A QoE and Visual Attention Evaluation on the Influence of Spatial Audio in 360 Videos

Amit Hirway
Department of Computer & Software Engineering
Athlone Institute of Technology
Athlone, Ireland
a.hirway@research.ait.ie

Yuansong Qiao
Software Research Institute
Athlone Institute of Technology
Athlone, Ireland
ysqiao@research.ait.ie

Niall Murray
Department of Computer& Software Engineering
Athlone Institute of Technology
Athlone, Ireland
nmurray@ait.ie

***Abstract*— Recently, there has been growing interest from academia and industry on the application of immersive technologies across a range of domains. Once such technology, 360° video, can be captured using an omnidirectional multi-camera arrangement. These 360° videos can then be rendered via Virtual Reality (VR) Head Mounted Displays (HMD). Viewers then have the freedom to look around the scene in any direction they wish. Whereas a body of work exists that focused on modeling visual attention (VA) in VR, little research has considered the impact of the audio modality on VA in VR. It is well accepted that audio has an important role in VR experiences. High quality spatial audio offers listeners the opportunity to experience sound in all directions. One such technique, Ambisonics or 3D audio, offers a complete 360° soundscape. This paper reports the results of an empirical study that looked at understanding how (if at all) spatial audio influences visual attention in 360° videos. It also assessed the impact of spatial audio on the user's Quality of Experience (QoE) by capturing implicit, explicit, and objective metrics. The results suggest surprisingly similar explicit QoE ratings for both the spatial and non-spatial audio environments. The implicit metrics indicate that users integrated with the spatial environment more quickly than the non-spatial environment. Users who experienced the spatial audio environment had a higher maximum mean head pose pitch value and were found to be more focused towards the sound-emitting regions in the spatial audio environment experiences.**

Keywords—360° Video, Spatial Audio, Ambisonics, QoE, Audio-Visual Attention

I. INTRODUCTION

Virtual Reality (VR) applications, including 360° videos [1] have gained significant interest in recent years. Head Mounted Display (HMD) technology is a popular way to experience 360° videos as the user can view and experience the videos through the display in 6 degrees of freedom (DOF). Based on the orientation, the HMD displays the current field of view (FoV) which is a fixed-size region, in the range of 90° to 110°, depending on the HMD being used. This level of interaction facilitates more immersive and realistic experiences. However, streaming these videos to HMDs is extremely challenging. Since a viewer never sees the full 360° video at the same time, streaming the entire video at full resolution is wasteful in terms of resources, including bandwidth, storage and computation. A significant amount of research has therefore been undertaken to understand visual attention (VA) in order to optimize existing 360° video streaming applications.

However, sound is also an important part of an immersive experience, it contributes to immersion and presence [28]. Relatively speaking, significantly more work has focused on VA analysis in immersive media compared to works that have considered the audio modality. Although public datasets [2][3][4] with user viewing behaviors (head-tracking, eye-tracking) while watching 360° videos are available, these are video-only datasets or non-spatial audio datasets. Furthermore, most of the previous research on "Audio-Visual Attention" have focused only on audio-visual attention in traditional (non-360°), non-spatial sound videos [5] [6] [7]. Recently, industry and academia has begun to see spatial audio as a key factor in VR experiences, due to its advanced features (e.g. more realistic features, natural listening experience, better positional accuracy). Adding spatial audio to the VR environment may completely change the way how users watch the videos: how they move their heads; directions in which they focus; and what content they can remember after each session.

Ambisonic audio [11] has recently gained prominence with the rise in VR and 360° videos. Ambisonics is a 360° audio capture and playback method. The most common format in Ambisonics is the 4-channel format called Ambisonics B-format. It uses as few as four channels to replicate the full sound sphere. The four first-order B-format channels are called W, X, Y and Z. Whilst the first-order B-format provides spatial immersion with a higher resolution than conventional surround technologies, higher-order B-format audio can provide even higher spatial resolution, offering more channels with distinct polar patterns. For example, second order Ambisonics uses 9 channels, third order Ambisonics employs up to 16 channels, with sixth order Ambisonics employing 49 channels.

Quality of experience (QoE) is critical to the success of immersive VR applications. It is defined as “the degree of delight or annoyance of the user of an application or service. It results from the fulfilment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user’s personality and current state [20]”. Due to the enormous growth and popularity of multimedia services and applications over the last decade, the perceived end-user QoE has become a vital criterion [9]. Various methods have been investigated to measure QoE which have involved the capture of explicit, implicit and objective metrics.

In the context of 360° videos and ambisonic sound, this paper presents the results of an empirical study to understand how non-spatial and spatial audio (third order ambisonics) can affect users Quality of Experience and influence Visual Attention in 360° videos. The 360° videos with spatial audio span a range of content categories: Opera, Instrumental, Riding, and Exploration. Our study involved executing user trials to produce a series of multimodal datasets of implicit, explicit and objective metrics: head movements and other physiological signals (heart rate - HR, electrodermal activity -

EDA) of users watching 360° videos with non-spatial and spatial audio on an HMD. Part of these trials also involves users self-reporting QoE via post experience questionnaires.

The rest of this paper is structured as follows: section II presents research works that have performed evaluations of visual and audio-visual attention in non-spatial environments; section III describes the experimental setup; section IV outlines the methodology ; section V presents the results with a discussion, whilst section VI presents the conclusion.

II. Related Work

In [19], Almquist et al., studied viewing behavior of subjects watching several 360° videos. While the subjects watched the videos, data related to their head orientation and rotation speed (captured via sensors in the HMD) was collected. Each video was classified into one of a group of categories. The aim of the categories was to help understand if different video contents produced different viewing patterns. An important inference was that the viewing angle distribution depended heavily on the video content. The viewers spent much of the time looking to the front of the video for categories of Static Focus and Rides, while an almost linear distribution was discovered for the Exploration and Moving Focus categories. It was also reported that the yaw-rotation was the most common rotation in comparison to pitch and roll rotations.

In [3], Lo et al. gathered 360° videos from YouTube with diverse features. The 360° videos were divided into 3 categories: (i) Computer Generated, fast-paced (ii) Natural Image, fast-paced, and (iii) Natural Image, slow-paced. They used an open-source head tracking tool to record viewer orientations from the HMD sensors including yaw, pitch, and roll. They claimed to have created a unique dataset with both content data (such as image saliency maps and motion maps) and sensor data (such as positions and orientations).

In [7], Min et al. investigated the circumstances under which sound could influence visual attention. A set of YouTube videos were used to perform eye-tracking experiments with different test conditions: videos with (AV) and without (V) soundtracks. By assessing the differences in eye movement data collected in AV and V conditions, they concluded that influence of sound depended on the consistency between visual and audio signals. Audio has little impact on visual attention when the sound sources are exactly the salient objects in the video. But when the sound sources differ from salient objects, they are likely to draw attention. The emphasis, however, was on the effect of non-spatial audio on visual attention, and the videos were of a non-360° type.

A public dataset, consisting of head and eye movement data was presented in [4] by David et al. It was obtained from a free-view experiment of participants wearing a VR headset with an integrated eye tracker. The 360° videos were played without audio. In addition to the videos, the dataset had related gaze fixation and head trajectory data in the form of saliency maps and scan paths.

The eye tracking dataset proposed by Marighetto et al. in [8] included the eye positions obtained during four eye tracking experiments. Observers were recorded when exploring video between various audio conditions (with or without sound) and visual categories (moving objects, landscapes and faces). The authors reported that there was often a lower audiovisual dispersion than visual dispersion. The presence or absence of sound appeared to affect the spatial distribution of the eye gaze in some visual categories. Nevertheless, the media used in this analysis were non-360° videos with non-spatial audio.

Milesen et al. discussed in [10] how stereoscopic video and ambisonic sound related to the perceived QoE of the users. A questionnaire with ranking of subjective quality metrics was given to participants. The questionnaire was based on imagery, sound, presence, and motion sickness. Participants were asked to rate specific aspects of their experiences on a Likert scale. The research did not note the level of ambisonic sound used. Also, it did not aim to measure the visual or audiovisual attention of participants.

Considering existing literature, the novelty of the work presented in this paper lies in the evaluation of both VA and QoE, by capturing and analyzing implicit, subjective and objective data in 360° environments when accompanied with spatial audio.

III. Experiment Setup

This section provides details on the immersive multimedia system and the devices used to capture physiological responses. Details of the hardware and software components used for the experiment are described in Table I.

A. Sensing Technologies & Systems

1) 360° Video Player

The GoPro VR player [15] is a free 360° video player which is used to play 360° videos (on PCs and the HTC Vive or Oculus Rift). The VR player streams 360° video playback information such as camera orientation, video URL, playback status, and playback position to a port on the system on which it runs. The viewer can watch the 360° videos at any orientation which can be recorded as yaw, pitch and roll (refer Fig.1).

2) HMD and Pose Acquisition script

The HTC Vive with an integrated Tobii pro eye tracker was used for presentation of the 360° videos and also to capture participant gaze information. The Tobii Python SDK [16] was used to develop a script to accept details such as participant id, audio condition (non-spatial/spatial), video category (indoor/outdoor) and video sequence to play during the test. Upon execution, the script plays the selected 360° video sequence in the HMD using GoPro VR player and records the camera orientation (yaw, pitch and roll) from the HMD sensors.

The actual viewing angle is determined by how far the headset has rotated in relation to the 0° axis. Fig.1 gives a pictorial representation of the yaw, pitch and roll angle measurements based on participants head movement. For yaw, the 0° is parallel to the direction of the sensor and for pitch and roll, parallel to the ground. For yaw, the angle is measured by values between 0 and -179 when rotated to the left of the 0° axis, and by values between 0 and +179 when rotated to the right. Similarly, pitching up or rolling to the left for pitch and roll produces negative values, and pitching down or rolling to the right provides positive values. Initially, the velocity is represented as radians per second which is converted to degrees per second. Positive and negative velocity are the same directions as the angles.

3) E4 Wristband

The E4 wristband is a wearable device offering real-time physiological data acquisition and in-depth analysis and visualization software. In this work, the E4 was used to capture EDA and HR during the experiment.

B. 360° video presentation

Ten 360° videos with third order ambisonic sound were selected for the experiment. The selection was based on the duration of the recording; content; categories; resolution; and order of the ambisonic sound from the many files located at [18]. The videos were categorized broadly as Indoor and Outdoor (5 in each category). These were further subcategorized as Opera, Instrument, Riding and Exploration. The selected video files were processed using *ffmpeg* tool to; a) set the duration of each video to 60-sec; b) stitch the videos; c) convert ambisonic sound to stereo for the non-spatial audio experience . The videos were stitched together to create 5-min (300-sec) segments which were presented in random order to remove participants' bias. There were no narratives or subtitles in the videos. Fig.2 shows the representative frames for Indoor and Outdoor categories.

IV. Research Methodology

The assessment method employed in this research is experimental and is inspired by [23] [26] [27]. The method is also informed by ITU-T Recommendation P.913 [12].

A. Participants

A convenience sampling approach was used to recruit 20 participants for this study. They had an average age of 27 years with 11 males and 9 females. From the twenty participants who took part, eight had used VR before.

B. Assessment Protocol

The assessment was categorized into five key phases: a 10-minute informative phase; a 10-minute screening process; a 5-minute training phase; a 15-minute testing phase; and 5-10 minutes to answer a questionnaire. The participant was provided with an information sheet during the information phase which described the experiment. If the participant had queries regarding the experiment, these were answered before signing the consent form.

In the screening phase, the participant's visual and auditory acuity, and colour perception were assessed. A Snellen test was administered for visual acuity. Deficiencies in red green colour were screened using an Ishihara test. The auditory test available at [25] comprised of playing sounds in the frequency range 250-8000 Hz through the headphones. The accuracy of this hearing test is estimated at around 10 dBHL (decibels hearing level), enough to diagnose a mild, moderate or severe hearing loss. Upon completion of the visual and auditory tests, baseline metrics of HR and EDA were captured over a 5-min period using the E4 wristband.

In the training phase, the participant viewed a 360° video of 60-sec duration with non-spatial audio to get familiar and comfortable with the VR environment. Finally, in the testing phase, the participant viewed two 360° videos, each of which were 5-minutes in duration, with either stereo (ST) or spatial (third order ambisonic - HO) sound. One video was recorded in an Indoor setting and the other in an outdoor location. The sequence in which videos were played and the accompanying sound condition was randomized for every participant to eliminate possible biases. The HR and EDA data was captured throughout the assessment. In the interest of safety and to allow exploring the full 360° field of view, participants were seated in a rotating chair. The participant then completed a subjective questionnaire. On average, the assessment took 45-50 minutes.

TABLE I. Experiment Setup Components

Component	*Manufacturer/Origin*	*Used For*
HMD	HTC with Tobii Pro VR Integration [13]	Watching 360° videos
Headphones	Beyerdynamic DT 990 Pro [14]	Listening to non-spatial/ spatial audio
Wristband	Empatica E4 [17]	Recording EDA and HR
360° Player	GoPro VR Player [15]	Playing 360° videos to the HMD, obtaining head orientation as yaw, pitch and roll
360° videos	[18]	Audio-visual presentation to participants

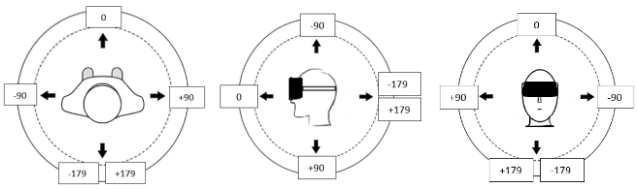

Fig. 1. Yaw, Pitch and Roll angles adapted from [19]

C. Questionnaire and Rating Scale

With inputs from [21] [22], a questionnaire [29] with twenty questions was developed to evaluate participants perception of presence (7 questions), immersion (7), and spatiality of sound (6) after watching the stimuli. The participants were asked to rate each question using the absolute category rating (ACR) system as outlined in [12]. The rating system used a five-point Likert scale to determine if a user agreed or disagreed with the statements.

V. Results and Discussion

This section outlines findings with respect to the objective, implicit and explicit data captured during the experiment.

A. Objective Metrics: Head Pose

For the head pose, data on 'pitch', 'yaw' and 'roll' were used to plot the subject's head orientation when watching the stimuli. A generated timestamp and position are used for synchronization. Since the head pose data is collected at a rate of 120 samples per second, for such a large sample, inspection of the individual data values does not provide a meaningful summary, and summary statistics are necessary. For this reason, the mean or average of the yaw, pitch and roll readings for each of the 20 participants, split across the two sound conditions (ST and HO) and categories (Indoor and Outdoor) were calculated using the IBM statistical analysis software package SPSS [24].

Table II presents the average of these mean values for the head pose data for comparison. From Table II, one can notice that the yaw is spread over a larger range across both ST and HO conditions, and for both Indoor and Outdoor categories. However, the yaw values are higher for the ST condition for both the Indoor and Outdoor categories as compared to the HO condition. It may possibly mean that participants were more focused in the HO condition due to realism and clarity of the sound as also confirmed by statistical analysis of self-reported measures (see Table IV).

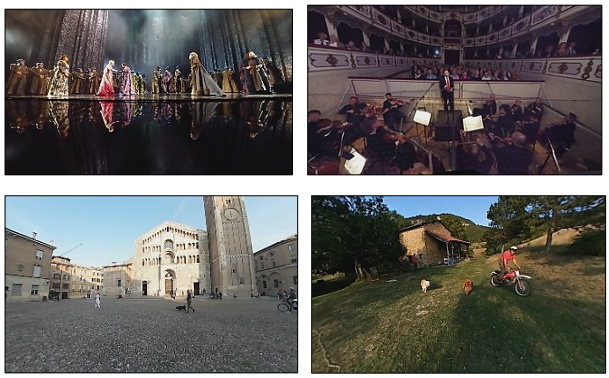

Fig. 2. Representative frames for Indoor and Outdoor scenes (Top – Opera Stage, Orchestra; Bottom – Town square with clock tower, Man riding a motorbike with two dogs following)

TABLE II. HEAD POSE DATA (MEAN VALUES) IN DEGREES

Category	Pitch	Roll	Yaw
Indoor Stereo	-11.167	-1.248	-14.206
Indoor_HighOrder	-6.705	0.091	-7.735
Outdoor Stereo	-0.395	-4.992	5.707
Outdoor_HighOrder	2.071	-0.673	2.918

Further, one can find that the mean value for pitch for the Outdoor HO condition is much higher than for the Outdoor ST condition. Videos in the Outdoor category had sound-emitting objects above the viewer's horizontal field of view, for e.g. a clock tower. Thus, a strong reason for this higher value could be the vertical sense of sound provided by the HO condition that prompted the participant to move their head vertically when viewing the videos. This finding prompted us to further investigate the distribution of the pitch angles and the head pose scan path for both sound conditions across the Indoor and Outdoor categories. Fig. 3 and 4 show the normal distribution of angles for pitch for the Indoor and Outdoor categories respectively. These are the minimum and maximum angles observed for each of the participants across the two sound conditions and categories, calculated using the IBM SPSS package.

1) Angle distribution

Videos in the Indoor category played an Opera with actors on an elevated platform and an orchestra performing below the platform. The camera was positioned between the platform and the orchestra. In the ST condition, the average pitch angle utilization was between -36° and -59° (up) and between 10° to 65° (down). For the HO condition, the utilization was between -26° to -59° and between 19° to 72°. Fig. 3 reveals that the pitch angles are less varied and more concentrated for the HO condition than for the ST condition which suggests that participants were more focused when the environment had spatial sound.

For the Outdoor category, the videos were exploratory in nature with no clear object of visual interest that the user should focus on. The videos had sound emitting objects, some of which were stationary (e.g. clock tower, person playing musical instrument while seated), while some moving (e.g. people talking while walking, ducks quacking while wading in water). In the ST condition, the average pitch angle utilization was between -27° and -76° and between 18° to 49°. For the HO condition, the utilization was between -21° to -56° and between 28° to 52°.

Fig.4 shows that that the pitch angles not only vary less but are more concentrated for the HO condition when looking in the upward direction. This means that the participants did not have to explore much in order to locate the source of the sound coming from above their heads. With regards to looking down, there is no obvious difference in angular distribution for both the ST and HO conditions. One possible reason could be that the ambisonic recording microphone and sound emitting objects were at the same level on the ground.

2) Head Pose Scan Path

Fig. 5 and 6 presents the complete head movement of participants with yaw and pitch angles as they explored the virtual environment. This essentially illustrates where users tend to focus their viewpoint in the non-spatial and spatial audio conditions and across the Indoor and Outdoor categories. The figures show diversity between both the sound conditions and scene categories and that they scan the environment quite differently. For videos of the Indoor category, one can find that the participant was more focused towards the front and looking slightly up when the video had HO sound. This is precisely the location where the actors were singing the Opera in the Indoor scenes.

Participants did explore the environment as revealed by the spread of the yaw angle and also looked down where the orchestra was playing as observed from spread of the pitch angle. However, the focus clearly was more towards the singers. In contrast, the head movement in the ST condition was quite scattered with no clear focus. Also, the spread of the yaw angle was more as compared to the HO condition. For videos of the Outdoor category, again, head movement of the participant was less scattered in the HO condition with pitch angles extending further in both upward and downward direction, possibly indicating attention towards the sound sources located above and below them.

Thus, from both the angle distribution and head pose scan path across the virtual environment, we can find that though the yaw-rotation is the most dominant rotation, the pitch-rotations vary across both sound conditions.

B. Implicit findings: HR and EDA

As outlined previously, the physiological metrics considered as part of this work were HR and EDA. The results are presented in Table III for EDA and HR respectively. Table III presents the mean values for each of the baseline, and testing phases, for both the ST and HO groups.

In Table III, we note that the average skin conductivity value increased across the baseline (0.642) to the testing (0.678) phase for the ST group. However, the HO group saw a decrease in value from 0.755 during the baseline phase, to 0.625 during the testing phase. This data suggests that the HO group experienced lesser stress compared to the ST group as they progressed from the baseline to the testing phase. Again, this could be possibly related to clarity of sound offered by the HO condition that could have made it easier for the participants to explore the 360° environment.

Table III also shows the readings for the average heart rate values for ST and HO groups during the baseline and testing phases. For the ST group, the average heart rate values were 95.35 and 90.25 beats per minute (BPM) for each of the two phases respectively. The minimum and maximum heart rate

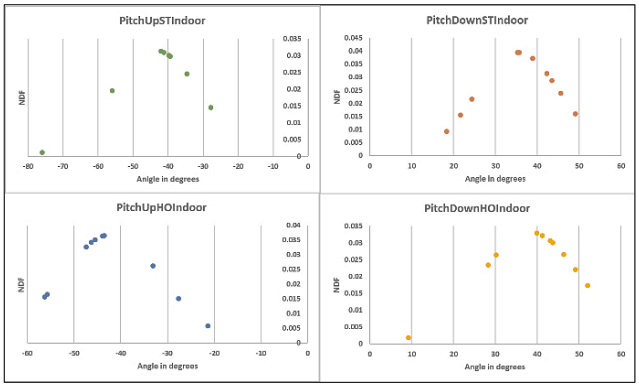

Fig. 3. NDF (normal distribution function) of angles for Pitch (Indoor category) for ST and HO conditions. Each value represents a participant.

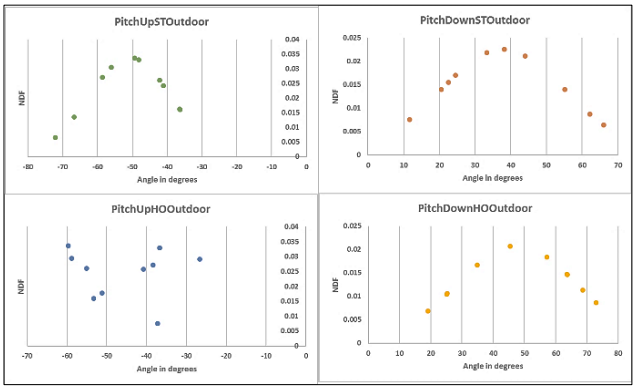

Fig. 4. NDF (normal distribution function) of angles for Pitch (Indoor category) for ST and HO conditions. Each value represents a participant.

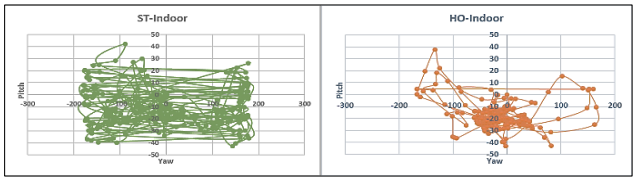

Fig. 5. Head movement (Indoor category) for ST and HO conditions

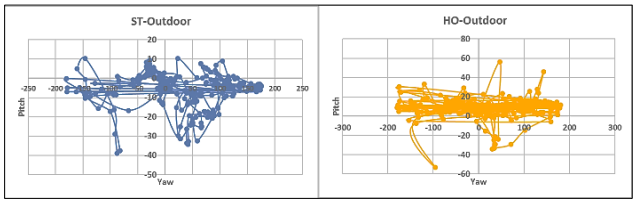

Fig. 6. Head movement (Outdoor category) for ST and HO conditions

values during each of these phases was 68.40/103.55 and 77.35/109.44. For the HO group, the average heart rate values were 88.79 and 85.137 BPM for the phases outlined. The minimum and maximum HO values during each of these phases was 66.51/105.58 and 75.88/109.05.

Both the ST and HO group's heart rate continued to drop across the two phases which was unexpected. This suggests that both groups acclimatized to their immersive experience without any significant difference.

C. *Explicit findings: Likert Questionnaire*

Table IV presents the significant results of the MOS self-reported measures captured via the post-test questionnaires discussed in Section III along with the statistical analysis.

TABLE III. AVERAGE EDA AND HR DURING THE BASELINE AND TESTING PHASES

	EDA		HR	
	Baseline	**Testing**	**Baseline**	**Testing**
Stereo	0.642	0.678	95.35	90.25
High Order	0.755	0.625	88.79	85.13

TABLE IV. STATISTICAL ANALYSIS OF SELF REPORTED MEASURES WITH 95% CONFIDENCE LEVEL

	ST		HO				
	MOS	**SD**	**MOS**	**SD**	***F***	***df***	***Sig. (2-tailed)***
Q16	4.0	0.471	4.4	0.471	4.160	18	0.087
Q18	3.6	0.843	4.3	0.483	2.881	18	**0.035**

Since the participants were part of two independent groups, an independent samples t-test was performed on the data with a 95% confidence level using the IBM statistical analysis software package SPSS. Of the twenty questions asked, only question 18, which asked participants about their perception of clarity of the sound, reported a statistically significant difference between the ST and HO groups with a two-tailed p value of $p=0.035$, $p<0.05$. The ST group reported a MOS rating of 3.6 whereas the HO group reported a MOS rating of 4.3 as per Table IV. Question 16 was the only other question that had a p-value of 0.087, making it close to being statistically significant. It aimed to discover the participants perception of realism in the sound while experiencing the system.

For each of the remaining questions, differences between the ST and HO group were not found to be statistically significant. Prior to the experiments, it was hypothesized that the HO experience group would have higher awareness of sound identification, localization, and clarity as well as higher level of presence and involvement than the ST group.

VI. CONCLUSION

In this paper, we have presented the findings of a study to 1) understand differences in head pose of users viewing 360° videos in non-spatial (ST) and spatial (HO) audio conditions; 2) understand the influence of both ST and HO audio on the user's QoE. Through assessing the differences of head pose data collected in ST and HO conditions, the yaw is spread over a larger range across both ST and HO conditions as compared to pitch and roll. However, the spread is lower for the HO condition, which could possibly mean that the participants were more focused in comparison to the ST condition due to the realism and clarity of the sound. We also report that the mean pitch value for Outdoor HO is much higher than for Outdoor ST. One reason for this could be the vertical sense of sound provided by the HO condition that prompted the participants to move their head vertically when viewing the videos. The analysis in terms of EDA, suggests that HO users experienced lesser stress compared to the ST group.

Further, the statistical analysis on the MOS self-reported measures reported a statistically significant difference between the ST and HO groups in terms of their perception of clarity of the sound.

In future work, we intend to add more sound conditions in addition to stereo and third order ambisonic sound for the current video categories. With this addition, we hope to get a wider comparison across sound conditions. Further, more categories of video could be introduced to understand whether presence of these sound conditions in different video content produces different viewing patterns. Also, it would be worth exploring the users head movement at different time intervals across the duration of the scene exploration. Future analysis will also consider the influence of various other factors on QoE such as age and gender.

ACKNOWLEDGEMENT

This research is funded by Athlone Institute of Technology President's Seed Fund and supported by Science Foundation Ireland and the ADAPT Centre under Grant Number 12/RC/2106.

REFERENCES

[1] Jung, Timothy & Tom Dieck, M. Claudia. (2017). Augmented Reality and Virtual Reality: Empowering Human, Place and Business.

[2] X. Corbillon, F. De Simone, and G. Simon, "360-degree video head movement dataset," in Proceedings of the 8th ACM on Multimedia Systems Conference, Taipei, Taiwan, June 2017, pp. 199–204, ACM.

[3] Lo, Wen-Chih & Fan, Ching-Ling & Lee, Jean & Huang, Chun-Ying & Chen, Kuan-Ta & Hsu, Cheng-Hsin, (2017), 360° Video Viewing Dataset in Head-Mounted Virtual Reality, 211-216.

[4] David, E.J., Gutiérrez-Cillán, J., Coutrot, A., Silva, M.P., & Callet, P.L., (2018), A dataset of head and eye movements for 360° videos, MMSys '18.

[5] Coutrot, Antoine & Guyader, Nathalie & Ionescu, Gelu & Caplier, Alice, (2013), Video viewing: Do auditory salient events capture visual attention?, Annals of Telecommunications. 69. 1-9.

[6] Coutrot, Antoine & Guyader, Nathalie & Ionescu, Gelu & Caplier, Alice, (2012), Influence of soundtrack on eye movements during video exploration, Journal of Eye Movement Research, 5, 1-10 .

[7] Min, Xiongkuo & Zhai, Guangtao & Gao, Zhongpai & Hu, Chunjia & Yang, Xiaokang, (2014), Sound influences visual attention discriminately in videos, 2014 6th International Workshop on Quality of Multimedia Experience, QoMEX 2014, 153-158

[8] Marighetto, Pierre & Coutrot, Antoine & Riche, Nicolas & Guyader, Nathalie & Mancas, Matei & Gosselin, Bernard & Laganiere, Robert, (2017), Audio-Visual Attention: Eye-Tracking Dataset and Analysis Toolbox.

[9] Akhtar, Zahid & Siddique, Kamran & Rattani, Ajita & Lutfi, Syaheerah & Falk, Tiago, (2019),Why is Multimedia Quality of Experience Assessment a Challenging Problem?,IEEE Access, PP.1-1.

[10] Milesen, Madsen & Lind, V., 2017. Quality Assessment of VR Film - A Study on Spatial Features in VR Concert Experiences. Masters. Copenhagen: Aalborg University.

[11] ZOTTER, F., & FRANK, M. (2019). Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality.

[12] International Telecommunications Union. 2016. P.913: Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment. [ONLINE] Available at: https://www.itu.int/rec/T-REC-P.913/en. [Accessed 28 May 2020].

[13] Tobii Pro. 2018. Tobii Pro VR Integration – based on HTC Vive Development Kit Description. [ONLINE] Available at: https://www.tobiipro.com/siteassets/tobii-pro/product-descriptions/tobii-pro-vr-integration-product-description.pdf/?v=1.7.[Accessed 28 May 2020].

[14] Beyerdynamic. 2020. Beyerdynamic DT990 Pro. [ONLINE] Available at: https://europe.beyerdynamic.com/dt-990-pro.html. [Accessed 28 May 2020].

[15] GoPro. 2020. GoPro VR Player for Desktop FAQ. [ONLINE] Available at: https://gopro.com/help/articles/block/gopro-vr-player-for-desktop-faq. [Accessed 28 May 2020].

[16] Tobii Pro. 2020. Tobii Pro SDK - Develop eye tracking applications for research. [ONLINE] Available at: https://www.tobiipro.com/product-listing/tobii-pro-sdk/. [Accessed 28 May 2020].

[17] Empatica. 2019. E4 Wristband - Empatica Support. [ONLINE] Available at: https://support.empatica.com/hc/en-us/categories/200023126-E4-wristband. [Accessed 28 May 2020].

[18] Angelo Farina. 2020. Index of /Public. [ONLINE] Available at: http://www.angelofarina.it/Public/. [Accessed 28 May 2020].

[19] Almquist, Mathias and Viktor Almquist, "Analysis of 360° Video Viewing Behaviours", (2018).

[20] "Qualinet White Paper on Definitions of Quality of Experience (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Patrick Le Callet, Sebastian Möller and Andrew Perkis, eds., Lausanne, Switzerland, Version 1.2, March 2013."

[21] J. M. Rigby, S. J. J. Gould, D. P. Brumby, and A. L. Cox, Development of a questionnaire to measure immersion in video media: The Film IEQ, TVX 2019 - Proc. 2019 ACM Int. Conf. Interact. Exp. TV Online Video, pp. 35–46, 2019.

[22] U. C. Lab, "Sheet_PRESENCE QUESTIONNAIRE(PQ)," 2004.

[23] Conor Keighrey, Ronan Flynn, Siobhan Murray and Niall Murray, (2017), A QoE Evaluation of Immersive Augmented and Virtual Reality Speech & Language Assessment Applications, June 2017, in Erfurt, Germany.

[24] "IBM SPSS - IBM Analytics," IBM, [Online]. Available: https://www.ibm.com/analytics/us/en/technology/spss/. [Accessed 28 May 2020].

[25] D. Pigeon, "Online Hearing Test & Audiogram Printout", Hearingtest.online, 2020. [Online]. Available: https://hearingtest.online/. [Accessed: 28- May- 2020].

[26] E. Hynes, R. Flynn, B. Lee and N. Murray, "A Quality of Experience Evaluation Comparing Augmented Reality and Paper Based Instruction for Complex Task Assistance," 2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP), Kuala Lumpur, Malaysia, 2019, pp. 1-6

[27] D. Egan, S. Brennan, J. Barrett, Y. Qiao, C. Timmerer and N. Murray, "An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments," in Quality of Multimedia Experience (QoMEX), 2016

[28] Poeschl-Guenther, Sandra & Wall, Konstantin & Doering, Nicola, (2013), Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence, Proceedings - IEEE Virtual Reality, 129-130

[29] https://drive.google.com/file/d/1EnPjAPHFqCMPWceJ2iL9nh2o_FZhVzDo/view?usp=sharing