Immersive Audio Recording for Virtual and Augmented Reality

0 Analysis Overview

All test data was recorded using a Google Form and collated into a large document upon finishing the testing period. All data was then imported into MATLAB for analysis. Table 1 shows the average spatial attribute results for each microphone across each spatial attribute. The highest scores are highlighted in green. The directional OCT microphone array from microphone position and viewing position B scored highest for 'Locatedness' and scored the highest along with the Hamasaki Cube, a diffuse-field microphone array located at position C for 'Sense of Space'. The Hamasaki Cube when viewing from position A scored the highest for 'Externalisation' and 'Envelopment' which results in the highest over all spatial attribute average score. A break down of timbral attribute scores is covered in section 0.4

The following six sections will be used to further break down and analyse the data:

Analysis 1:   Does viewing position affect Spatial Attribute rating?

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

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

Analysis 4:   Does perception of timbre change with difference viewing positions?

Analysis 5:   Is there a correlation between spatial attribute score and selected timral attributes?

Analysis 6:   Is there a correlation between the enjoyment rating and spatial audio attributes?

This will be followed by an summary section with conclusions draw from each analysis section.

0.1 Analysis 1: Does viewing position affect spatial attribute scores?

To measure the effect that viewing from a different position within the room might have on the participants perception of the given spatial attributes, the test data was split into four groups containing the average results for

each spatial attribute, each of which was split into two sub groups corresponding to the data obtained when viewing from position A or B. The data is illustrated in figure 1.

It can be seen that the average spatial attribute scores for viewing position A and B 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 spatial attribute scores are extremely close with an overall average for both being 7.1. The Two-Sample T-Test returned $p > 0.05$ for all data groups indicating that the probability of these results recurring is not unlikely and therefore the results shown are not statistically significant.

0.2 Analysis 2: Does the choice of microphone array affect Spatial Attribute score?

Breaking down the data showing in figure 1, the average spatial attribute score across all used microphone configu-

Viewing Position	Mic Position	Microphones	Locatedness	Sense of Space	Externalisation	Envelopment	Average
A	A	Spots	7.26	6.16	6.47	7.11	6.75
		Eigen	7.21	7.42	6.84	7.53	7.25
		ESMA	7.63	6.74	6.95	7.37	7.17
		ORTF	7.58	6.74	6.68	7.16	7.04
		ST450	7.53	6.74	7.00	7.32	7.14
	C	IRT	7.42	7.21	6.79	7.21	7.16
		Hamasaki	7.26	7.26	7.58	7.84	7.49
		ST456	6.84	7.26	6.89	7.42	7.11
B	B	Spots	7.08	6.16	6.66	6.87	6.69
		OCT	7.76	7.53	7.18	7.42	7.47
		PCMA	7.53	7.18	6.84	7.47	7.26
		Ambeo	7.53	6.95	6.47	7.45	7.10
	C	IRT Cross	7.45	6.32	6.55	7.16	6.87
		Hamasaki	7.34	7.53	7.03	7.34	7.31
		ST450	7.47	6.89	7.18	7.03	7.14

Table 1. Table containing the average spatial attribute scores for all microphone with on over all average spatial attribute score. Highest scoring microphones are highlighted in green.

rations is shown in figure 3. 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 the average scores between any of the samples were significantly different. All groups returned $p > 0.05$ other than 'Sense of Space' which returned $p = 0.0227$. As determined in section 0.1 there is no statistically significant difference between the data due to different 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 $p = 0.0035$.

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

Conclusion

Looking at the scores for the OCT and Hamasaki cube across all spatial attributes, it is apparent that they typically perform well among other microphone configurations. It is therefore probably more appropriate to look at this result as a significantly poor performance from the spot microphones whilst compared to microphone arrays that consistently perform well as opposed to an significantly exceptional performance from the two microphone arrays. Across all other spatial attributes it has been shown that the difference between performance is not statistically significant.

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

Section 0.2 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
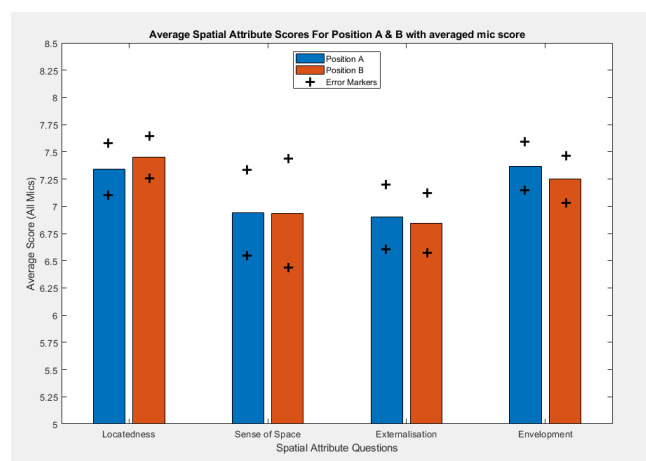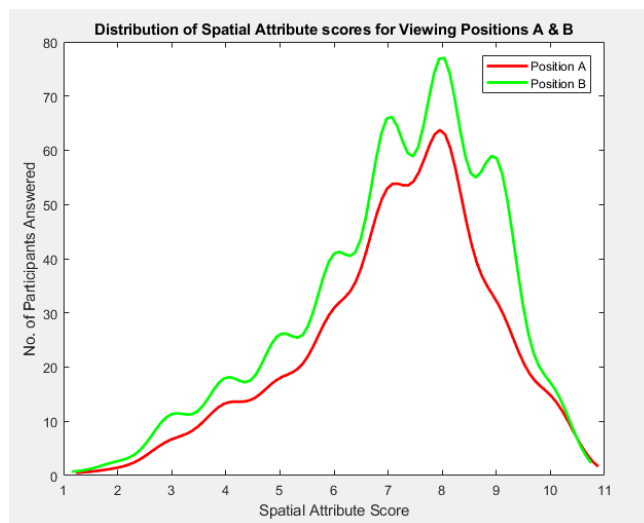

Fig. 2. Histogram showing the distribution of spatial attribute scores for viewing position A and B. This indicates that viewing position has little effect on spatial qualities in a VR environment.

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.

Conclusion There appears to be no significant effect of using either a directional or diffuse-field microphone array providing that spot microphones alone are not used

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

The effect of different viewing positions on participants perception of timbre can be assessed by comparing the data collected for microphones that were shared across both viewing positions which includes the spot microphones and microphone arrays from position C. Figure 5 shows the percentage of participants that selected each timbral attribute for each microphone. Table 2 presents the corresponding data showing the percentage difference between each microphone pair (each microphone at both positions) calculated by subtracting the results for microphones at position B from the same microphones at position A. For

Fig. 1. Bar chart showing average spatial attribute score for each spatial attribute at viewing position A and B
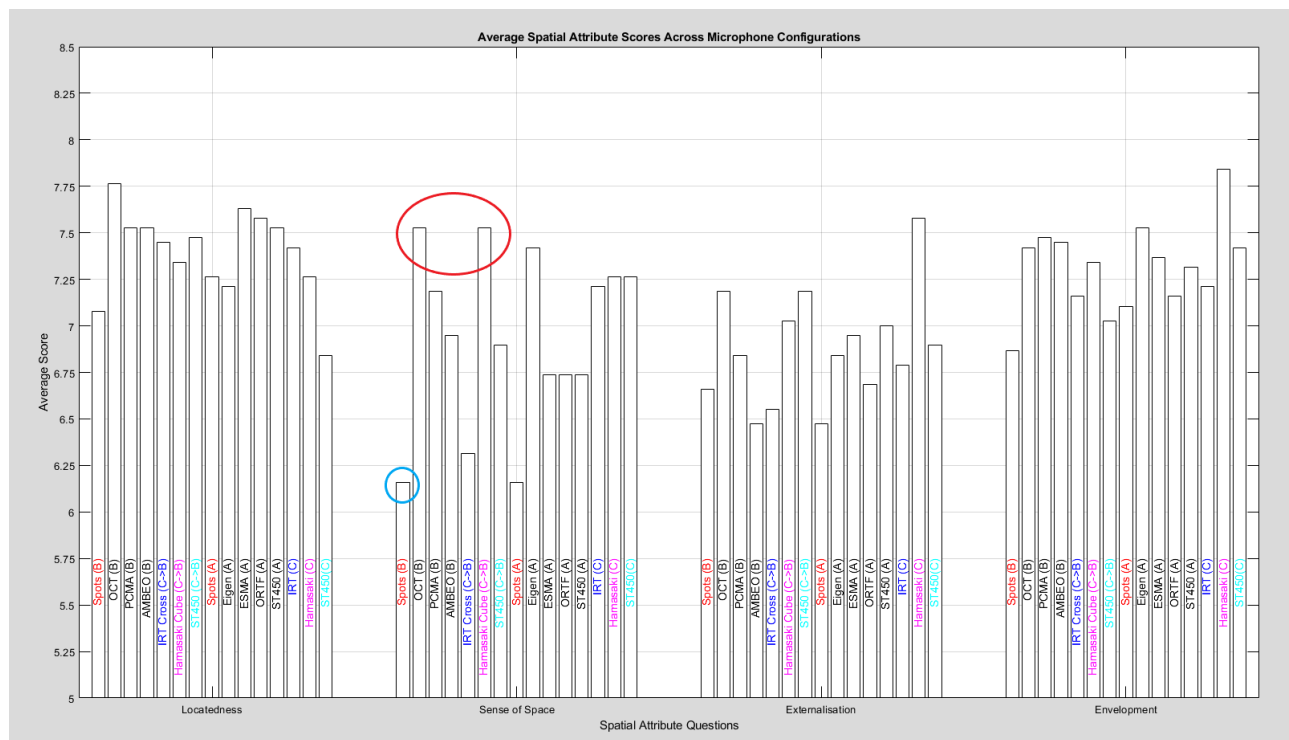
Fig. 3. Bar chart of the average spatial attribute score across all microphone configurations where *microphone(X)* indicates the microphone configuration and location. The microphone names are displayed on their corresponding bar where (C−>B) indicates a microphone from position C whilst viewing from B and (C) indicates a microphone from position C whilst viewing from position A. Microphones that were shared across both viewing positions are highlighted in matching colours.

each timbral attribute, the average results column indicates whether the attribute was selected by more participants for viewing position A (positive number) or position B (negative number). By looking at the table it can be seen that across all microphones the timbral attributes 'Realistic' and 'Loud' share the trend that they were selected by an equal or greater percentage of participants for viewing position A. The timbral attribute 'Realistic' experienced the greatest variation between viewing positions with an average difference of 19%. Though the increased frequency of participants selecting 'Loud' is only by a minor amount, it is possible that the consistency is caused simply by the fact that viewing position A is closer to the musicians thus increasing the possibility for participants to perceive the

audio as 'Loud'.

Conclusion

In terms of timbre there appears to be little difference made by watching the performance from either position. The attributes effected most by changing viewing positions are a sense of whether things sounded realistic for which position A showed a preference. This increased difference however is mostly influenced by the large difference in spot
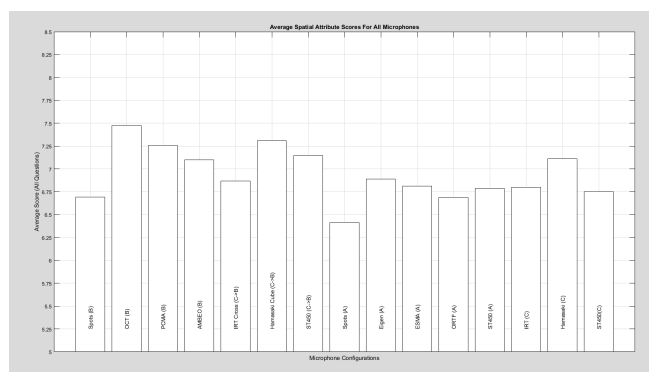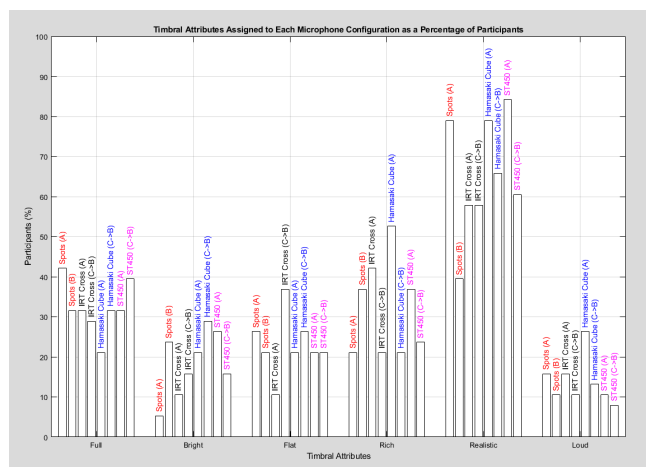

Fig. 4.

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

Attributes	Microphones (%)				Averaged
	Spots	IRT Cross	Hamasaki Cube	ST450	
Full	10.53	2.63	-10.53	-7.89	-1.32
Bright	-18.42	-5.26	-7.89	10.53	-5.26
Flat	5.26	-26.32	-5.26	0	-6.58
Rich	-15.79	21.05	31.59	13.16	12.5
Realistic	39.47	0	13.16	23.68	19.08
Loud	5.2	5.2	13.16	2.63	6.58

Table 2. Table showing the percentage difference between each pair of microphones calculated by $A - B$

microphone scores where there is a 39.5% increase from the low 39.5% of participants regarding the spot mics at position B to sound realistic to the 79% of participants regarding the spot mics at position A to sound realistic.

It is hypothesised by the author that this is due to the unnatural lack of room acoustics heard when positioned at a distance where the sound of the room is expected to be heard. As the participant is standing much closer to the musicians at position A where the direct to reverb ratio is naturally much higher, the lack of room acoustics may be less noticeable and still be perceived as realistic. As a consequence of a lack of room acoustics present in the spot microphone mix, sound sources can be perceived more spatially isolated from one another. The unnaturalness of this is possibly less obvious when fully surrounded by the musicians as experienced at position A as this introduces an unnatural separation of the musicians for a listening experience. When viewing from position B however, as the musicians can be seen closer together, the unnatural sound source separation perceived may stand out to participants and further affect their judgement.

0.5 Analysis 5: Is there a correlation between spatial attribute score and selected timbral attributes?

Correlation coefficients were calculated for each combination of spatial attribute score against timbral attribute score. Most calculations returned weak correlations with $p > 0.05$ other than two timbral attributes specifically when comparing against the 'envelopment' spatial attribute. Figure 6 shows the line of best fit for all timbral attribute scores against the spatial attribute scores for 'envelopment'. The dashed lines indicate a statistically significant correlation as found with the timbral attributes 'Full' and 'Realistic'. The graph indicates that there is a significant positive correlation between the increase in sense of envelopment in the virtual environment with the sense of the virtual environment sounding realistic.

The data also indicates a negative correlation between participants sense of envelopment and their perception of the timbre sounding 'Full'. The reason as to why this is is not exactly clear. It could be said that participants perception of 'Full' may mean an unnatural abundance of bass frequencies which may sound unrealistic. If this is the case, as we can see from the positive correlation between a sense of envelopment and a sense of the environment sounding

realistic, that an unrealistic sounding environment would lead to a decrease in participants sense of envelopment.

Conclusion

The only spatial attribute to show a significant correlation with timbral attribute data is 'Sense of Space' with a positive correlation for a perception of the mix sounding 'Realistic' and a negative correlation for the perception of the mix sounding 'Full'.

0.6 Analysis 6: Enjoyment Rating

The purpose of both VR and music is primarily to entertain. If users do not enjoy what they are experiencing then the cause should be addressed. For this purpose participants were asked to rate on a scale of 1-10 how much they felt they enjoyed the experience. Figure 7 shows the line of best fit for average enjoyment score against average spatial attribute score for each microphone for both viewing position A and B.

Both graphs show a strong positive correlation between the scores for spatial attributes and enjoyment ratings with a correlation coefficient $r = 0.81$ and $r = 0.96$ for viewing position A and B respectively.

Conclusion

This suggests that when provided with higher quality spatial audio for an experience such as the one presented, viewers are more likely to enjoy the experience. It is important to address that this VR experience differs from the currently most popular form of VR experience available which is gaming. VR games offer the possibility of movement and interaction which may distract users from the quality of spatial audio. However, in an experience such as the one presented in this paper, where the focus of the experience is listening to music and lateral movement is not
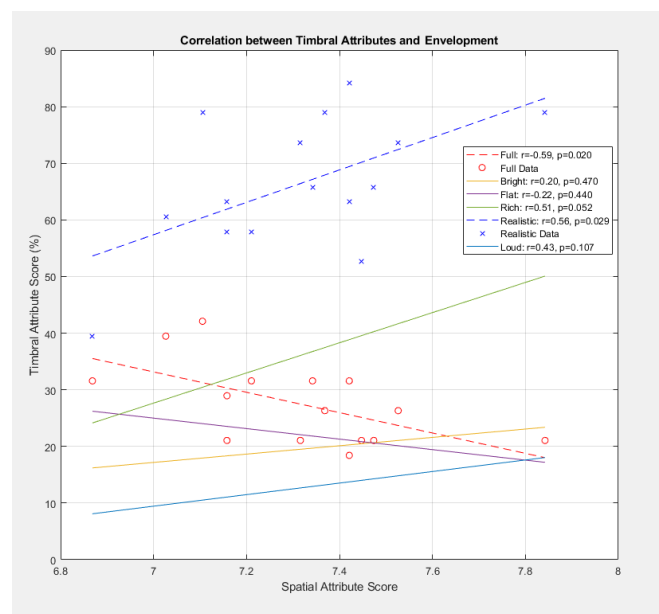
Fig. 6. Graph showing correlations between Timbral Attribute and Envelopment scores where dashed lines indicate a statistically significant correlation for which their individual data points have also been plotted.

an option, the users focus is not taken away from the spatial audio quality. It is therefore important for listening to music in VR that the spatial audio quality is high to ensure the listening experience is enjoyed.

1 Analysis Summary

Ana1: There is no significant difference on SA score between viewing positions

Ana2: OCT and Hamasaki Cube were found to be statistically significant with regards to the SA 'Sense of Space'. This can be seen as a significantly poor performance from the spot mics at viewing position B. 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.

Ana3: Results are inconclusive. There appears to be no significant difference between using either a direct or diffuse-field microphone array so long as spot microphones alone are not used.

Ana4: Timbre itself seems to be unaffected by view position. However participants perception of 'realistic' and 'loud' appears to be the most affected with a higher score for viewing position A. It is hypothesised that this is due to the unnatural room acoustic and sound source seperation caused when listening to the spot mic mix that skews the result this way.

Ana5: Statistically significant positive correlation between sense of space and 'realistic'. Also significant negative correlation between SoS and the perception of the timbre sounding 'Full'.

2 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`
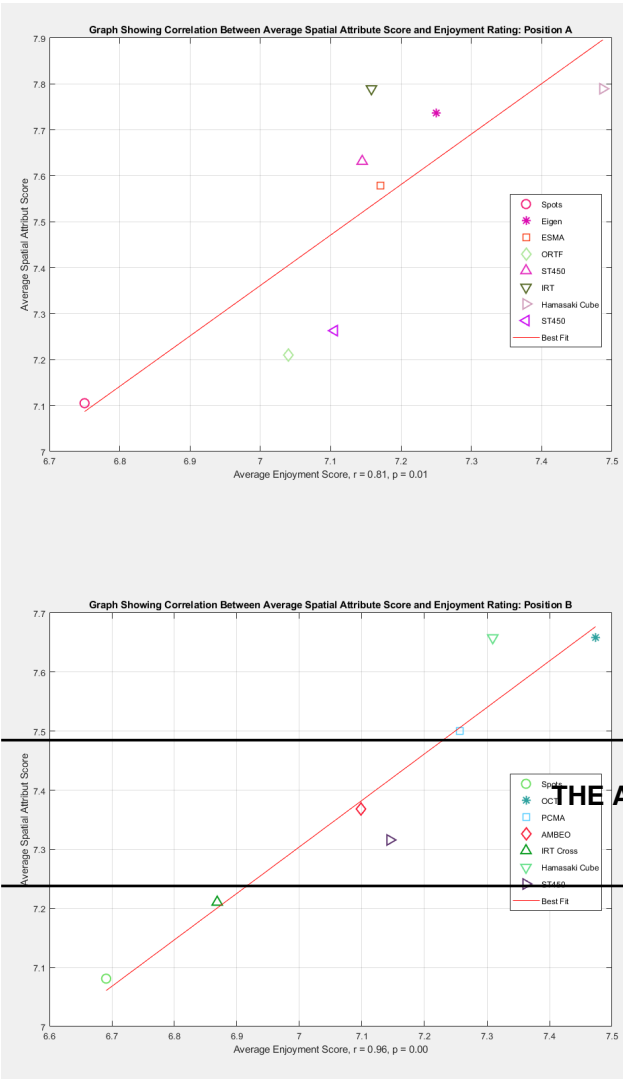

Fig. 7. Two graphs showing a positive correlation between enjoyment rating and spatial attribute scores for viewing positions A (Top) and B (Bottom).

THE AUTHORS