ON THE SUBJECTIVE NATURE OF BINAURAL EXTERNALISATION

A. Plail Acoustics Research Centre, The University of Salford, Manchester, UK
B.M. Fazenda Acoustics Research Centre, The University of Salford, Manchester, UK

1 INTRODUCTION

Externalisation (the sensation of hearing a sound object outside the head) on headphone reproduction is a recurrent topic in modern research for binaural sound reproduction. One of the inherent issues with testing for externalisation is its quantification. Generally, subjects are asked to indicate if a source sounds 'more external' or 'more distant' in order to determine the degree to which they are externalising. Herein lies an issue with the response's subjective nature. It cannot be certain that what a subject believes to be external is a genuine perception of externalisation, and a source seeming further away doesn't necessarily imply it is being localised farther from the head. There is also an issue with cohesion between subjects: it cannot be assured that two subjects that indicated a source to be more external by the same level actually perceived it to the same level as externalisation is an immeasurable response. The motivation therefore exists for a means of confirming that a subject is in fact externalising a source. This was attempted using considerations of the cocktail party effect in previous research[1,2,3,4].

An aspect that has received little attention in past research regards the situation in which the stimulus is presented. It follows from Brookes[6] and Wersényi[7], who's findings both involved measures taken to 'fool' the auditory system into perceiving the auralisation as a natural event, that the stimulus being realistic in the context of the current situation is an important factor. The effect that the plausibility of a given stimulus occurring has on its externalisation is hence of particular interest.

The main objectives of the study presented here are thus as follows:

1. To investigate an objective measure of externalisation via binaural simulation of room acoustics and the use of a speech intelligibility performance task.
2. To investigate the effect that the plausibility of a stimulus location has on its externalisation when reproduced binaurally.

2 EXPERIMENTAL DESIGN

An auralisation model that processes sound to induce externalisation when reproduced binaurally was developed and implemented during subjective tests. The testing incorporated the two objectives of the study as explained subsequently.

2.1 Objective Measurement Of Externalisation

A means of confirming whether or not the subject is actually externalising is proposed using a speech intelligibility test. If higher intelligibility is perceived when two speech sources are separated in space (along an axis of origin at the centre of the listener's head) as opposed to coincidentally positioned, it could be inferred that some cognitive process is responsible for this disassociation on a perceptual level. This task relies on the cognitive processing known as cocktail party effect[2,3].

Three experimental conditions were considered: *unmodelled, coincident* and *segregated* (Figure 1) where the speech streams were played simultaneously with synchronized onset times. For the *unmodelled* condition, the stimulus consists simply of simultaneously presented control and experimental sources, unaltered from the speech corpus[7]; no spatial or room artefacts are therefore present. For the *coincident* condition, the control and experimental source are modeled at exactly the same point in space: 1m from the centre of the listener's head, at a given azimuth and no elevation. For the *segregated* condition, the experimental source is modeled at a distance of 1.5m from the centre of the listener's head and presented simultaneously with the control source modeled at 1m.

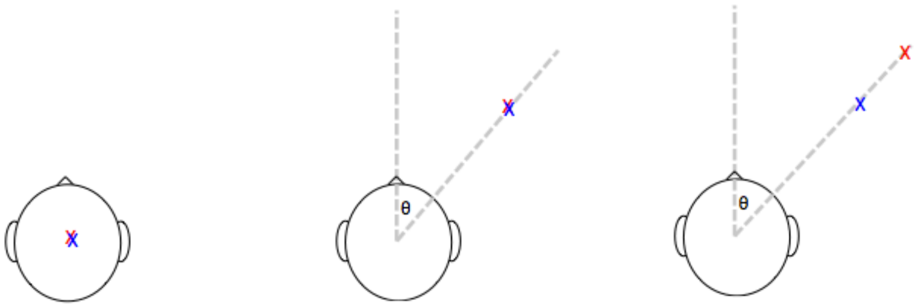

Figure 1: Illustration of unmodelled (left), coincident (centre) and segregated (right) conditions, where a blue cross represents the control source and a red cross represents the experimental source.

2.1.1 Speech Sources

The Coordinate Response Measure speech corpus[7] was used for the test sources. The database contains eight different speakers, four male and four female, each reciting 256 variations on a sentence that follows the format, "Ready [callsign], go to [colour] [number] now." This allows the callsign to be used as a flag to the listener to indicate which of two simultaneously presented sources they should pay attention to, and the colour and number to act as the intelligibility test variables.

While the corpus contains eight callsigns, five were selected for use that were phonetically dissimilar to a sufficient degree; easily confusable callsigns could prompt the listener to listen for the wrong sentence, hence confounding results. The callsigns were "Charlie," "Ringo," "Tiger," "Eagle" and "Baron." Similarly, only numbers one to six were used (out of eight) as the word "Seven" contains two syllables, giving it an increased chance of intelligibility. "Talker 3," one of the male speakers, was also disregarded due to a significantly slower delivery in many cases, causing simultaneous sentences to fall out of sync and relative intelligibility to be affected. The four colours, "blue," "red," "white" and "green" were all included as all have similar experimental 'difficulty.'

2.1.2 Test Setup

The test was designed to last no more than ten minutes. It consists of sixty trials of simultaneous sentences: twenty speech pairings under three conditions. Each trial should rarely exceed ten seconds.

Same gender pairings were considered to be spectrally similar; these were hence used in all experimental cases. Significant difference in the frequencies of the first and second formants between male and female speakers can be seen in Figure 2, this being a salient factor in speech intelligibility[8]. Male-female pairings were thus included to act as ‘dummy’ trials, considerably easier trials that serve to both keep subjects interested and act as an indicator to the subject’s understanding of the task: if they scored significantly poorly in these trials, their results would be discounted.

Sources were modelled at two azimuths, 10° and 70°, representing frontal and lateral azimuths respectively. Callsigns, colours, numbers and appropriate talkers were randomly paired using a MATLAB script.

Counterbalancing was implemented in order to prevent learning being accountable for an overall increase in intelligibility under any condition.

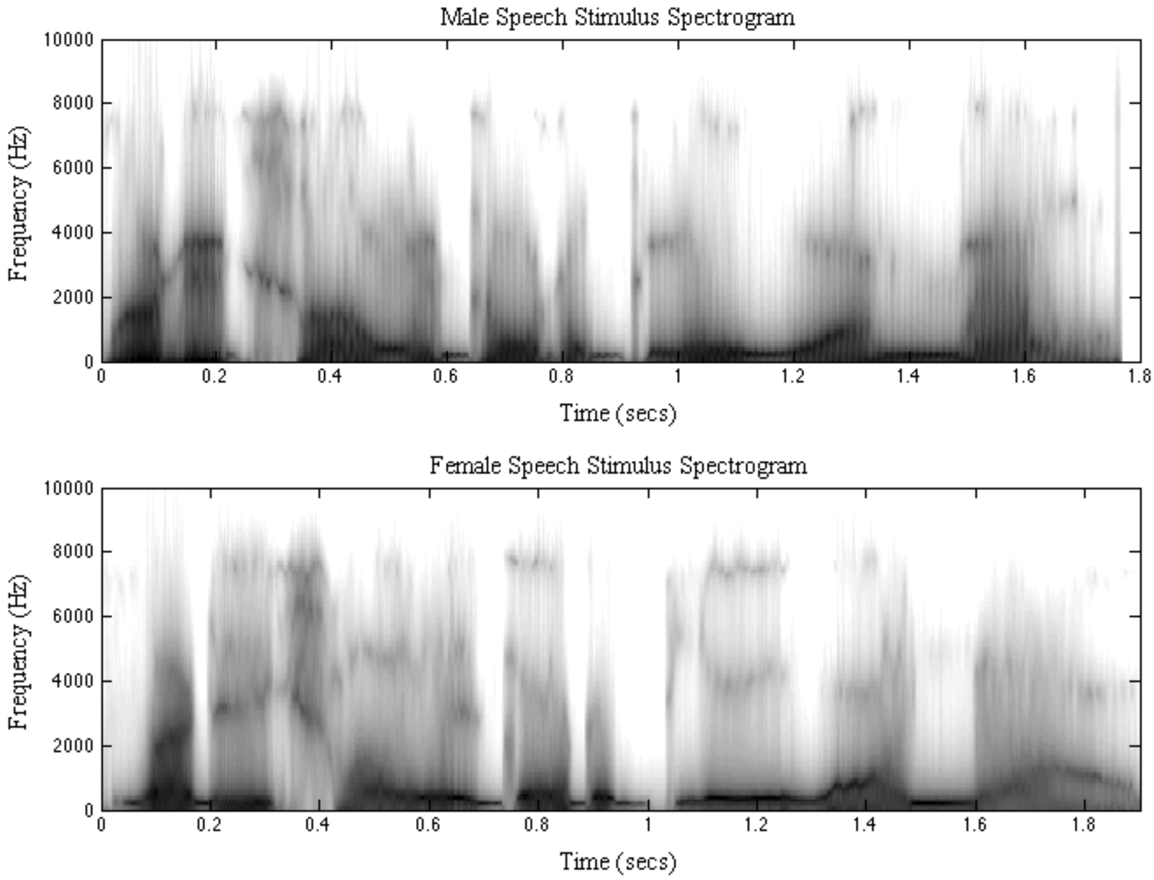

Figure 2: Spectrograms of the sentence “Ready Charlie go to blue one now” spoken by one of the male (top) and female (bottom) speakers

2.2 The Effect Of Plausibility

This aspect of the study explored the effect that a sound being plausible in the context of the current situation had on its perceived localisation (internal or external). Paramount to testing the effect of plausibility on externalisation is the set up of a stimulus that is believable and natural within the experimental situation. By recording an auditory event that had a plausible chance of occurring during the experiment and processing it using the auralisation model, it was possible to monitor the subject’s reaction to the stimulus presentation under different experimental conditions.

Three conditions were used: *unmodelled, correct azimuth* and *incorrect azimut*h. During the *unmodelled* condition, the stimulus is presented to the subject without having been processed by the auralisation model; there is therefore no spatial coding. For the *correct azimuth* condition, the stimulus is processed at the same azimuth as the event could realistically occur, and for the *incorrect azimuth* this azimuth is inverted about the median plane.

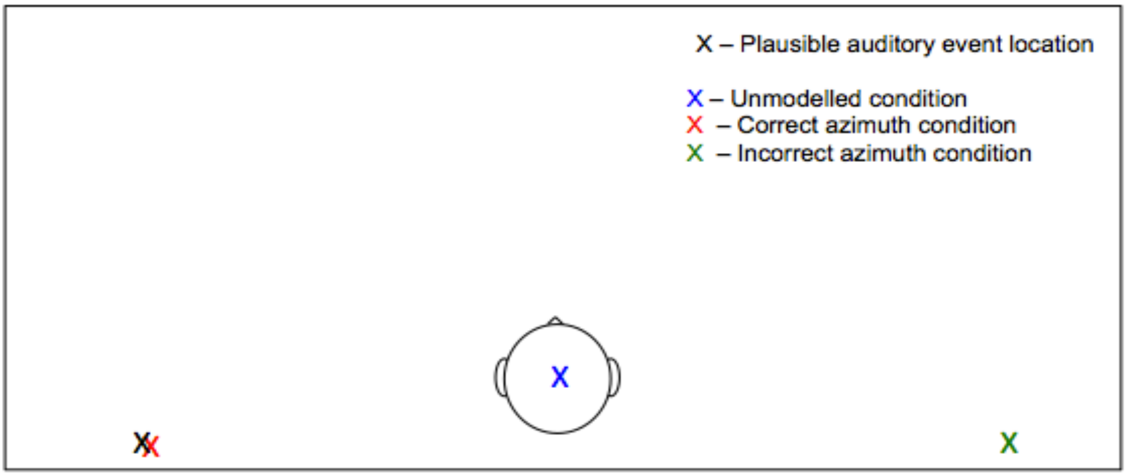

Figure 3: Illustration of the three conditions used to test the effect of plausibility

2.2.1 Experimental Stimulus

An umbrella falling over was used as the stimulus. An umbrella was placed in the test room leant up against the wall at a position just out of the subjects peripheral vision. The presence of the umbrella was brought to the subject's attention by knocking it over, seemingly accidentally, during the subject's briefing. This also serves to allow the subjects to learn the sound that is made when the umbrella falls over in the current acoustic environment. The umbrella was then picked up and leant against the wall again, establishing a scenario where it would be plausible to hear an umbrella falling over.

The stimulus was recorded using a Røde NT2-A, set to a cardioid polar pattern, positioned about 30cm from the source. It was crucial to capture as few room artefacts as possible, hence the use of a comparatively directional polar response and close positioning to capture a large D/R. Ideally the stimulus would have been recorded in anechoic conditions, but the nature of the source rendered this logistically impossible; the recording was undertaken in the same room as the experiment took place, which as the live room of a studio, had reasonable damping.

A particular effort was made to induce two separate stages to the auditory event: the sound of the umbrella starting to fall and the sound of it hitting the floor. The sound of it starting to move was characterised by a 'scraping' of the umbrella tip against the wall as it slowly overcomes inertia and begins to fall. As well as being more believable to the subject than a sudden fall, this 'scraping' signals to the subject that something is about to happen in their surrounding environment, hopefully inciting the requirement to turn and see what is happening. This sets up the possibility for the subject's reaction to the stimulus to indicate whether they believed the auditory event to be real, and if so, where they believed it to be happening.

The transfer function of the propagation of sound from outside to inside the headphone cup was convolved with the left and right channels of the modeled and unmodelled umbrella falling. This was done because outputs from the model provide auralisations of the subject sitting where they are within the test room without wearing headphones. In testing the effect of plausibility, the sound of the stimulus occurring in the room while headphones are being worn needed to be induced.

2.2.2 Test Logistics

As the plausibility test requires the stimulus to be unexpected, each subject was tested under only one condition. Once it has occurred once, they are aware that something else is going on and would most likely react differently to a second presentation of the stimulus. There is hence independence of observations as no participant is in more than one group.

Once all 60 speech intelligibility trials have been completed, the callsign screen is again presented as if another trial is beginning. While the callsign is being displayed, the plausibility stimulus is presented and the subject's reaction is recorded by taking frames from a webcam every 0.2 seconds for 4 seconds. This is followed by a series of questions intended to further assess the subject's perception of the sound they were just presented:

Question 1: Initially, the sound appeared to be coming from… (in the room/in the headphones)
Question 2: From which direction did it appear to be coming from? (Left/Centre/ Right)
Question 3: What do you believe the sound was?

The questions were presented separately by three consecutive figures with the intention of revealing less about the intent of the questions, thus hopefully prompting more accurate answers.

3 AURALISATION MODEL DESIGN

MATLAB was used to process an anechoic source such that, when reproduced binaurally, it is as if it is being excited from a point within the room and external to the head. This was executed via the combined implementation of room modeling acoustics and perceptual modeling to form a binaural auralisation.

The MIT database of KEMAR dummy head diffuse field equalised HRTF measurements[9] was used. The database contains HRIRs (head related impulse responses) measured at 1.4m in 710 different positions, covering the full 360° in azimuth and elevations from -40° to +90°.

As the HRTFs are discretised, it was necessary to round the angles of arrival of the modelled reflections. At 0° elevation, the HRTFs are discretised to every 5° azimuths.

As the HRTFs have been measured at a distance of 1.4m, an ILD exists within the response that is inherent to 1.4m; it was therefore necessary to remove this cue and re-introduce the correct ILD.

3.1 Modelling The Room

The dimensions and surface absorption characteristics were modelled based on the room in which the listening tests took place, the live room of a studio, with some approximations made. In the name of simplification, the room was assumed to be a perfect cuboid with a uniform surface absorption coefficient. The room was measured to have dimensions of 2.4×4.6×2.5m. An absorption coefficient of 0.47 was used in the early reflection calculations, and octave band absorption coefficients of 0.15, 0.32, 0.5, 0.6, 0.55 and 0.7 from 88Hz to 5680Hz (centre frequencies of 125Hz to 4kHz) were used for late reverberation calculation. These values were estimated from a published table of absorption coefficients for common materials[10]. The precise materials for which values were taken for the walls/floor/ceiling were fairly arbitrary; it was deemed unnecessary to precisely model the attributes of the boundary materials in the test room due to the many simplifications being made. It was taken into consideration that the space was relatively dead for a room of its size however, since it was a studio's live room.

3.1.1 Early Reflections

Image source modelling was used to determine the time and intensity of reflections. Reflections up to the 4th order were used (it is suggested that only first order reflections are necessary in the evocation of externalisation[11]). From the image source locations, the time of arrival and relative amplitude of each reflection could be inferred.

In order to encode the directional characteristics of each reflection, the azimuth and elevation relevant to the centre of the listener's head are returned, allowing the most accurate HRTF to be retrieved and convolved with the source. Each reflection can then be scaled and combined using the previously calculated amplitude factors and arrival times.

3.1.2 Late Reverberation

It can be approximated that the energy field is diffuse from 100ms of the reverberant tail; the late reverberation was hence modelled using decaying white Gaussian noise. Also, Völk[12] found that reflections arriving later than 100ms did not improve externalisation, detailed modelling is therefore not deemed to be required in this case.

Sabine's equation was used to generate separate reverberation times (T_{60}) for six octave bands from 125Hz to 4kHz using the absorption coefficients stated in Section 3.1. Although this does not cover the full audible spectrum, it does span the range that speech typically occupies. Longworth-Reed[13] used the same method when creating BRIRs for use with speech; he noted that the technique produces perceptually similar results to those derived from real measurements[14]. The T_{60} calculations were then used to derive a decay function for each of the octave bands.

Each channel of stereo white Gaussian noise is filtered into the six octave bands using 256th order FIR filters. The relevant decay function is then applied to each before they are summed to form one channel again.

3.1.3 Compiling The BRIR

The direct sound was added to the BRIR in the same fashion as each early reflection, by determining its amplitude and time of arrival according to the distance from the listener and scaling and adding it to the array accordingly.

The decaying white Gaussian noise was scaled such that the RMS level of the direct sound matched the RMS level of the decaying noise at the same point in time. The first 80ms of the decaying noise was then removed (replaced by zeros), and then added to the direct sound and early reflections to form the reverberant tail of the BRIR.

4 RESULTS

Eighteen subjects took part in the subjective testing, although one was discounted from the confirmation of externalisation test as they performed poorly in the dummy trials (37.5% correct).

4.1 Objective Measurement Of Externalisation

The model condition and azimuth of incidence were tested as two independent variables against the percentage correct dependent variable. Subjects' mean scores under each condition of the independent variables are shown in box-and-whisker plots (Figure 4a).

The *unmodelled, coincident* and *segregated* conditions each returned normality following a Shapiro-Wilk test; Mauchly's test of sphericity also indicated no significant violation. A one-way analysis of variance (ANOVA) with repeated measures was performed using the model conditions as samples, returning strong significance ($p < 0.0005$); a significant difference therefore exists between the means of at least one pair of conditions. A Bonferonni post-hoc test found strong significance between both the *unmodelled* and *segregated* conditions and the *coincident* and *segregated* conditions (p-values of 0.002 and <0.0005 respectively).

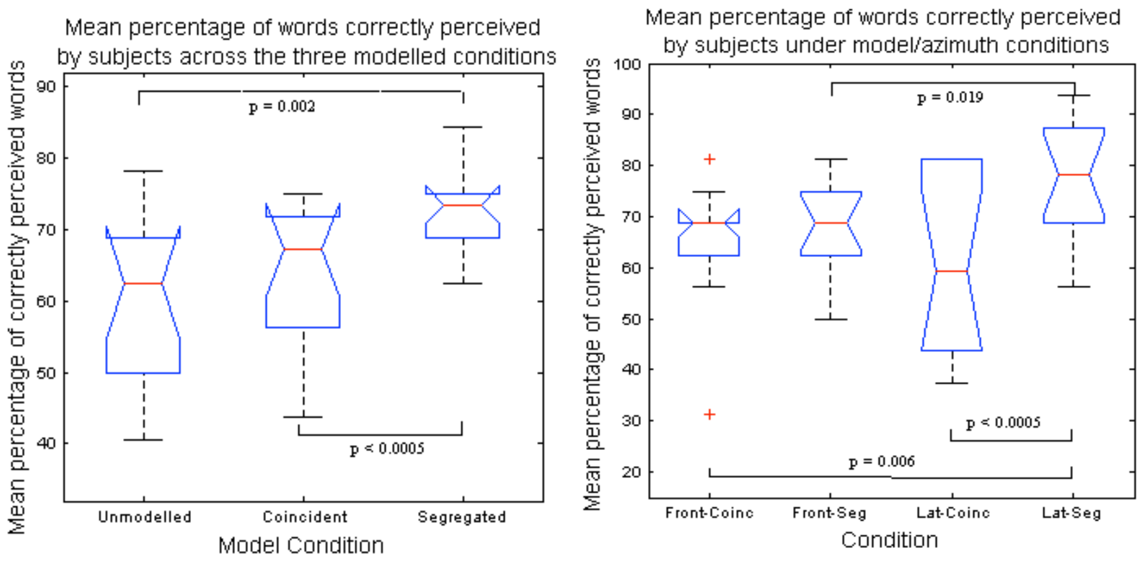

*Figure 4: Box-and-whisker plots showing a) the mean percentage of correctly identified words for each modelled condition b) the mean percentage of correctly identified words for model*azimuth factors. P-values are indicated where significant differences exist*

A two-way ANOVA with repeated measures found a p-value of 0.006 when testing the within-subject contrasts for model/azimuth conditions. A subsequent multiple comparisons test, shown in Figure 4b, found that the increase in correct responses under the *segregated* condition only occurs for *lateral* sources. Results under all conditions except the *frontal-coincident* case were normally distributed according to the Shapiro-Wilk test; this was neglected due to the insignificance effects found relating to this group.

4.2 The Effect Of Plausibility

Subjects' reactions to the presented stimulus as filmed by the webcam proved inconclusive in attempting to determine how they perceived the sound. Generally a reaction of confusion was exhibited, an expression that conveyed their acknowledgment that something odd had been perceived; this involved the 'screwing' of the face, looking up from the screen, and often surveying the room through the movement of their eyes. Drastic reactions were only exerted by two subjects, one who turned their head towards where the umbrella was placed, and one who turned around 180° to look through a window into the room where the experimenter was sitting. It was also common for no reaction whatsoever to be exerted (although, as one subject mentioned, this may have been because they did not want to look away from the callsign that they were being presented onscreen). Despite the inconsequential reactions of the subjects, perceived realism did seem existent; it even prompted one subject to enquire if the umbrella was being controlled from the experimenter's room.

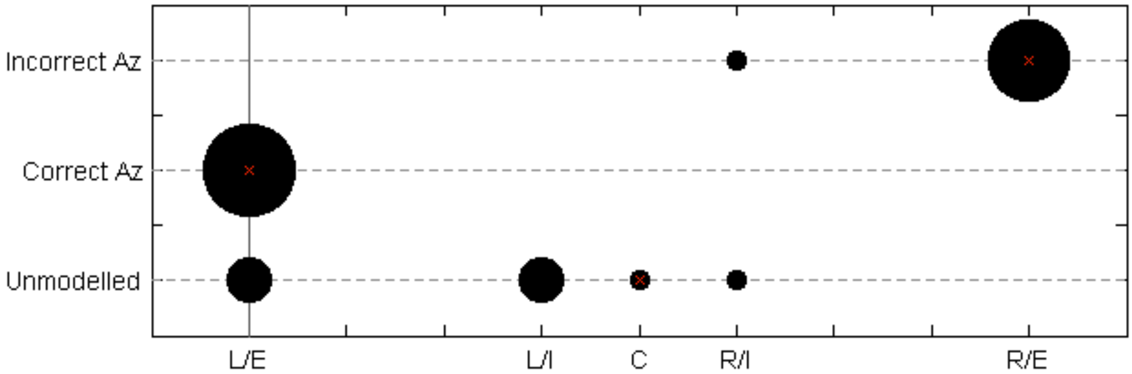

Figure 5: Visualisation of the subjects' responses. L, C and R represent left, centre and right, and I and E represent internal or external. The vertical grey line represents the location of the plausible stimulus and the red cross represents the location of the modelled stimulus. The size of the black dots represents number of subjects

Whilst it seems the plausibility stimulus had some effect in inducing externalisation, the model proved to be the stronger cue. When the two cues worked in conjunction under the *correct azimuth* condition, 100% of subjects perceived the source externally.

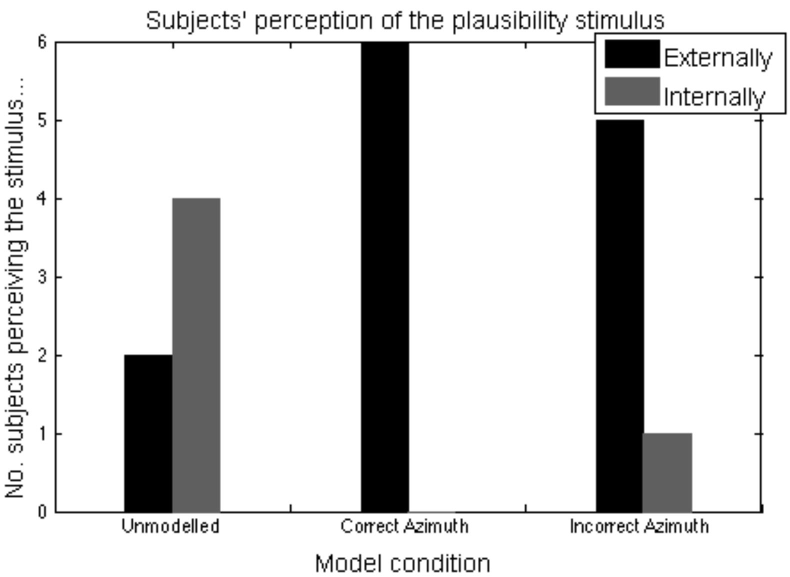

Figure 6: The number of subjects that perceived the plausibility stimulus internally/externally under each condition

When simply considering the perception as either internal or external (Figure 6), a chi-squared test returned a significant effect ($p = 0.027$). A significant number of subjects therefore externalised the stimulus, however not enough subjects took part in the study to allow for effects specific to each condition to be determined

5 DISCUSSION

The main deductions from this investigation are:

1) The increased intelligibility under the *lateral-segregated* condition was not due to externalisation, but rather some other cue present in the early reflections.
2) It is unclear whether subjects didn't externalise the speech sources at all, or whether they externalised sources at both distances to a similar degree.
3) The plausibility of a source seems to have implications on whether or not it is perceived externally, but binaural cues rather than its physically plausible location appears to have greater importance.

5.1 Increased Intelligibility For *Lateral-Segregated* Sources

While speech sources presented under the *segregated* condition improved the intelligibility of the farther source, this alone does not confirm that the effect was caused by subject externalisation.

The lack of significance between the *unmodelled* and *coincident* conditions implies that the addition of the room and HRTFs had a neutral effect on speech intelligibility. The significance associated with the *segregated* condition must have been due to the distance between the two speech sources, as this is the only factor that has been altered; the increased intelligibility is therefore caused by this separation.

Figure 4b shows that the effect only exists for *lateral* sources; this corresponds with findings that suggest that sources are externalised to a greater extent when positioned laterally[15,16]. However, Shinn-Cunningham[17] argues that the most important perceptual distance cues are source intensity, ILD and the D/R. The intensity and D/R cues are not dependent on source azimuth, and cannot therefore be responsible for the improved cognitive separation of the sources under the *lateral-segregated* condition. It seems to be generally regarded that a distance dependent ILD cue only exists for sources within a metre of the listener's head[17]; as the distances of the modelled sources in this study were 1m and 1.5m, it follows that the ILD cue cannot be attributed to the observed trend either. Conversely, the distances used are too small for spectral cues to be significant, particularly as speech doesn't contain high frequency components. It is unlikely, then, that the cognitive separation of sources is due to the experimental source being perceived as more distant (or more external) than the control source.

Considering the model, the only part of the auralisation anatomy that was affected by the change in distance was the early reflections (apart from a slight decrease in direct sound intensity); the cause of the effect must therefore result from this. Whilst it is unclear what specific aspect of the early reflections caused the improved intelligibility when lateral sources were segregated, the results do seem to suggest that the cognitive separation of two streams with the same angle of incidence is possible.

5.2 Deducing Whether Or Not Externalisation Occurred

Begault[11] found early reflections to be a salient factor in evoking the externalisation of speech stimuli when comparing reverberant and anechoic conditions. Early reflections cause a reduction in cross-correlation of the signals at each ear (relative to the direct sound) [18]; Begault[11] suggests that the resulting increase in differentiation of the binaural signal over time causes the external perception. Regarding the D/R cue, Zahorik[19] and Larsen[20] have found difference limens of 2-6dB, suggesting up to a doubling of source distance can be required in order for it to be perceived. Similar imprecision is therefore no doubt inherent to externalisation as a response, quite possibly to an even greater degree. Given that the experimental source was modelled at 1.5 times the distance of the control source, it is possible that the relative distances were too small to induce any noticeable effect; the distances were, however, constrained by the dimensions of the test room.

Had individualised HRTFs been used, greater precision in distance perception would have possibly resulted. Møller[21] found subjects localised speech sources to the same degree of accuracy as in real life when using individualised HRTFs, but to a lesser degree when using non-individualised HRTFs; this effect was significant around the median plane, analogous to the *frontal* condition in this study. This suggests that the improved cognitive separation of sources under the *lateral-segregated* condition may have been in part due to the use of non-individualised HRTFs being less of an issue for sources that are lateral to the median plane.

5.3 The Effect Of Plausibility On Externalisation

There is evidence to suggest that measures taken to make the auralisation seem 'natural' play an important role in its externalisation. For example, Wersényi's[6] simulation of consciously imperceptible head movements and Brookes'[5] decorrelation of binaural signals both introduced added elements of realism to the experiment, and both found an increase in externalisation. The study of plausibility introduced realism into the auralisation by using a stimulus that was relevant in the context of the current situation. The plausibility cue is similar to the vision cue in that it provides further, non-aural information as to the location of a source.

Of particular interest is the relative weighting of the use of the model and plausibility cue in localising the source. Under the *correct azimuth* modelled condition, the model and plausibility cue operated in conjunction, resulting in 100% subject externalisation. Under the *unmodelled* condition, only the plausibility cue was present, the results of which confirm that it is enough of a cue to induce externalisation in some. The *incorrect azimuth* modelled condition is particularly interesting as the plausibility cue and the model conflict; the results show that the model is unanimous in 'overwriting' the plausibility cue, however 100% externalisation was not observed. These findings therefore suggest that, while plausibility does provide some cue in externalisation, the modelling of the source is a much stronger factor.

6 CONCLUSIONS

On attempting to confirm whether a subject is in fact externalising a presented sound source, speech intelligibility was used as a means of objective quantification. A significant effect was only found when laterally incident sources were separated, with no difference observed for frontally incident sources. The three main distance cues are intensity, D/R and ILD; as the only cue that changes depending on source azimuth is the ILD, and ILD is not distance dependent for the distances used, the increased intelligibility cannot have been due to it being perceived more externally. While it is unclear what caused the effect, it is apparent by deduction that it resulted from the change that varying source distance had on the early reflections.

Due to the imprecise nature of distance perception, it is possible that the distance separation was not enough to invoke a perceptible difference, meaning the subject would have been unable to cognitively separate the sources. It is not clear, then, whether the subjects simply externalised sources modelled at 1m and 1.5m to the same degree, or didn't externalise at all. It was found that subjects found the task of trying to understand the speech quite intensive, suggesting that it might have dominated the cognitive processing and made it harder to localise the source externally.

The source stimulus being plausible within the context of the current environment was all that was required to induce externalisation in some subjects, over all conditions, a significant amount of subjects externalised the stimulus. How the response was modelled seemed to have a more important effect on its perception, however, as the model was found to be prevalent when the cues conflicted.

The results observed suggest certain implications outside of the study. The improved intelligibility when lateral sources were separated suggests that the ability to model the segregation of sources would result in their perception as more discrete. This has implications in music production, for example, where two similarly panned tracks could be modelled apart, helping to segregate the streams and create a better sounding mix when listening over headphones. Also, plausibility as an externalisation cue could have implications in inducing the external perception of sound effects produced over headphones whilst watching a film or playing a video game, particularly if the plausibility and model cues worked concurrently.

7 FURTHER WORK

A further study could manipulate individual aspects of the auralisation early reflections in order to try to determine the cause of the effect seen under the *lateral-segregated* condition. If the effect is common to other types of sources it could have useful implications, for example the mixdown stage of music production. A larger and less ‘dead’ room might be used, providing the possibility to use greater source distances and to provide a more prominent reverberant part. By implementing a greater separation of sources, it could be deduced whether the absence of increased speech intelligibility was due to the sources being externalised to the same extent or not at all. It would have been interesting to measure if, when subjects inputted incorrect answers to the speech intelligibility test, they had answered with what the control speaker had said rather than the experimental. This would provide further information as to whether the sources were perceived separately or whether they were ‘mashed together’ so that it was completely unintelligible. A similar test using sources other than speech would be useful, however the problem here lies in objectively measuring separation.

Of relevance for the application of externalisation is as an alternative to surround sound systems used for film and gaming. In this context it would be interesting to test for ‘vicarious plausibility,’ whereby it is tested if there is a cue in evoking externalisation when the stimulus is plausible in the context of what the subject is watching rather than their current environment.

8 REFERENCES

1. H Cherry, E. C.. Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustic Society of America*, 25 (5), 975-979 (1953).
2. Spieth, W., Curtis, J. F and Webster, J. C. Responding to one of two simultaneous messages. *Journal of the Acoustic Society of America*, 26 (1), 391-396 (1954).
3. Hirsch, I. J. Relation between localization and intelligibility. *Journal of the Acoustical Society of America*, 22(2), 196–200 (1950).
4. Durlach, N. I. and Colburn, H. S.. Binaural phenomena. In E C. Carterette and M. P. Friedman, editors, *Hearing*, volume IV of *Handbook of Perception*, chapter 10. Academic Press (1978).
5. Brookes, T. and Treble, C. The effect of non-symmetrical left/right recording pinnae on the perceived externalisation of binaural recordings. *The Audio Engineering Society,* 118th Convention (2005).
6. Wersényi , G. Effect of Emulated Head-Tracking for Reducing Localization Errors in Virtual Audio Simulation. *IEEE,* 17 (2), 247-52 (2009).

7. Bolia, R. S., Nelson, W. T., Ericson, M. A. & Simpson, B. D. A speech corpus for multitalker communications research. *Journal of the Acoustical Society of America.* 107, 1065–66 (2000).
8. Thomas, I. B. The influence of first and second formants on the intelligibility of clipped speech. *Journal of the Audio Engineering Society,* 16 (2), 182-85 (1968).
9. MIT. HRTF Measurements of a KEMAR Dummy-Head Microphone (1994)
10. Cox, T. J. and D'Antonio, P. *Acoustic absorbers and diffusers.* London: Spoon Press (2004).
11. Begault, D. R., Lee, A. S., Wenzel, E. M. and Anderson, M. R. Direct comparison of the impact of head tracking, reverberation, and individualised head-related transfer functions on the spatial perception of a virtual speech source. *The Audio Engineering Society,* 108th Convention (2000).
12. Völk, F. Externalization in data-based Binaural Synthesis : Effects of Impulse Response Length" *NAG/DAGA International Conference on Acoustics,* 1075-78 (2009).
13. Longworth-Reed, L., Brandewie, E. and Zahorik, P. Time-forward speech intelligibility in time-reversed rooms. *Journal of the Acoustical Society of America*, 125(1) (2009).
14. Zahorik, P. Perceptual scaling of room reverberation. *Journal of the Acoustical Society of America*. 115, 2598 (2004).
15. Kim, S. and Choi, W. On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach. *Journal of the Acoustic Society of America,* 117 (6), 3657-65 (2005).
16. Völk, F., Heinemann, F. and Fastl, H. Externalization in binaural synthesis: effects of recording environment and measurement procedure. *Journal of the Acoustic Society of America,* 123 (5), 3935 (2008).
17. Shinn-Cunningham, B. G. Distance cues for auditory space. *IEEE*, Special edition on virtual auditory space, 227-30 (2000).
18. Blauert, J. *Spatial hearing: the psychophysics of human sound localization*. Cambridge, MA: MIT Press (1983).
19. Zahorik, P. Assessing auditory distance perception using virtual acoustics. *Journal of the Acoustic Society of America,* 111 (4), 1832-46 (2002).
20. Larsen, E., Iver, Nandini., Lansing, C. R. and Feng, A. S. On the minimum audible diference in direct-to-reverberant energy ratio. *Journal of the Acoustic Society of America,* 124 (1), 450-61 (2008).

21 Møller, H., Sørensen, M. F., Jensen, C. B. and Hammershøi, D. Binaural technique: do we need individual recordings? *Journal of the Audio Engineering Society,* 44 (6), 451-469 (1996).