THE PSYCHOACOUSTIC EFFECTS OF STIMULI PLAUSIBILITY ON HEADPHONE EXTERNALISATION

B. Fazenda Acoustics Research Centre, University of Salford, UK
J. Newton Acoustics Research Centre, University of Salford, UK

1 INTRODUCTION AND LITERATURE REVIEW

In the world we live in, sound events are typically associated with a physical event happening within our surroundings. In this case, the event carries a degree of plausibility in that it exists in the physical world and should therefore have associated with it all the aural cues that allow our hearing system to place it somewhere in the space around us. Throughout life, one's ability to locate such sounds is learnt through incorporating Interaural Time Differences (ILDs), Interaural Level Differences (ILDs), Head Related Transfer Functions and Room Reverberation of the signals arriving at our ears[1].

The field of spatial hearing is well established and the importance of the various acoustic cues for realistic auditory perception has been researched in considerable depth[1,2,3,4,5,6,7,8]. When listening on headphones, these perceptual features are not present, causing sounds to have an apparent location inside the head. Much research has been devoted to counteract this effect by processing headphone audio in such a way that it contains a form of the features which are usually present in everyday listening, and thus evoke out-of-head localization or *externalisation*[9,10,11,12,13].

In the area of Human perception, seminal studies have demonstrated that there exists a multimodal integration between the senses of sight and sound[14,15] that is crucial for our sense of reality. Despite this, there seems to be little attention given to the aspects of expectation and plausibility of stimuli, in eliciting out-of-head effects when the sound events are presented over headphones.

A previous study presented at RS 2013 looked at factors contributing to the percept of externalization over headphones[16]. It was found that correctly modeled binaural cues could elicit the effect of externalization. However, an interesting find revealed that the plausibility of a stimuli existing in the space where the audition was taking place was also a significant effect in creating the percept of externalization.

In the study presented here, this effect has been researched further using a novel experimental methodology. In this case, a number of sound events have been enacted within the room where a subject (listener) is engaged in a task (playing a computer game). The subject was asked to pay attention to the events taking place, mainly by listening to the sounds of those events. The plausible sound sources in the room were acted silently (by mimicking) whilst their sounds were replayed solely over the headphones. This was carried out with 2 listener groups, each of them being played the sound events with or without binaural cues. The sound events were plausible in the sense that they were associated with objects placed and moved around the room. Additional sound events, used as experimental controls, were played over the headphones but these did not have their plausible physical counterparts in the room. The perception of externalisation was assessed through a questionnaire at the end of the test where subjects were required to indicate whether they heard the sounds within the headphones or in the room and, in the latter case, where the sound event was located.

The study thus poses the following research question:

What is the relative importance between sound event plausibility and correct binaural cues in eliciting the percept of externalization when listening on headphones?

Our motivation is to find evidence that may be used by producers and sound designers concerned with the creation of believable audio scenes in 3 dimensional audio rendering systems.

2 TEST METHODOLOGY AND EXPERIMENTAL DESIGN

Subjects were invited to participate in an experiment reported to test their attention capabilities. However, in reality, the experiment was designed to assess the effect of the presence of a plausible sound event on the perception of headphone externalisation. Recorded samples were modified into test stimuli using a binaural auralisation model. The auralisation model used in the experiment has been described in[16]. Stimuli with and without binaural cues were then used in the experiment to test the effect of the plausibility of a sound on externalisation.

2.1 Recording

The sound events had to occur plausibly in the testing room. Therefore, the sounds selected were part of everyday type sounds that could occur in either domestic or working environments. The samples produced were a *mobile phone ringing*, *the sound of a mug being placed on a table*, *the sorting and placing of a ream of paper*, and *the sound of finger joints 'cracking'*. These were chosen due to the impulsive nature, and the ease of which the sound could be 'acted out' and thus appear 'plausible'. Non-plausible sounds, used as an experimental control, were selected as sounds not normally heard in a testing room. The sounds chosen were that of a car horn, a bike bell, a heavy wooden door opening and a microphone stand falling over.

As all room cues, including air absorption were to be modeled and added to the dry sample, it was vital the recordings contained minimal room effects such as reverberation and room colouration. Therefore, where possible, the original sound samples were recorded in anechoic conditions.

2.2 Methodology

The test was designed with an emphasis on gathering a large amount of data per subject in order to obtain a high level of statistical power; this was achieved by creating a test which collected 8 answers from each subject, consisting of 4 'plausible' and 4 'not plausible' sound events. All test stimuli were played through the headphones and no sounds actually existed in the room.

Test stimuli was divided into two categories: 1) stimuli with binaural cues and; stimuli without binaural cues. The stimuli was further subdivided into 'plausible' and 'not plausible' referring to whether the auditioned stimuli was accompanied by an action taking place in the room or not.
Each sound is thus modeled for each of the categories and subcategories within:

- Plausible with Binaural Cues
- Plausible without Binaural Cues
- Not Plausible with Binaural Cues
- Not Plausible without Binaural Cues

A group of subjects was tested with binaural cues implemented in their sound events whilst a second group was tested with stimuli that had no binaural cues implemented.

2.3 Test Set-up

Subjects were brought into a room and asked to sit at a desk and play a computer game. The game was simple enough for every subject to be able to complete it, yet absorbing enough to keep the player focus and gaze on the screen. A simple 'snake game' was designed in which subjects controlled a snake object to collect items with a scoring system, as shown in Figure 1. Subjects were told their performance would be analysed.

Upon entering the testing room, subjects were made aware of the objects used for the plausible samples. This was accomplished by the ‘actor’ obviously clearing the desk at which the test was to be carried out at. The phone was moved to a table at their back right, the ream of paper placed to their left and the mug moved opposite them (see Figure 2). Before the test was started, the ‘actor’ cracked their joints to make subjects aware of the sound. During the test, subjects were asked to play a computer game. This was used to stop subjects expecting sounds, and thus ‘over analyse’ their answers, saying where they thought the sound was located, as opposed to where they perceived a sound to be located. The distraction also allowed for any slight miss-timings between the actions in the room and the sound being played to the subject over headphones, which could occur due to a miscalculation of the timing of the action.

Throughout the test, the plausible sounds were acted out silently in their simulated locations. In order to make the actions silent, a padded base was added to the mug, and the ream of paper was placed down onto a soft surface. First the ringing phone sound was played to the subject, to which the ‘actor’ reacted by looking over at where the phone was placed (Figure 2(1)). The ‘actor’ then placed the mug, which they held in their hand, on the table (Figure 2(2)) and walked over to the phone. After ‘rejecting the call’ on the phone, the ‘actor’ walked over to the ream of paper next to the subject, picked it up, flicked through it, and placed it on a table to the back left of the subject (Figure 2(3)). Finally, the ‘actor’ walked to the front right of the subject, and acted out cracking their fingers (Figure 2(4)). These mimes were synced with the audio played to the subject through the use of a secondary screen placed directly behind and not visible to the subject, with the track waveform and a timer.

The subjective test required a method to confirm whether a subject experienced a sound played to them as externalised. This was done in a two tiered approach: At the end of the test, each subject was shown a list of sound events that had been tested (including the non-plausible group of sounds); for each sound event the subject was asked whether the sound appeared to be played in the headphones or in the room; for every sound event perceived in the room, the subject was further asked to indicate which position in the room did they hear the sound from - by analysing a subject’s localization of a particular sound event. If a subject experienced the sound event as coming from the direction of the plausible sound source then it was considered that the subject perceived the sound event as externalised.

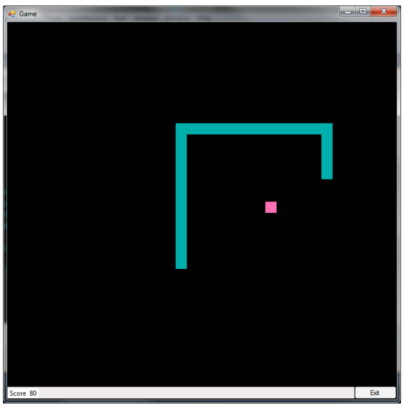

Figure 1 – Snake game used to keep subject’s gaze and attention on the computer screen n front of them. Subject played the game whilst wearing headphones through which the test stimuli was being played.

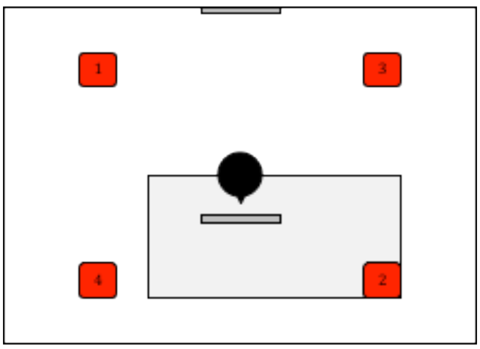

Figure 2 – Testing room scenario. The location of the ‘plausible’ sound sources is indicated with numbers. The subject’s computer screen as well as the actor’s cue screen used to sync the plausible actions are shown.

20 subjects took part in the study. Subjects were divided into two groups: Group 1 took the test with ‘plausible with cues’ and ‘not plausible without cues’ stimuli; Group 2 took the test with ‘plausible without cues’ and ‘not plausible with cues’. With this arrangement, each group has been exposed to both variables, i.e. ‘plausible’ vs ‘not plausible’ and ‘with cues’ vs ‘without cues’. Each subject was asked to indicate their perceived externalisation of 4 sounds. The total number of data points collected, N, is 160.

3 RESULTS AND DISCUSSION

3.1 Subject Reactions

Firstly, subject’s reactions to sounds were observed and noted during the experiment. Upon hearing earlier stimuli with no spatial cues, the majority of subjects would look up from the computer, often looking at the experimenter with a confused expression. Later in the experiment, subjects would have little to no reaction to hearing stimuli without cues. Upon hearing stimuli with spatial cues, subjects would look at the location of the apparent sound, either moving their entire head to look at the apparent sound source, or just their eyes with a small head movement. A number of subjects questioned whether the experiment should continue after hearing a phone ringing.

3.2 Data Analysis

Data was classified in terms of successful (“Yes”) or unsuccesful (“No”) externalisation according to subject’s responses. If a subject indicated the sound as being perceived on the headphones then no externalisation was achieved. If a subject indicated that the sound was perceived in the room from the correct direction then externalisation was achieved. A report of a sound in the room from an incorrect direction was considered as an unsuccessful externalisation.

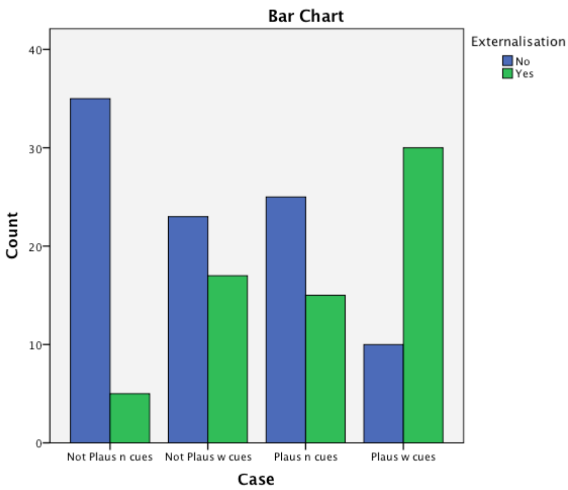

Figure 3 - Frequencies of successful and unsuccessful externalisation for all cases.

Figure 3 shows the number of successfully and unsuccessfully externalised sounds across all cases tested. The cases are labelled as *Not plausible without cues*, *Not plausible with cues*, *Plausible without cues* and *Plausible with cues*. The *Plausible* variable refers to whether the sound event had been enacted within the room during the test and the *cues* variable refers to whether the sound event had been rendered over headphones with auralisation cues. No sounds actually existed in the room as they were all rendered to the subject over headphones regardless of whether they are classified as plausible or not plausible.

The data shows that the number of successful externalisations is larger for the case where a plausible sound with auralisation cues is presented. A Pearson Chi- Square test reveals that the effect of *case* is significant with a moderate effect size, suggesting that the variables for *Plausibility* and *Cues* appear to be responsible for the effect of externalisation (*N=160,* X^2*=32.53, df=3, p=0.000, Cramer's V=0.451*).

One sample binomial tests were run for each case, where the null hypothesis, H_0, was stated as *the probabilities of successfully (Yes) and unsuccessfully externalised (No) are 0.5.* This provides a measure of whether the difference between the number of *Yes* and *No* is statistically above chance.

Case	p	Decision
Not Plausible without cues	0.000	Reject H_0
Not Plausible with cues	0.429	Retain H_0
Plausible without cues	0.155	Retain H_0
Plausible with cues	0.003	Reject H_0

Table 1 – One sample binomial tests for each case in Figure 3

Table 1 confirms that cases are significant for *Not Plausible without cues* where externalisation is not successful and *Plausible with cues* where externalisation is successful. The results clearly indicate that the presence of correct auralisation cues on the rendering of plausible sounds will elicit the percept of a sound event which is external to the headphones. Conversely, if the sounds being rendered are not plausible and the rendering system does not use auralisation cues, the percept reverts to the well known in-head localisation. Perhaps more interestingly, the existence of either plausibility or auralisation cues on their own are not sufficient to elicit a clear percept of externalisation across the group. It is only when the two perceptual aspects are combined that the externalisation effect is significant.

Further tests were carried out by collapsing data into one of the tested variables (Plausible or Cues). Figure 4 shows the effect of *Plausibility* on externalisation after collapsing the data across the auralisation cue cases. A Chi Square test reveals that plausibility of a sound event is significant in eliciting externalisation with a moderate effect size (*N=160,* X^2*=13.58, df=1, p=0.000, Cramer's V=0.3*).

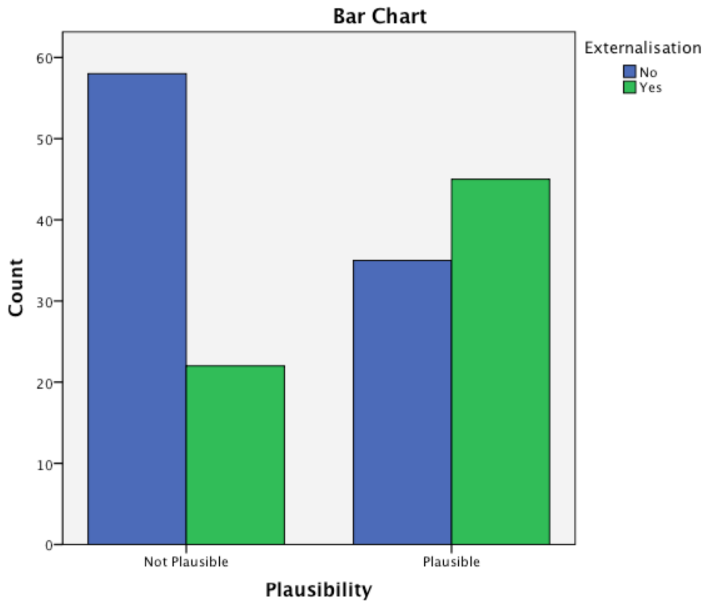

Figure 4 – Effect of plausibility on externalization disregarding auralisation cues.

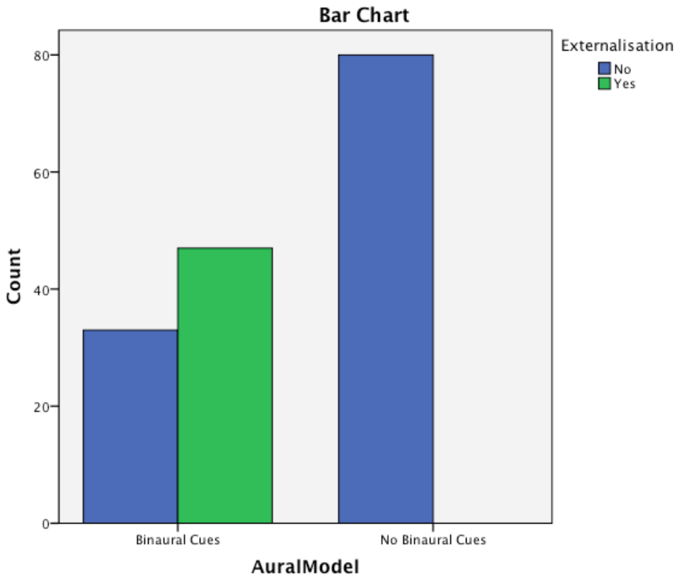

Figure 5 – Effect of auralisation cues on externalization disregarding plausibility.

Collapsing the data across plausibility and testing for the effect of auralisation cues, Figure 5 shows that in the absence of auralisation cues, there is no elicitation for the percept of externalisation. This is a statistically significant effect with a moderate effect size, albeit larger than the one measured for the plausibility cues (*N=160, X^2=66.54, df=1, p=0.000, Cramer's V=0.6*).

4 CONCLUSIONS

The significance of stimuli plausibility on externalisation was assessed and compared to the significance of binaural cues. It was found that the plausibility of a stimuli has a strong positive effect on reports of externalisation, though this effect was found to be dependent on the presence of binaural cues. Moderate effect sizes have been measured for both plausibility and binaural cues, although the latter was measured as twice as large, indicating that binaural cues remain the most important factor of the two in eliciting externalisation. The existence of stimuli plausibility (or binaural cues) on its own does not appear to be significant in unequivocally eliciting externalisation. A combination of the two variables appears to be required for a convincing effect.

This study further advances our understanding on effecting convincing externalisation over headphones and what aspects, beyond the usual physical reproduction accuracy, play a part in achieving it. This may be used by producers and sound designers in their quest to achieve believable audio scenes for the general user.

REFERENCES

1 Blauert, J. (1997). Spatial hearing: the psychophysics of human sound localization. MIT press.

2 Shinn-Cunningham, B. G. Distance cues for auditory space. IEEE, Special edition on virtual auditory space, 227-30, 2000.

3 Brungart, D. S. and Rabinowitz, W. M. Auditory localisation of nearby sources. Head related transfer functions. Journal of the Acoustical Society of America. 106(3), 1465–79, 1999.

4 Coleman, P. Dual role of frequency spectrum in the determination of auditory distance. Journal of the Acoustic Society of America, 4 (2), 631-2, 1968.

5 Little A D, Mershon D H, Cox P H. Spectral content as a cue to perceived auditory distance. Perception 21 (3), 405 – 416, 1992.

6 Mershon, D. H. and King, E. Intensity and reverberation as factors in the auditory perception of egocentric distance. Percept. Psychophys., 18, 409-415, 1975.

7 Mershon, D. H., Bowers, J. N. Absolute and relative cues for the auditory perception of egocentric distance. Perception, 8 (3), 311-322, 1979.

8 Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. Auditory distance perception in humans: A summary of past and present research. Acta Acustica, 91 (3), 409-420, 2005.

9 Begault, Durand R. 3-D sound for virtual reality and multimedia. Vol. 955. Boston etc: AP professional, 1994.

10 Fukuda, T., Horiuchi, T., Hokari, H. and Shimada, S. Relative distance perception by manipulation of the ILD of HRTFs. Acoust. Sci. & Tech., 24 (5), 325-6, 2003.

11 VORLÄNDER M., *Auralization*, Springer, Berlin 2007.

12 Zahorik, P. Assessing auditory distance perception using virtual acoustics. Journal of the Acoustic Society of America, 111 (4), 1832-46, 2002.

13 Lentz, T., Binaural Technology for Virtual Reality, Doctoral Thesis, RWTH Aachen University, Germany 2008

14 McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices.

15 Knudsen, E. I., & Brainard, M. S. (1995). Creating a unified representation of visual and auditory space in the brain. Annual review of neuroscience, 18(1), 19-43.

16 A.Plail, B. Fazenda, ‘On The Subjective Nature Of Binaural Externalisation’, IOA Proceeedings, Vol35-2, Reproduced Sound Conference, Manchester, UK, 2013