# 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 Test Material Preperation

### 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. Both tests recruited participants both from the University of York and Abbey Road Studios. All participants 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:

| Recruted From | Test 1 | Test 2 |
|---:|:---:|:---:|
| UoY | 15 | 29 |
| Abbey Road | 4 | 9 |
| **Total** | **19** | **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

Participants was asked to wear and Oculus Rift DK2 headset and a pair of Audio Technica MH50x headphones and watch the first 80 seconds of the chosen take from their respective positions (A or B). Once the clip was finished participants would answer a questionnaire. This involved rating on a scale of 1 - 10 the level at which they experienced each of the spatial attributes listed in table 1 and to list any timbral attributes they felt best described the timbre of the recording.

To ensure that all participants understood each of the attributes in the same way a short training exercise was con-

ducted 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 correlation between SA score and selected timral attributes?

Analysis 5:  Is there a in perception of timbre with difference viewing positions?

### 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.
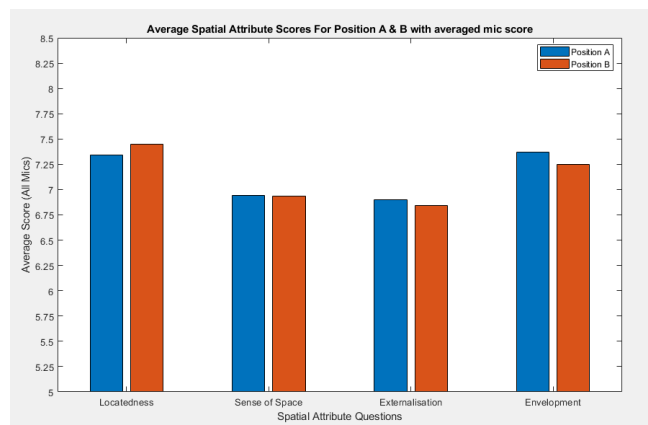
**Conclusion**

The bar chart indicates that the average SA score are extremely close for each SA 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 a p-value of $p = 0.0227$. 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 a p-value $p = 0.0035$.

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. The two microphone arrays that showed statistical



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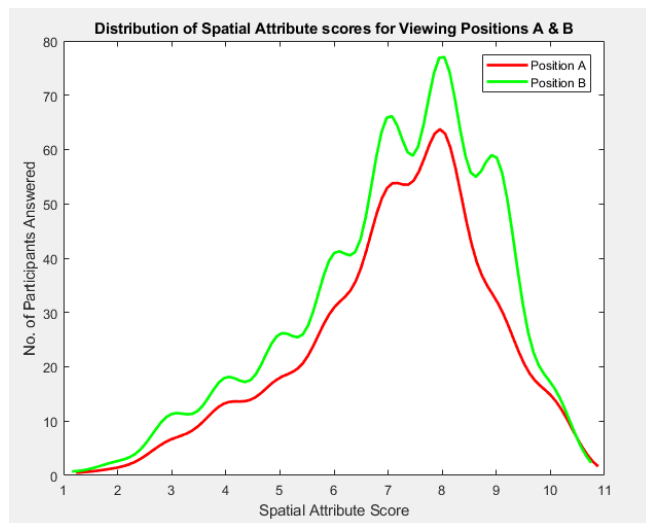


Fig. 2. Histogram showing the distribution of spatial attribute scores for viewing position A and B. This indicates that viewing position has little effect on spatial qualities in a VR environment.

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, particularly the OCT and Hamasaki cube make a significant difference to listening to just the spot mics on their own.

### 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 just the spot mic mix against mixing the spot mics with either the OCT array or the Hamasaki Cube. As both of these microphone arrays are different from each other, one was used as a directional array and the other as a diffuse field array, and that they 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.

It can also be seen that the OCT scores are always the highest other than for envelopment, where it comes as a close third.

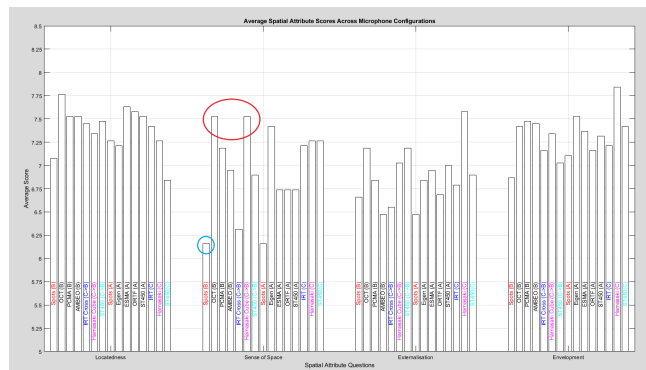Analysing the bar chart in figure 3 however it is possible to come to some conclusions about particular micro-



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

phone configurations. For example, looking at the scores for Sense of Space, the three diffuse field microphone in position C whilst viewing from position A all score relatively similar, an average of 7.2 and three of the directional microphones at position A (ESMA, ORTF, ST450) all score the same average score of 6.7. However it is the Eigenmike that levels the difference between these two groups, with an overall higher average higher score.

### 4.4 Analysis 4: Is there a correlation between SA score and selected timbral attributes?
### 4.5 Analysis 5: Is there a in perception of timbre with difference viewing positions?
### 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**