Immersive Audio Recording for Virtual and Augmented Reality

Abstract here

0 INTRODUCTION

This work follows on from [1]...

1 Recording at Abbey Road Studios

2 Post-Processing

2.1 Video

As two different 360°cameras were used during recording, two different methods of spherical video encoding were used. **[Look at has's AES/thesis for this?]**

2.1.1 Position A - 360°perspective

Writing

2.1.2 Position B - 180°perspective

2.2 Audio

3 Listening Tests

Two rounds of listening tests were conducted for viewing position A (test 1) and B (test 2). The procedure was identical however the data used, such as video and microphone configurations used were different for each. Participants were recruited from the University of York and Abbey Road Studios for both tests, all of who were required to have some previous experience with mixing/producing and/or spatial audio. The number of participants recruited for each test were as follows:

	UoY	Abbey Road	Total
Test 1	15	4	19
Test 2	29	9	38

3.1 Attributes Focus Group

The aim of the listening test was to assess the performance of each microphone array configuration for a VR environment in terms of its spatial and timbral quality. Due to the subjectivity of such a test, a focus group was assembled with the purpose of producing a list of mutually agreeable adjectives to use to describe certain spatial and timbral attributes. The attributes chosen to use within the listening tests are shown in table1.

Attribute	Description
Spatial	
Locatedness	How easily you can locate a sound source within the VR environment
Sense of Space	How well the space where the recording was made is perceived
Externalisation	Perception of sound coming from all around your head
Envelopment	Whether the sounds are perceived to originate inside of outside of the head
Timbral	
Full	Abundance of low frequencies present
Bright	Abundance of high frequencies present
Flat	Lack of high and low frequencies present
Rich	The mix sounds good with both high and low frequencies
Realistic	The sounds heard in the VR experience are realistic (sound like real instruments) and timbral characteristics have been preserved.
Loud	The perceived level sounds high

Table 1. Table of Spatial and Timbral Attributes

3.2 Procedure

Using an Oculus DK2 headset and a pair of Audio Technica MH50x headphones, participants were presented with an 80 second VR sample of the recording session which included the songs intro, verse and chorus using one of the microphone arrays appropriate for the samples viewing position. Once the clip was finished participants would answer a questionnaire that was spit into two main sections. The first asked them to rate on a scale of 1 - 10 each of the spatial audio attributes listed in table 1 where 1 indicated they did not experience that particular attribute well and 10 being that they experience that attribute very well. The second section asked them to select as many of the timbral

attributes they felt best described the overall timbre of the clip. Participants were also asked to rate on a scale of 1 - 10 how much they enjoyed the VR experience. This procedure was repeated a number of times depending on the number of microphone arrays to be presented (8 times for test 1 and 7 times for test 2). The order in which samples were presented was randomised per participant.

To ensure a uniform understanding of the list of attributes that were used in the questionnaire a short training exercise was conducted before each test. This involved taking the participants through each of the attributes with audio examples.

4 Analysis

To best analyse the results, five analysis targets were defined:

- Analysis 1: Does viewing position effect Spatial Attribute rating?
- Analysis 2: Does the choice of microphone array effect Spatial Attribute score?
- Analysis 3: What is the effect of using Directional or Diffuse-Field Arrays?
- Analysis 4: Is there a in perception of timbre with difference viewing positions?
- Analysis 5: Is there a correlation between SA score and selected timral attributes?

4.1 Analysis 1: Does viewing position effect spatial attribute scores?

To assess the potential effect of viewing position on spatial attribute score, data was grouped into 8 sections: An average score for each spatial audio attribute (4 groups) each split into the average score for viewing position A and B (8 groups) illustrated in figure 1.

It can be seen that the average spatial attribute score for viewing position A and B for each spatial attribute are close with the difference in overall mean score being 0.02. Running a Two-Sample T-Test between each of the four spatial attribute groups (e.g A vs B for locatedness etc) indicates no statistical significance. Running the same test for the averaged combined spatial attribute score (average of all four spatial attribute scores) also indicates no statistical significant between viewing position. This is made clear in figure 2 illustrating the overall similarity in the distribution of scores for viewing position A and B.

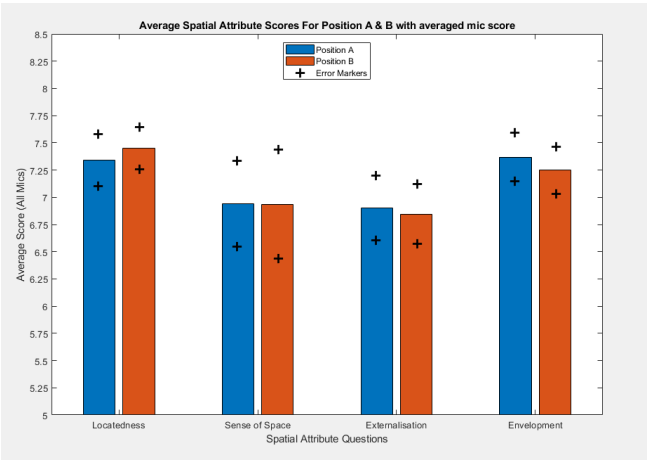

Fig. 1. Bar chart showing average spatial attribute score for each spatial attribute at viewing position A and B

Conclusion

The bar chart indicates that the average spatial attribute scores are extremely close with an overall average for both being 7.1. The Two-Sample T-Test indicates that the probability of these results recurring is not unlikely ($p > \alpha(0.05)$) and therefore the results shown are not statistically significant.

4.2 Analysis 2: Does the choice of microphone array effect Spatial Attribute score?

Breaking down the data showing in figure 1, figure 3 shows the average spatial attribute score across all used microphone configurations. The Anderson-Darling test was used to determine that not all of the sample data (participants scores per microphone) is normally distributed. Due to non-normally distributed data the Kurskal-Wallis (K-W) ANOVA was used to determine whether any of the samples were significantly different.

All groups returned a p-value greater than 0.05 other than 'Sense of Space' which returned $p = 0.0227$. As determined in section 4.1 there is no statistical significant difference between the data from viewing positions A and B. Therefore the data was separated according to their viewing positions and another K-W test was conducted on each group. This indicated a significant difference within the group of data from viewing position B, returning $p = 0.0035$.

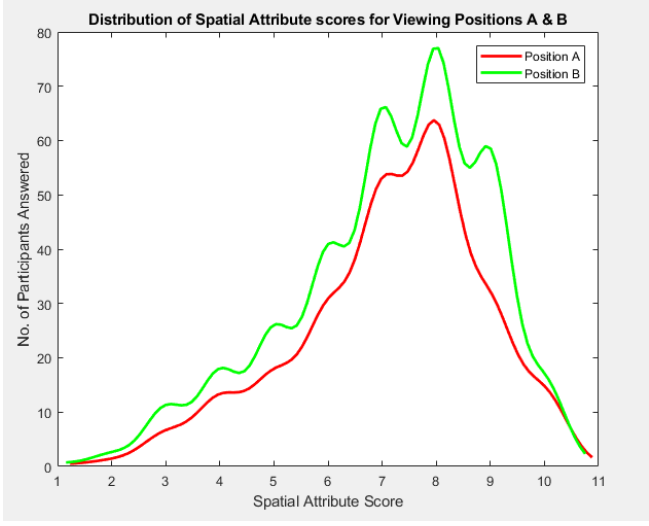

Fig. 2. Histogram showing the distribution of spatial attribute scores for viewing position A and B. This indicates that viewing position has little effect on spatial qualities in a VR environment.

Using MATLABs *multcompare* function, a post-hoc test was conducted to determine that the sample data for two microphone arrays, OCT and Hamasaki Cube were significantly different to the sample data for the spot microphones, circled in red and blue respectively in figure 3.

Though little statistical significance was found between the difference microphone arrays, by analysing figure 3 we can determine which microphones are overall the best choice with regards to this listening test. The two microphone arrays that showed statistical significance, the OCT and Hamasaki Cube also happen to consistently be among the top scoring microphone arrays across all spatial attributes.

Conclusion

For most spatial attributes, differences in mic choice is not statistically significant. However when it comes to sense of space, mixing in the microphone arrays with either the OCT or Hamasaki cube makes a significant difference relative to a pure spot mic mix.

4.3 Analysis 3: What is the effect of using Directional or Diffuse-Field Arrays?

Section 4.2 revealed no significant difference between using any of the different microphone arrays apart from when it comes to a 'Sense of Space'. However the difference was only found between using the spot mic mix against mixing the spot mics with either the OCT array or the Hamasaki Cube. As both of these microphone arrays belong to different groups (OCT is used as a directional array and the Hamasaki Cube as a diffuse field array) and were not significantly different from each other, it can also be stated that the use of directional or diffuse field arrays is also not statistically significant.

Analysing the bar chart in figure 3 however it is possible to come to some conclusions about particular microphone configurations. For example, looking at the scores for Sense of Space, the three diffuse field microphone in position C whilst viewing from position A can be said to objectively perform worse than the three of the directional microphones at position A (ESMA, ORTF, ST450). However as the Eigenmke scores higher than all of them, drawing a conclusion that one microphone type is superior would be a incorrect. A more collated visualisation of overall microphone configuration performance can be seen in figure 4, highlighting the narrow lead of the OCT microphone configuration.

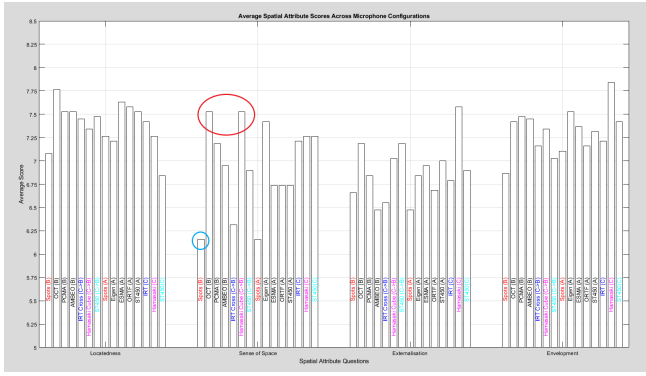

Fig. 3. Bar chart of the average SA score across all microphone configurations where (X) indicates the microphone location. The microphone names are displayed on their corresponding bar chart where (C-¿B) indicates a microphone from position C whilst viewing from B and (C) indicates a microphone from position C whilst viewing from position A

4.4 Analysis 4: Is there a difference in perception of timbre with difference viewing positions?

To assess whether viewing angle affected peoples perception of timbre, microphones that were used across both tests (spot mics on their own and microphones from position C) were directly compared for each of the timbral attributes. Figure 5 shows a the percentage of participants that selected each timbral attribute for each microphone the figures for which is shown in table 2.

The sway column is calculated by summing the percentage difference for each microphone and

$$Sway = \frac{\Sigma_{n=1}^{P} p}{N}$$

The 'Sway' column is calculated by summing the individual percentage difference for each timbral attribute resulting in a scalar representing which viewing position had a greater number of participants choosing each of the timbral attributes, where a positive number represents A and a negative number represents B.

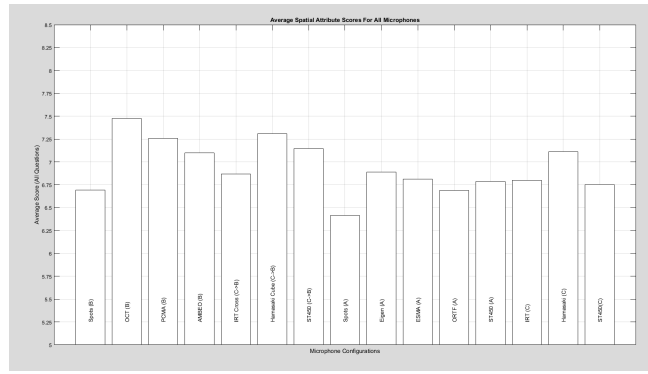

Fig. 4.

Attribute	Average Percentage Difference	Sway
Full	7.89	-1.32
Bright	10.53	-5.26
Flat	9.21	-6.58
Rich	20.39	12.5
Realistic	19.08	19.08
Loud	6.58	6.58

Table 2. Percentage difference

4.5 Analysis 5: Is there a correlation between SA score and selected timbral attributes?

5 REFERENCES

[1] H. Riaz, M. Stiles, C. Armstrong, A. Chadwick, H. Lee, G. Kearney, "Multichannel Microphone Array Recording for Popular Music Production in Virtual Reality," presented at the *Audio Engineering Society Convention 143* (2017 Oct), URL http://www.aes.org/e-lib/browse.cfm?elib=19333.

THE AUTHORS

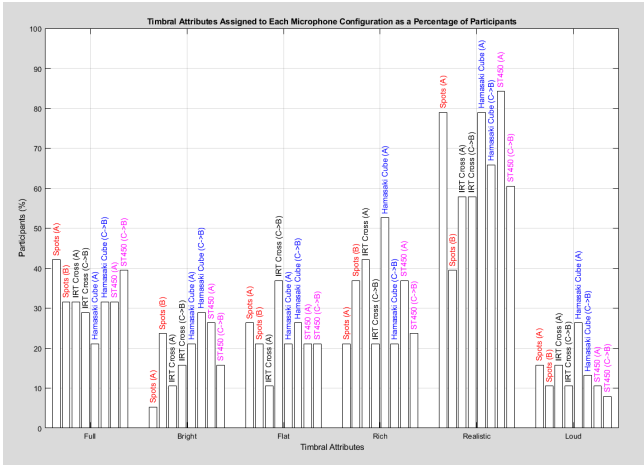

Fig. 5. Bar chart showing the timbral attributes chosen for each microphone array shared between viewing position A and B as a percentage of participants

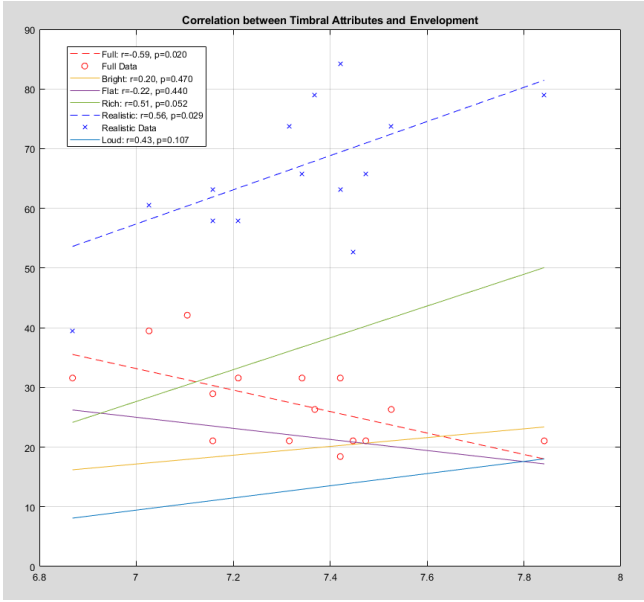

Fig. 6. Graph showing correlations between Timbral Attribute and Envelopment scores where dashed lines indicate a statistically significant correlation for which their individual data points have also been plotted.