2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)

Evaluating the User in a Sound Localisation Task in a Virtual Reality Application

Adrielle Nazar Moraes[1], Ronan Flynn[1], Andrew Hines[2], Niall Murray[1]

Department of Electronics and Informatics[1]
Athlone Institute of Technology
Athlone, Ireland

University College Dublin[2]
School of Computer Science
Dublin, Ireland

email: a.nmoraes@research.ait.ie[1], rflynn@ait.ie[1], andrew.hines@ucd.ie [2], nmurray@research.ait.ie[1]

***Abstract*— Virtual reality (VR) has proven to be a powerful tool enabling the development of immersive multimedia experiences. Initially focused on entertainment, industry and academia have begun to adapt and develop immersive applications for the healthcare domain, with opportunities in terms of condition assessment, diagnosis and intervention. In the context of immersive applications, audio, and in particular spatial audio, plays an important role on the immersion level. In order to process this information, the auditory cortex uses spatial cues encoded in the sound to provide relevant information about the distance, intensity and direction of the sound source. However, many different types of listening disorders can affect this capability. One condition, central auditory processing disorder (CAPD), significantly affects a user's ability to discriminate between different sound sources. People who suffer with this condition, are incapable of processing sounds properly, which may be stressful and frustrating when doing tasks with complex sounds or in noisy environments. This can have a significant impact on a person's quality of life.**

In this paper, an immersive VR spatial audio application is presented. It enables us to evaluate the ability of users to specify or localise the source of a sound. An integrated sensing system continuously collects relevant data from the user in order to fully understand how to quantify and evaluate spatial auditory skills from a quality of experience (QoE) perspective. QoE gives insight into a user's state and behaviour. To perform a detailed QoE evaluation of the listening task, implicit and explicit metrics were collected from the user. These included: self-reporting questionnaires, localisation performance, and physiological metrics. Data collected from this QoE evaluation gives an insight into a user's abilities to localise sound sources in VR, and also provides information on behaviour and effort (workload) in performing the task.

Keywords—QoE, spatial audio, CAPD, sound localisation, virtual reality

I. INTRODUCTION

Virtual reality technology is fast becoming the accepted platform for delivering high quality immersive audio-visual experiences [1]. It has the potential to enable truly personalised experiences due to its adaptability and flexibility. Although a lot of the discussion on VR has focused on the visual aspect of the experience, but audio also plays a decisive role in terms of user immersion [2]. With the advancement of audio technologies, it is possible to render and spatialise audio with a high level of quality, matching human resolution to discriminate sounds in terms of depth, elevation and lateralisation [3]. Spatialised audio is very important in industries like gaming and for communication tasks, and in environments that require humans to constantly give attention to a specific sound in space [4][5][6].

Spatial audio is accurately reproduced in VR using a set of head-related transfer functions (HRTFs). These functions are a mathematical representation of the sound in space. HRTFs contain information about the magnitude and phase shift of a sound [5]. HRTFs are strongly dependent on the localisation of the sound source and the spectral information of both ears [7]. In addition, a HRTF also contains relevant information for sound localisation on the azimuth plane and represents the differences between the right and left ears. The interaural level difference (ILD) represents the difference in amplitude of the sound at both ears and the interaural time difference (ITD) represents the difference in the arrival time of the sound at each ear [3]. These characteristics are important in terms of how sounds are perceived. For instance, if the sound is coming from the right, it is expected that the right ear would perceive the sound faster and louder than the left ear. However, if the sound is coming from the front or back of the listener, both the ILD and ITD can assume the same value. For these two cases, a third cue is used by the brain to eliminate confusion as to whether the sound is coming from the front or back. This cue is specific to each person as it is related to spectral cues encoded in the shape of the pinnae and head [8]. In addition to front-back confusion, there is the phenomenon of localisation blur, which can occur when the listener perceives the sound source to be at a location close to the actual sound source [9]. Given both error conditions, the challenge for spatial audio is how to generate generic functions to simulate accurately the acoustic environment. In [10] it has been shown that it is possible to reproduce sound with high reliability using generic HRTFs instead of having to create personalised ones for each individual, which is a time consuming and expensive task.

In addition to the application domains of gaming and communication, there are also opportunities for VR and spatial audio in health from intervention and diagnostic perspectives. For example, central auditory processing disorder (CAPD) is an umbrella term to describe the brain's inability to process complex sounds [11]. It can affect a listener's auditory memory, auditory sequencing, and the ability to distinguish and locate sounds in space [12]. Consequently, this has a negative influence on a person's quality of life (QoL). It can be frustrating to perform daily tasks such as driving, going to school or to a place with many sound sources. Since it is an umbrella term, there is no standard test to diagnose this condition. However, it is possible to perform a battery of tests in order to evaluate auditory processing skills considering different listening conditions. For the sound localisation task, the listen in

978-1-7281-5965-2/20/$31.00 ©2020 IEEE

spatialized noise (LISN) test was developed [13], it is illustrated in Fig. 1.

In this context, this work presents an evaluation of an immersive VR and spatial audio-based application, inspired by LISN, to evaluate a user's ability to localise sounds. The goal of the LISN test (as per [13]), is to evaluate, identify and understand the target speaker's voice, while ignoring distractors. The test is composed of several target sentences and the listener's threshold is measured by lowering or raising the target volume according to the listener's performance localising the sound. For this test, administered with headphones, the target is always positioned in front of the listener and the distractors appear at the same position as the target, or at ±90 degrees apart from it. The immersive spatial audio application implemented here achieves and extends this test by increasing the number of possible positions for the sound sources from 3 to 24. Additionally, the sound sources are positioned over a full range of 360 degrees.

To evaluate the proposed spatial audio-based application, the quality of experience (QoE) framework was employed. QoE aims to measure and evaluate the user's degree of annoyance or enjoyment after or while experiencing an application or service [14]. It has been used in many applications to give insight into user behaviour and state. This can be accomplished by the measurement of different metrics from the user: self-reporting questionnaires, performance data and physiological data [15]. This data can be useful for health care, as it provides additional information as an assessment or diagnosis support tool. The use of advanced sensor systems also allows researchers to generate reports with data collected from the user in a continuous and ecologically valid manner. The focus of the work presented in this paper is threefold: (i) to assess a user's ability to localise spatial auditory sources; (ii) to compare two different VR interaction methods while performing the assessment; (iii) to understand user QoE of these two VR interaction methods through user performance metrics and explicit metrics captured by post experience questionnaires.

This work presents the results of a QoE evaluation of a novel VR application designed to assess spatial auditory ability in the presence of distractors. The following sections explain how the experimental methodology was adapted from the standard assessment tools used in the health domain. The results and discussion provide insights into the users ability to localise sound sources as well as the user QoE of the different interaction methods evaluated.

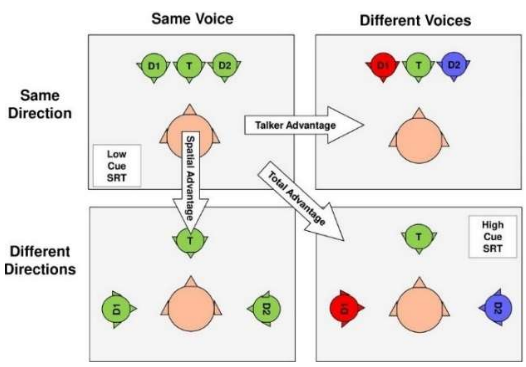

Fig. 1: LISN test conditions. For each condition there is a specific cue. When it comes to the different voices and directions, the threshold is expected to be lower (high cue speech related threshold (SRT)) when compared with the other conditions.

II. Related Work

The aim in [16] was to develop a VR application to perform training on sound localisation ability using a generic HRTF. The experimental design included 17 subjects, a tutorial phase and three main phases with 48 trials each. The pre-test (phase 1) was used as a reference for comparison between possible improvements in the post-test (phase 3). Training (phase 2) was the only phase with visual feedback in order to assist users with the localisation task, which included a visual representation of the 24 possible locations of the sound sources surrounding the listener. The authors reported that it was possible to improve the auditory localisation ability with training, achieving a smaller error angle and reaction time for the post-test phase in comparison to the pre-test phase.

The work presented in [17] reported a methodology and results of an evaluation on the effects of multimodal stimuli (visual and audio) on sound localisation abilities. Two experiments, with ten participants, were conducted using individualised HRTFs for each subject. The first experiment was conducted to investigate the effects of using VR environments with non-expert subjects on the localisation task. Listeners were asked to locate the target until 95% of the targets were located, to a precision within a root-mean-square error angle of 2 degrees, in three consecutive blocks of 100 trials. Participants needed more than 500 trials to achieve the goal. The second experiment investigated the influence of training. The task was divided in blocks of 100 trials, with up to 2,200 trials to be completed over a span of 20 days on average. The authors discovered that when listeners were immersed in a VR environment, the number of front-back confusions decreased, and the horizontal precision increased. However, significant improvements were observed only for the first 400 trials.

In [18], the researchers aimed to develop a VR application to evaluate user behaviour in response to spatial audio. The experiment included 29 children diagnosed with autism spectrum disorder (ASD). They were divided into two groups, spatialised and non-spatialised audio. Participants were free to move around a tracked space of 1.6m x 1.6m while wearing the head mounted display (HMD). The experiment consisted of 8 auditory events distributed in the VR environment. The performance was measured by the distance between the listener and the sound source. The authors concluded that spatialised audio significantly improved the performance on the localisation task. In addition, a multimodal experience with visual and audio cues increased the positive outcomes of VR-based therapy for persons diagnosed with ASD.

Considering the related work, the novelty of this work lies in: (i) evaluating localisation abilities of spatial audio in VR with distractors; (ii) a comparison of interaction methods to make the selection using explicit and objective performance metrics to accomplish this goal. The latter includes questions to understand interaction, usability and immersion, as well as the NASA TLX questionnaire to understand the mental workload. In addition, the implicit metric of physiological data was also recorded, including a set of behavioural data (eye tracking, head position and user performance).

III. EXPERIMENT SETUP

A. Virtual reality and spatial audio display technology

The virtual environment was designed using a Unity game engine (version 2018.2.15f1) [19] with a Steam Audio set of generic HRTFs [20] to render the spatialised audio. The immersive HMD used was the HTC VIVE with Tobii eye tracking integrated [21]. This headset has a field of view of 110°, refresh rate of 90Hz and a resolution of 1,080 x 1,200 pixels per eye. The audio was reproduced by Beyerdynamic DT 990 PRO Studio Headphones [22], with diffuse-field equalisation. The system was designed to run wirelessly in order to allow the listener to naturally interact (look around) with the virtual environment, with 6 degrees of freedom (DoF), and search for the correct sound source. Before the beginning of the tests, the headset initial position was calibrated with the same reference point for all participants. Additionally, headset adjustment and eye calibration were performed to ensure that the content was displayed correctly.

The VR environment, as shown in Fig. 2, consisted of an open field with 24 equally spaced spheres (15º apart from each other) on a circle of radius 10m surrounding the listener. The objective of the evaluation was for the user (positioned at the centre of the circle of spheres) to identify the sphere from which the sound was coming. The source selection involved two potential interaction modes, namely, eye gaze pointing and wand pointing.

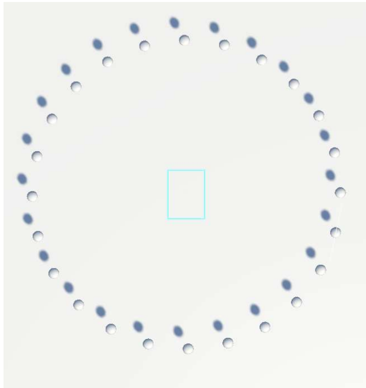

Fig. 2: Top view of the VR environment designed for the system. The listener has to locate the sound in space by pointing with either eye gaze or with the controller after a sound stimulus is reproduced in the VR environment. Each sphere represents one of the possible sound source positions.

B. Interaction methods

One of the experiment aims was to evaluate user interaction with the system. This involved comparing two interaction methods: eye gaze pointing and pointing with a controller (the wand).

1) *Eye gaze pointing (GP):* to perform the interaction with the VR environment, the first option to be evaluated was to use the data from the eye gaze sensor. The user task selected the option by looking at the sphere and then confirmed the selection by clicking on the HTC VIVE wand controller button.

2) *Wand pointing (WP):* the second interaction option involved the user making their selection using the wand only. They achieved this by pointing at the sphere they chose to select with the wand (a laser beam was emitted from the controller) and then confirmed the selection by clicking the controller button.

C. Sound localisation assessment

The immersive VR sound localisation task consisted of three testing phases, inspired by the LISN test:

1) *Target only.*

2) *Target + distractor (same location).*

3) *Target + distractor (±90º apart from target).*

Each testing phase is illustrated on Fig. 3 and the order of the testing phases was the same for all participants. The user was required to select the sound source a total of 72 times (24 trials for each testing phase). The experiment protocol was designed to guarantee that each sphere will reproduce a sound 3 times in total, one for each testing phase. In order to evaluate localisation blur and front-back confusions, the order and location of the sound stimulus is randomized but considered the previous stimulus source. The sound stimulus is a 2.5 s duration white noise presented at a constant sound pressure level of 55 dB SPL. The distractor sound stimulus consists of a low-frequency, 1,500 Hz pure-tone sine wave presented at the same SPL as the target stimulus. This frequency value represents the cut off frequency that best describes the interaural level difference and interaural time different cues [23]. Each testing phase represents a different challenge for the listener in terms of spatial localisation and target discrimination.

D. Sensor system

After the screening process (see Section IV), participants were asked to wear the E4 Empatica wristband [24] in order to collect physiological data (electrodermal activity, interbeat interval, temperature and acceleration). This data is used to measure the participant's level of arousal and valence throughout the experiment.

During the experiment, gaze data (pupil size, gaze direction and head pose) is sampled at a rate of 120 Hz. In addition, the listener's performance data recorded includes: the number of selection attempts; the number of correct selections; the informed location of the target; the real position of the target and the completion time for each sound source.

IV. ASSESSMENT METHODOLOGY

The experimental assessment method is composed of 6 phases, an approach inspired by [25] and [26]. Ethical approval for this study was obtained prior to participant recruitment. In phase one, all information about the test is given to the participant with a general explanation of how the system works and the objectives of the experiment. After this procedure, the participants were screened in phase two. This phase included visual and listening screening. Phase three was a 5-minute resting period for baseline physiological data collection. Subsequently, the listener was given video-based training (phase four) on how to interact with the VR environment. After that, the volunteer was asked to wear the headset and the headphones. They experienced a tutorial scene with the goal of learning how to interact with the system and using the controllers to select a sphere. After the user has become familiar with the system, the test begins (phase five). The entire test took on average 43 minutes, with an average test phase duration of 17 minutes. After the listening task was completed, volunteers were required to

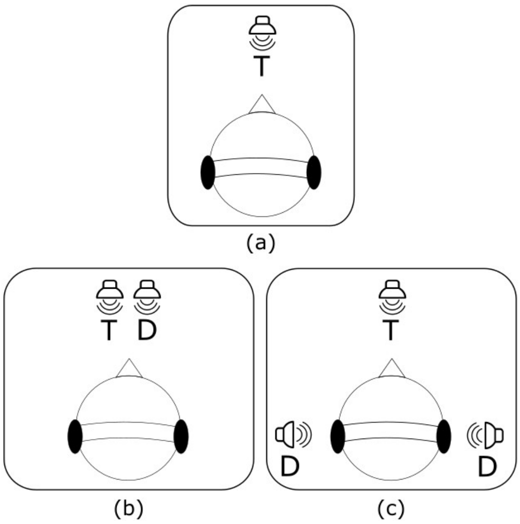

Fig. 3: Test phases: (a) target (T) only, (b) target and distractor (D) at the same location and (c) target and distractor separated by 90 degrees. The order of the test phases is the same for all of the participants. The testing phases are repeated 3 times for each sound source position.

complete a questionnaire related to their experience (phase six).

A. Participant screening

A total of twenty subjects took part in the experiment (10 in each interaction group). The average age was 29 years, with 8 female and 12 male subjects. Six of the participants had never experienced VR before. Each listener was screened at the beginning of the test to check for hearing impairments with an online tool developed by WIDEX [27]. In addition, participants were screened for visual impairments using a Snellen chart [28] and an Ishihara test [29] for colour blindness.

B. Laboratory Design

The evaluation of both interaction methods (eye gaze pointing and wand pointing) took place in the same laboratory. The room was prepared to run a wireless application thus allowing movement with 6 DoF. Before the test began, volunteers were asked to sit in a swivel chair. They were free to rotate in the chair but not to move the chair around the room. In addition, listeners were able to move their head during the experiment to locate the sound source.

C. Questionnaire

At the end of the experiment, listeners were asked to complete a questionnaire reporting their experience. In addition, they were also required to complete the NASA TLX questionnaire [30] to evaluate workload. Table I presents the category target for each of the presented questions. The questions are under three categories: interaction with the system [31], usability [32] and presence [33]. The chosen categories are key elements for evaluating QoE. The CAPD test is applied to measure cognitive abilities related to auditory memory and auditory attention. The NASA TLX questionnaire also evaluates the physical and mental demand from the user perspective. Since sound processing is a mental task (processed by the brain), there is a relationship between the outcomes of the questionnaire and the outcomes of a CAPD assessment.

V. Results and Discussion

This section of the paper presents the findings of this research, discussing the data collected from the questionnaires and the performance metrics.

Fig. 4 shows the distribution of mean opinion scores (MOS) for the questionnaire responses. Both the eye gaze pointing and wand pointing groups had similar responses for the questionnaire and no statistically significant differences between the group means (tested with an independent samples test with a 95% confidence level). However, users reported a good experience with the system in terms of learning how to use the controllers. The results for Q8 in Table I, related to the interaction aspect, indicate that for both interaction methods (eye gaze and wand) the task was easy to perform. In terms of the usability aspects, users reported the task to be easy to follow. As observed from the responses to Q9, listeners were confident enough to perform the localisation task again, being able to predict what would happen after a sound source is reproduced. Presence was expected to be high, since the system was built to be wireless and to allow 6 DOF. This is supported by the results from Q4, which show that users were not aware of their surroundings, being able to concentrate on the localisation task.

TABLE I. QUESTIONNAIRE CLASSIFICATION

Question	*Interaction*	*Usability*	*Presence*
Q1 - The task was complex		*x*	
Q2 - The system was difficult to use	*x*	*x*	
Q3 - I needed to learn a lot of things before I could complete the task	*x*		
Q4 - How aware were you of the real world surrounding while in the virtual world?			*x*
Q5 - It was not difficult to localise the sound source with distractors	*x*	*x*	
Q6 - I had no idea where the sound was coming from			*x*
Q7 - I was restricted in my movements using the system	*x*		*x*
Q8 - The interaction with the system was natural	*x*		*x*
Q9 - I was able to anticipate what would happen in response to the action that I performed		*x*	*x*
Q10 - I felt qualified enough in interacting with the virtual environment at the end of the experience		*x*	
Q11 - The spheres distracted me from performing the task			*x*

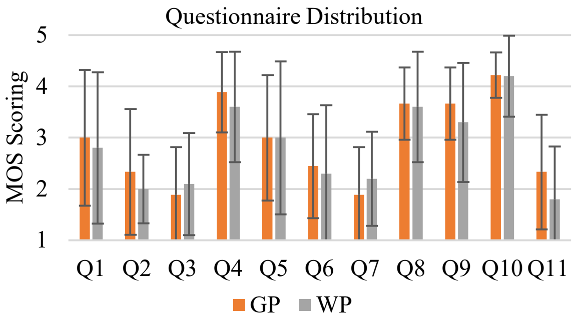

Fig. 4: Questionnaire distribution of MOS scores for the experiment for the eye gaze pointing (GP) and the wand pointing (WP) groups. Each boxplot is related to the questions in the Table I. The vertical values are the range of options from 1 (strongly agree) to 5 (strongly disagree).

The time-to-complete values for the test were evaluated in order to give insight into the learning curve for the localisation task (see Fig. 5). An interesting result is related to the trend of the completion time mean values across the three conditions. Since the complexity is increasing across the phases, it was expected that the time to complete each phase would increase. However, this number has decreased, a possible indication that listeners learned how to perform the localisation task. This finding can also be supported by the questionnaire responses related to the interaction factor (Q2, Q3, Q5, Q7 and Q8). The results from Table II represent the average error angles for each group along the testing phases. This table shows that both of the groups had lateralisation errors (angle error < 45°). Additionally, these values indicate that participants missed the target by one (15°) or two spheres (30°). The results from the table also indicate that front-back confusions were not frequent across the tests, with the majority of localisation errors classified as localisation blur.

In order to measure workload during the completion of the test, a post-test NASA-TLX questionnaire was completed. An independent samples test with a 95% confidence level was run to check for differences between means for both interaction groups. Results from the test are given in Fig. 6, with the impact percentage for each of the six workload factors (physical, temporal, performance, effort, frustration and mental). The mean for each group was similar for all the workload factors with the exception of the temporal factor. The temporal factor was found to be more relevant for the WP (58% ±23.11) interaction method than the GP (30.5% ±26.29) method, $t(18) = 2.484$, $p = 0.023$, $p < 0.05$. The temporal factor is related to the time pressure to complete the required task [30]. This result suggests that listeners from the GP group were comfortable with the time given to perform the localisation. This result is also justified by the time-to-complete values for the GP group, which were on average lower than the WP group.

According to the literature, the sound localisation task requires a high level of processing from the brain [11]. Therefore, the mental workload factor values were high for both groups, as expected given the nature of the task. For both groups, the frustration level was considered low. This result suggests that users were aware of the sound localisation by the system, since there was no feedback from their responses during the test.

The average measurements for electrodermal activity (EDA) for each group are given in Fig. 7.

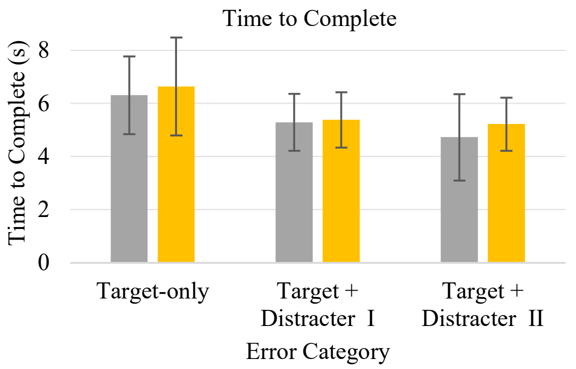

Fig. 5: Time to complete task for each selection by the eye gaze pointing (GP) and wand pointing (WP) groups.

Additionally, data was divided to indicate EDA measurement behaviour across test phases. The datapoints are a result of the normalization for each subject on a scale from 0 to 100% of the participant's maximum EDA level. According to the results, there was an increasing value of participant's EDA for both interaction groups. From the QoE perspective, this result indicates that there was an increase in the arousal level across the test [15], suggesting that participants were engaged and immersed in the system.

TABLE II. AVERAGE ERROR ANGLES FOR EACH GROUP (EYE GAZE POINTING (GP) AND WAND POINTING (WP)) ACROSS TESTING PHASES

	Testing Phase 1	*Testing Phase 2*	*Testing Phase 3*
GP	30 ±22°	31 ±22°	35±22°
WP	37±14°	28±17°	33±18°

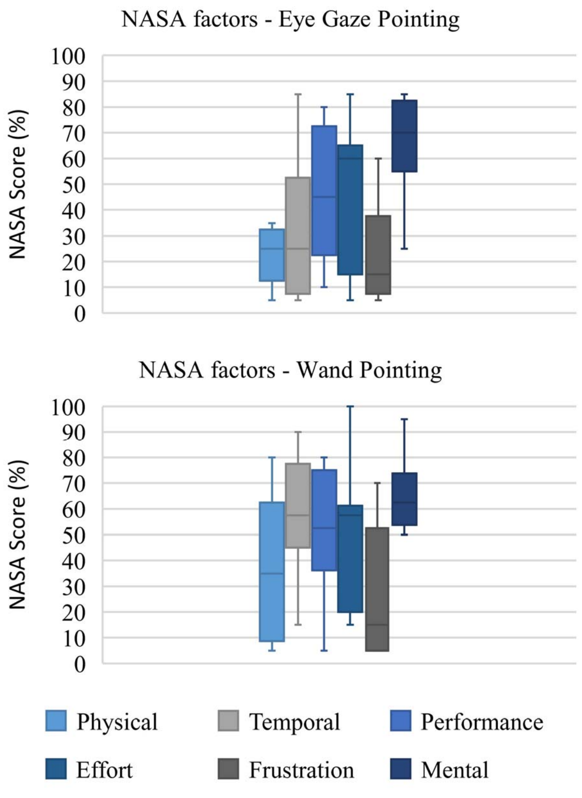

Fig. 6: NASA factors distribution for the eye gaze pointing group and the wand pointing group. Each bar represents one of the workload factors.

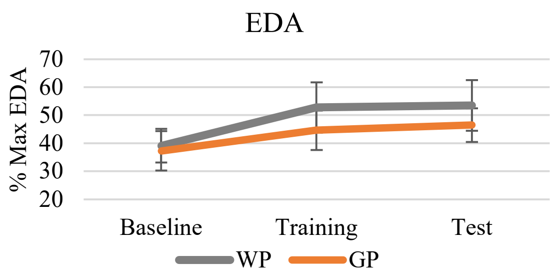

Fig. 7 EDA measurement for each interaction group. The graph contains values ranging from the beginning of the experiment, training phase and for the complete test (including each testing phase).

VI. Conclusion

Other than pilot studies of VR and speech localisation [34], there is a lack of subjective evaluation for immersive spatial audio applications in the literature and, in particular, very few works have considered or assessed spatial auditory factors and localisation abilities. This work presents an assessment methodology for auditory localisation abilities in VR. This application can be expanded to the health domain as a tool to improve assessment methodologies for people diagnosed with auditory processing disorders. Results from the experiments, show that users were able to understand the task and learn how to use the system. They also report that the two interaction methods evaluated, eye gaze pointing and wand pointing, are quite comparable. Results from the implicit metrics indicate an increasing level of immersion across test phases. Further investigation on the data collected during the experiment will be performed in order to improve the outcomes from the system regarding head movement analysis and performance. In addition, an analysis of the physiological and gaze data will be conducted to compare the benefits of the visual stimuli in VR.

Acknowledgment

This Research is supported by Science Foundation Ireland and the ADAPT Centre. It is co-funded by the European Regional Development Fund under Grant Number 12/RC/2106.

References

[1] C. Keighrey, R. Flynn, S. Murray, S. Brennan, and N. Murray, “Comparing User QoE via Physiological and Interaction Measurements of Immersive AR and VR Speech and Language Therapy Applications,” in *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*, 2017, pp. 485–492.

[2] T. Walton, “The Quality Of Experience Of Next Generation Audio: Exploring System, Context And Human Influence Factors,” Newcastle University, 2018.

[3] J. Broderick, J. Duggan, and S. Redfern, “The Importance of Spatial Audio in Modern Games and Virtual Environments,” in *2018 IEEE Games, Entertainment, Media Conference (GEM)*, 2018, pp. 1–9.

[4] G. Llorach, G. Grimm, M. M. E. Hendrikse, and V. Hohmann, “Towards Realistic Immersive Audiovisual Simulations for Hearing Research: Capture, Virtual Scenes and Reproduction,” in *Proceedings of the 2018 Workshop on Audio-Visual Scene Understanding for Immersive Multimedia*, 2018, pp. 33–40.

[5] O. Rummukainen, S. Schlecht, and A. Plinge, “Evaluating Binaural Reproduction Systems from Behavioral Patterns in a Virtual Reality—A Case Study with Impaired Binaural Cues and Tracking Latency,” in *Audio Engineering Society Convention 143*, 2017.

[6] R. Kilgore, M. Chignell, and P. Smith, “Spatialized audioconferencing: What are the benefits?,” *Cent. Adv. Stud. Collab. Res.*, pp. 135–144, 2003.

[7] C. C. Berger, M. Gonzalez-franco, A. Tajadura-Jiménez, D. Florencio, and Z. Zhang, “Generic HRTFs may be good enough in virtual reality. Improving source localisation through cross-modal plasticity,” *Front. Neurosci.*, vol. 12, 2018.

[8] K. D. Gardner,William G. and Martin, “HRTF measurements of a KEMAR,” *J. Acoust. Soc. Am.*, vol. 97, 1995.

[9] M. Cohen-McFarlane, R. Goubran, and F. Knoefel, “Design of a System to Measure Spatial Sound Localisation Abilities,” in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)* , 2018, pp. 1–6.

[10] D. N. Zotkin, R. Duraiswami, and L. S. Davis, “Rendering localised spatial audio in a virtual auditory space,” *IEEE Trans. Multimed.*, vol. 6, no. 4, pp. 553–564, Aug. 2004.

[11] American Academy of Audiology, “Diagnosis , Treatment and Management of Children and Adults with Central Auditory Processing Disorder.” American Academy of Audiology, 2010.

[12] G. Stavrinos, V. M. Iliadou, L. Edwards, T. Sirimanna, and D. E. Bamiou, “The relationship between types of attention and auditory processing skills: Reconsidering auditory processing disorder diagnosis,” *Front. Psychol.*, vol. 9, 2018.

[13] S. Cameron, H. Dillon, and P. Newall, “Development and evaluation of the listening in spatialized noise test,” *Ear Hear.*, vol. 27, no. 1, pp. 30–42, 2006.

[14] S. Moller and A. Raake, *Quality of Experience - Advanced Concepts, Applications and Methods*. Springer, 2014.

[15] U. Engelke *et al.*, “Psychophysiology-Based QoE Assessment: A Survey,” *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 1, pp. 6–21, Feb. 2017.

[16] Y.-H. Wu and A. Roginska, “Analysis and training of human sound localisation behavior with VR application,” in *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*, 2019.

[17] P. Majdak, M. J. Goupell, and B. Laback, “3-D localisation of virtual sound sources: Effects of visual environment, pointing method, and training,” *Attention, Perception, Psychophys.*, vol. 72, no. 2, pp. 454–469, 2010.

[18] D. Johnston, H. Egermann, and G. Kearney, “Measuring the Behavioral Response to Spatial Audio within a Multi-Modal Virtual Reality Environment in Children with Autism Spectrum Disorder,” *Appl. Sci.*, vol. 9, 2019.

[19] Unity Technologies, “Unity 3D.” [Online]. Available: https://unity.com/. [Accessed: 27-Jan-2020].

[20] Valve, “Steam Audio.” [Online]. Available: https://valvesoftware.github.io/steam-audio/. [Accessed: 25-Nov-2019].

[21] HTC Corporation, “HTC Vive.” [Online]. Available: https://www.vive.com/us/product/vive-virtual-reality-system/. [Accessed: 27-Jan-2020].

[22] BeyerDynamic, “DT 990 PRO.” [Online]. Available: https://europe.beyerdynamic.com/dt-990-pro.html. [Accessed: 24-Jan-2020].

[23] T. Potisk, “Head-Related Transfer Function.” University of Ljubljana, Jan-2015.

[24] Empatica, “E4 sensors.” [Online]. Available: https://www.empatica.com/en-gb/research/e4/. [Accessed: 27-Jan-2020].

[25] C. Keighrey, R. Flynn, S. Murray, and N. Murray, “A Physiology-based QoE Comparison of Interactive Augmented Reality, Virtual Reality and Tablet-based Applications,” *IEEE Trans. Multimed.*, 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,” in *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*, 2019, pp. 1–6.

[27] Widex, “Online Hearing Test.” [Online]. Available: https://global.widex.com/en/online-hearing-test#3. [Accessed: 24-Jan-2020].

[28] American Academy of Ophthalmology, “Eye Chart.” [Online]. Available: https://www.aao.org/eye-health/tips-prevention/eye-chart-facts-history. [Accessed: 27-Jan-2020].

[29] “Ishihara Test for Color Blindness.” [Online]. Available: https://www.color-blindness.com/color-blindness-tests/. [Accessed: 27-Jan-2020].

[30] L. E. S. Sandra G. Hart, “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research,” *Adv. Psychol.*, vol. 52, pp. 139–183, 1988.

[31] C. Keighrey, R. Flynn, S. Murray, and N. Murray, “A QoE evaluation of immersive augmented and virtual reality speech & language assessment applications,” in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)* , 2017, pp. 1–6.

[32] A. Bangor, P. T. Kortum, and J. T. Miller, “An empirical evaluation of the system usability scale,” *Int. J. Hum. Comput. Interact.*, vol. 24, no. 6, pp. 574–594, 2008.

[33] M. J. Witmer, Bob G. and Singer, “Measuring Presence in Virtual Environments: A Presence Questionnaire,” *Presence Teleoperators Virtual Environ.*, vol. 7, pp. 225–240, 1998.

[34] A. Siddig, W. Sun, M. Parker, and A. Hines, “Perception Deception: Audio-Visual Mismatch in Virtual Reality Using the McGurk Effect,” in *Proceedings for the 27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science*, 2019.