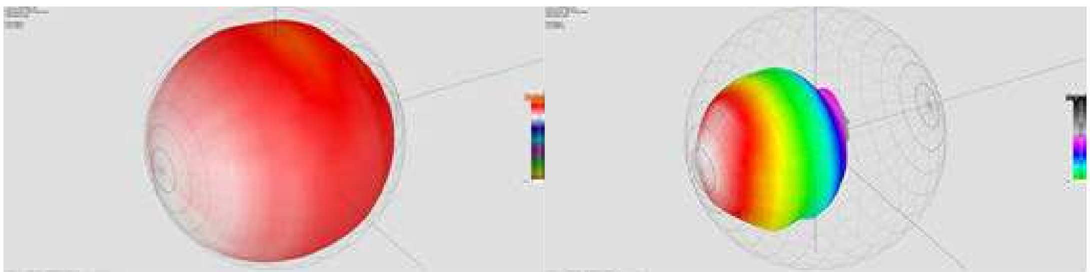

Figure 5 - JBL Control 25 attenuation balloon at 250Hz (left) and 8kHz (right)

Simulation of low frequency modes are not taken into account using the EASE™ software[12] as modelling is based on ray-tracing principles; the accuracy in auralising frequencies below the transition region (<300Hz) will therefore be affected and is suitable for investigation in further work. Research has identified that modal behaviour within a small space can be modelled with Boundary Element Methods (BEM), showing good similarity when compared with measurements of a space[13]. It has also been presented that modal excitation within small spaces is smoothed out through considered placement of dedicated low-frequency (subwoofer) speakers[1], therefore supporting an argument whereby observation of the processes employed within this paper should only be considered valid above a specific frequency; future testing is proposed to validate this concept.

2.2.2 Loudspeaker Positioning

Figure 6 show a plan view of the space and the 1st, 2nd and 3rd order arrays respectively. Logically the front speakers were offset from 0 (zero) degrees as would be dictated in a 'real' situation. To maintain the regular shape of the array the 2nd and 3rd order arrangements were limited to a radius of 2.5m to fit within the room width and not overlap with objects already positioned within the model. Speaker heights were set to 1.2m (ear height when seated) and orientated towards the listening position.

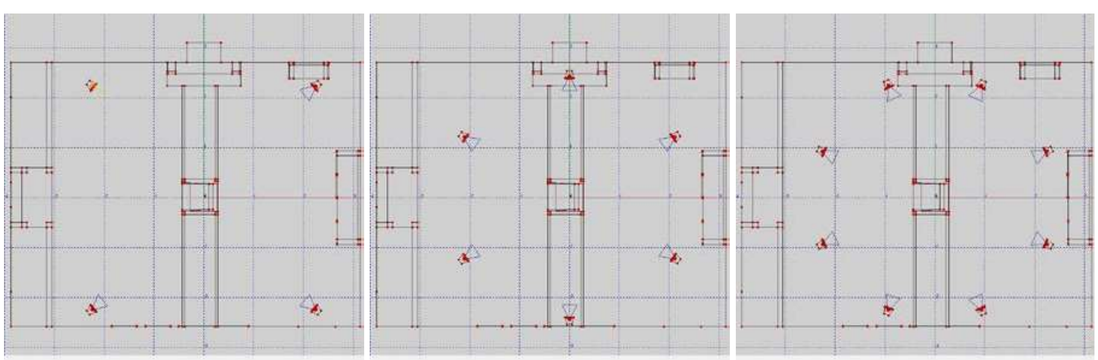

Figure 6 - Positions of the 1st (left), 2nd (middle) and 3rd (right) order loudspeaker arrays within the room

3 AURALISING THE ARRAYS

Auralising the system allowed for a mathematically accurate binaural simulation to be obtained simulating the sound at the two ears of a centrally seated listener.

Capturing an impulse response ensures that all individual components from source to receiver that are used within the system are represented; this concept doesn't change between real or simulated spaces. By convolving anechoic material with an impulse response we are able to 'place' the listener in that space and effectively perceive any audio 'out of the head'. Figure 7 demonstrates the process applied for a 1st order array.

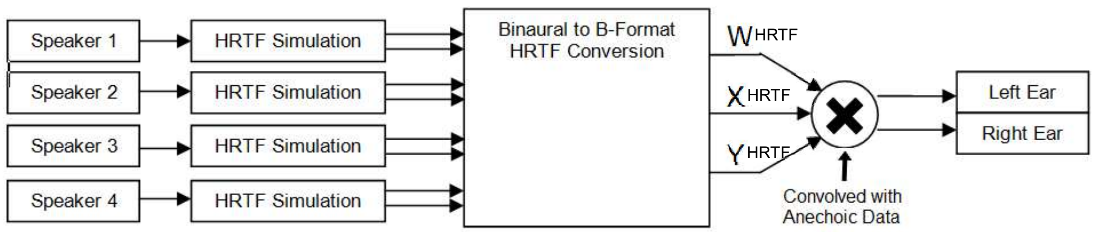

Figure 7 - The conversion process for the 1st order loudspeaker array

A series of binaural room impulses responses (BRIR) were generated within EASE using the mirror image impact ray tracing method for each speaker within the array. The mirror image method was chosen as it determines and computes all the impacts on a specified point (i.e. receiver position) up to a chosen time limit or reflection order[14]. Auralisation using AURA Response was available but it was felt based on the size of the model that the statistical mapping was sufficient and representative. Each BRIR is therefore the product of the speaker response, room reflections and dummy head head related transfer function (HRTF) for that position.

In order to simulate the Ambisonic loudspeaker array in the room, the loudspeaker outputs can simply be convolved with the corresponding loudspeaker HRIR. However, one efficiency saving that can be implemented is to roll up the Ambisonic decoder and HRIRs into a single operation, meaning that, instead of a pair of HRIRs per loudspeakers, one pair of HRIRs per Ambisonic channel are needed. For 1st order, this means 3 pairs of HRIR are needed (no matter how many speakers are modelled), 2nd order uses 5 pairs and 3rd order 7 pairs[11]. Auditory cues relating to amplitude and time differences along with head filtering are already embedded into the initial BRIRs and therefore this data is retained during conversion. With horizontal only arrays the script omitted code to calculate spherical harmonics that represent height.

In order to generate material for subjective testing, the B-Format IR's were convolved with anechoic material in Reaper and the output is again delivered in a binaural format.

4 TESTING THE SYSTEM

4.1 The Listening Test

A listening test was designed to subjectively analyse the performance of the simulated arrays delivered binaurally over headphones within the modelled space. It was decided that source localisation would be the parameter to observe, confirming if the process has the ability to deliver audio from pre-determined angles and also if the Ambisonic order and room response made a difference to the accuracy when determining source position. A total of 23 participants completed the test with the majority of those being unfamiliar with these practices.

4.1.1 Program Material

Two different anechoic test signals of a maximum duration of 20 seconds were selected for the listening tests; a time frame greater than that proposed by research in binaural audio presentation[15]. The change in source material is intended to excite the 'system' in different ways generating further data for evaluation. The signals used were:

- Trumpet, taken from the Bang and Olufsen CD made for the Archimedes project[16]. This material contains mid-high frequencies with a mixture of transient attacks and sustains[17].
- Male Voice, taken from the EASE Software. This material contains low-mid frequencies spoken at a consistent rate.

4.1.2 Orders and Position

6 (six) positions around the 360 degree horizontal array were chosen for each Ambisonic order (1st, 2nd and 3rd). The positon of the audio does not need to correlate to the known locations of the speakers; under anechoic conditions, as the Ambisonic order (and number of speakers) increases, it is known auditory cues up to higher frequencies are improved as more energy will be coming from the correct direction[11], therefore improvements in localisation will be expected. However, presenting audio from different locations within a room, the IR's will contain variations in spatial and timbre content that may impact on localisation.

4.1.3 Anechoic Conditions

An additional 36 test files that mimic the positions and sources of the test files mentioned above with the variation that the simulated room consisted only of materials with 100% absorption, therefore eliminating energy upon the first reflection. This was considered an appropriate method to simulate anechoic room conditions maintaining source-receiver level and position and utilising the same HRTF set in all test audio.

4.1.4 The Graphical User Interface (GUI)

A GUI was created, where the listener must specify the angle at which they perceive the sound to be positioned around the 360 degree azimuth, as shown in Figure 8. The audio clips are selected at random from the 72 files discussed previously, with the requirement to judge a minimum of 18 before the test is completed. A repeat function was included allowing unrestricted opportunities to listen to each audio clip. The limitations of the interface were such that the user was unable to place the puck closer or further away from the 'head', restricting decisions to source position rather than specific location.

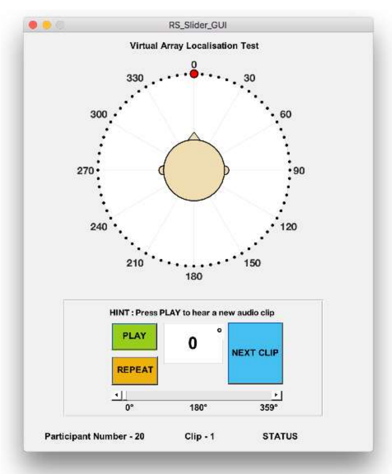

Figure 8 - The graphical user interface

Each listener conducted a short training period prior to experiencing the audio clips that did not contribute towards the results. The training period played back 4 (four) clips at angles of 0, 90, 180, 270 degrees with this knowledge divulged to the listener for the purpose of setting auditory points of reference and awareness of user/software interaction.

5 RESULTS AND DISCUSSION

5.1 Trumpet as Source Material

Figure 9 plots the selected position error data against the known source position for all conditions and Ambisonic orders when using the trumpet as the source material. Representing the data using box and whisker plots allows for observation of the localisation accuracy across all the tested data for all participants. The black dotted line marks the position of 0 degrees of error (i.e. participants have been 100% precise when selecting source position).

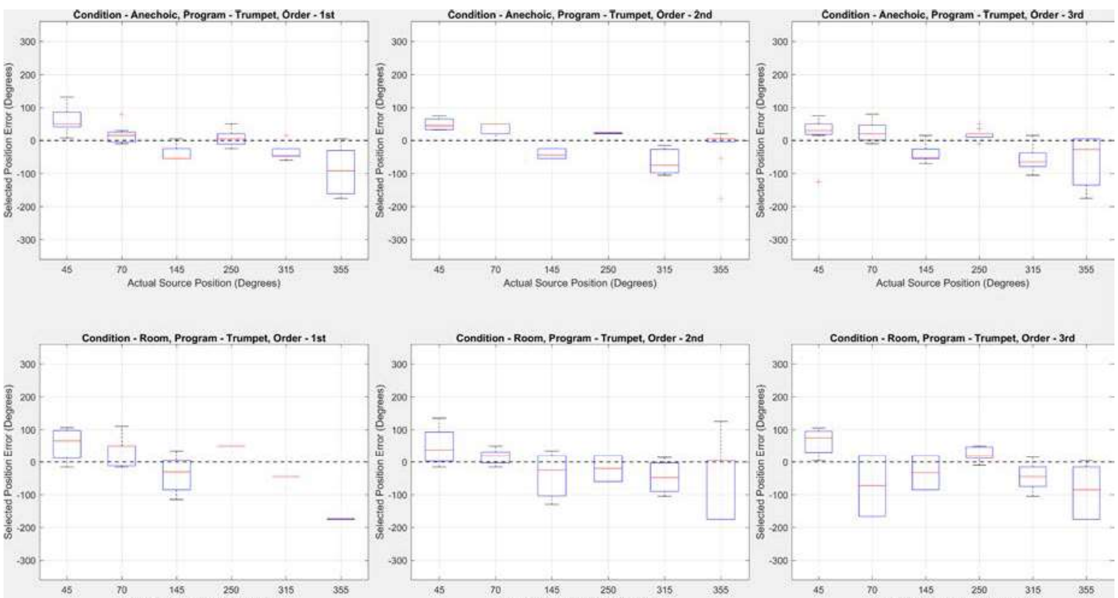

Figure 9 - Localisation error using trumpet as source material

Where the median (red line) is closest to the black dotted lines and length of the box is most compact, this is an indicator that for the occasions those variables were tested the participants have been more accurate; therefore the listeners were able to use the variables associated with those audio events to influence more precision in their choice of source position.

Observing the data in Figure 9 it is evident that localisation errors have occurred for all tested positions. This evidence suggests that for all conditions and orders, sound presented at 355 degrees resulted in the greatest errors, with some data points recording a maximum of 180 degrees in error.

Analysis of variance (ANOVA) using MATLAB was conducted on the data sets for observation of statistically different means between the varying factors applied within the test. For this project, the output of the ANOVA shows if the localisation errors reported by the listener are significantly different, which can be observed within a group (i.e. change in source position), or across multiple factors (i.e. condition and order), outputted by the process as a level of confidence (*p-value*).

Conducting one-way ANOVA for each of the groupings shown in Figure 9 it is confirmed that 4 (four) out of the 6 (six) data sets, 355 degrees is known to have the greatest mean localisation error. For the other two cases this event occurs at 45 degrees and at 315 degrees. These results highlight that for sound arriving from angles close to the front (0 degrees) the listeners had difficulty in localising its true position. For source positions presented with the room condition from a 2nd order array, the results show no significant difference in the localisation errors between the groupings, this is unlike the other sets; this perhaps indicates more favourable interactions for perceived source location.

Two-way ANOVA was used to observe how the outcomes from an individual source position compared between a change in the order, a change in the room condition and also between those two variables. Notably, when sound arrives from 70 degrees the data shows that both the condition and order have an effect. With the room condition applied from a 3rd order array the analysis shows this to be significantly different to the other groupings, therefore the room and order have altered the ability to localise the source position in a way that wasn't apparent when presented in anechoic conditions and also from the other array types. The analysis also shows the order of the array makes a significant difference on the sounds presented from 355 degrees; the 1st and 2nd order arrays do not provide comparable localisation errors, highlighting the 2nd order array in both room and anechoic conditions are more favourable when identifying source location.

5.2 Voice as Source Material

Figure 10 plots the selected position error data against the known source position for all conditions and Ambisonic orders when using the voice as the source material. On initial observation of the plots in Figure 10 the evidence suggests that the greatest inaccuracies are associated with the arrival of the sound from 170 degrees, where the box and whisker plots at this position are longer and the median results (red line) are further from the ideal (black dotted line).

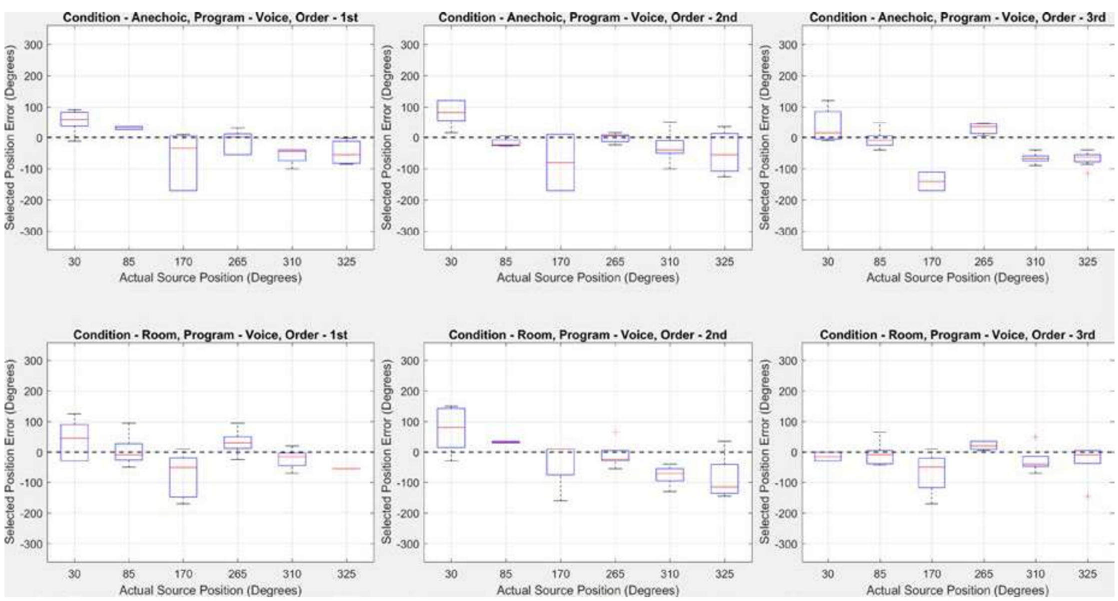

Figure 10 - Localisation error using voice at source material

Conducting one-way ANOVA for each of the groupings shown in Figure 10 it is confirmed that with the exception of the results with the room and 2nd order array, 170 degrees shows the greatest mean error; confirming the statement made earlier in the report. Room and 2nd order conditions indicate that the greatest mean error occurs when presenting sound from 325 degrees. It is observed these values are at positions closest to the front and rear of the listener of those tested. Across the 6 (six) data sets shown in Figure 10, all 3 (three) anechoic results report that audio material presented from 30 degrees was the most significantly different grouping, therefore the localisation errors recorded were least likely to be made when compared to the other angles; although this does not state if the difference was an improvement. This was also the case for room

condition, 2nd order, however the other 2 (two) room conditions with 1st and 3rd order arrays show that sounds arriving from 170 degrees as the significantly different data set.
Two-way ANOVA was used to observe how the outcomes from an individual source position compared between a change in the order, a change in the condition and between those two variables. The only results of note occur at 30 degrees where the difference between room and anechoic conditions are not significant; however the 3rd order data is shown to be significantly different from 1st order with the results showing an improvement in accuracy for the 3rd order array.

5.3 Discussion

From the analysis carried out for both the trumpet and voice material, it is evident that neither an increase in horizontal Ambisonic order nor an inclusion of the simulated room environment that the listeners were able to localise the source position with greater accuracy. Where two-way ANOVA has identified differences when changing the order for certain source positions (i.e. Trumpet: 355 degrees, Voice: 30 degrees), the trend in the data is not present and a greater number of participants would be required to make a conclusive judgement.

Most evident from the data is that as material is presented at positions closer to the front or rear of the listener, this promotes an increase in localisation error; this is not significantly improved with an increase in the order or between anechoic or room conditions. Research into binaural presentation of audio suggests that to improve localisation accuracy the listener requires the ability to move their head in relation to the sound field, as would occur in the natural environment, where sound presented either from the front or the back see the greatest improvement. Anecdotally, when receiving comments from the listeners many reported the confusion between the front and the back. Some test participants reported the clear difference between sounds that felt closer to the head and those which were presented at a distance; it is therefore considered that externalisation is readily perceived by the listener above the ability to localise the source. Two of the participants made comments that some of the audio clips appeared to arrive from an elevated position; this sensation would agree with research stating this effect can be caused from non-individualised binaural recordings[18].

The possibility that a change in Ambisonic order creates a ‘system’ change that is too subtle must also be considered. The analysis has shown that in certain data sets the Ambisonic order is attributed to variations; although a trend isn’t observed, the listeners have been responsive enough to recognise a change namely for 2nd order arrays; an increase in test participants would provide further validation.

6 CONCLUSION

Testing lateralisation accuracy, alone, does not seem to illicit statistically significant results between different Ambisonic orders of binaural reproduction or anechoic vs room modelled reproduction. Anecdotal evidence does suggest that not only do the systems sound different, but externalisation is present once the room response is introduced, although it’s interesting to note that this, on its own, doesn’t seem to improve front/back reversals for sources at or around 0 and 180 degrees.

6.1 Further Work

The experience gained from developing this methodology, compiling the data and completing the report has provided plenty of points for consideration in future work with this project being the baseline for future testing and development. Although the project hasn’t provided the expected results, with respect to increased accuracy with increased Ambisonic order, it is yet to be explored if the equivalent ‘real’ space would generate similar results, which is clearly important in the validation of this, and future work. Regarding the binaural system used in the testing, the project strives towards the use of low latency head-tracking techniques along with a higher resolution model of the room; where the speakers and room responses are modelled at 35th order (meaning ILD and ITD cues, and the HRTF data in general, will be correct above 10kHz) allowing for smooth head rotation to be formed accurately which can be modelled in EASE™.

7 REFERENCES

1. F.E. Toole, *"Loudspeakers and Rooms for Sound Reproduction - A Scientific Review",* Journal of the Audio Engineering Society, Vol.54 [6], 2006.
2. S. Olive and W. Martins, *"Interaction between Loudspeakers and Room Acoustics Influences Loudspeaker Preferences in Multichannel Audio Reproduction",* AES 123rd Convention, New York, 2007.
3. S. Linkwitz, *"The Magic in 2-Channel Sound Reproduction – Why is it so Rarely Heard?"*, [Online] Available at: http://www.linkwitzlab.com/The_Magic/The_Magic.htm, (Accessed: 27/09/17)
4. P. Schuck, S. Olive, et al., *"Perception of Perceived Sound in Rooms: Some Results of the Athena Project"*, AES 12th International Conference, Copenhagen, 1993.
5. D. Satongar, *"Simulation and Analysis of Spatial Audio Reproduction and Listening Area Effects"*, Ph.D. Thesis, University of Salford, 2016.
6. A. McKeag, D. McGrath, *"Using Auralisation Techniques to Render 5.1 Surround to Binaural and Transaural Playback"*, AES 102nd Convention, Munich, 1997.
7. British Standards Institution, *"BS6840-13: 1998, Sound System Equipment – Part 13: Listening Tests on Loudspeakers"*, London: BSI, 1999.
8. H. Kuttruff, *"Sound Fields in Small Rooms"*, AES 15th International Conference, Copenhagen, 1998.
9. D. Griesinger*, "Spaciousness and Localization in Listening Rooms and Their Effects on the Recording Technique"*, Journal of the Audio Engineering Society, Vol.34 [4], 1986.
10. B. Wiggins, *"An Investigation into the Real-Time Manipulation and Control of Three-Dimensional Sound Fields"*, Ph.D. Thesis, University of Derby, 2004.
11. B. Wiggins, *"Analysis of Binaural Cue Matching using Ambisonics to Binaural Decoding Techniques"*, 4th International Conference on Spatial Audio, Graz, Austria, 2017.
12. AFMG Technologies GmbH, *"EASE Parameters in Aura"*, [Online] Available at: http://ease.afmg.eu/index.php/ea-faq-all-en/items/Parameters_in_AURA.html, (Accessed: 27/09/17)
13. R. Cipriano, R. Hersberger, et al. *"Low Frequency Behavior of Small Rooms"*, AES 139th Convention, New York, 2015.
14. AFMG Technologies GmbH, *"EASE Version 4.3 User's Guide & Tutorial"*, [Online] Available at: http://www.afmg-support.de/SoftwareDownloadBase/AFMG/EASE/EASE_4.3_Tutorial_English.pdf (Accessed: 20/08/17)
15. J.S. Calle, A. Roginska, "Head Rotation Data Extraction From Virtual Reality Gameplay Using Non-Individualised HRTFs", AES 143rd Convention, New York, 2017.
16. Band and Olufsen, *"Music for Archimedes",* [CD] Track 36, 1992.
17. R. Mason, F. Rumsey, "An Assessment of the Spatial Performance of Virtual Home Theatre Algorithms by Subjective and Objective Methods", AES 108th Convention, Paris, 2000.
18. H. Moller, M.F. Sorensen, et al. "Binaural Technique: Do We Need Individual Recordings?", Journal of the Audio Engineering Society, Vol.44 [6], 1996.