A QoE Evaluation of Immersive Augmented and Virtual Reality Speech & Language Assessment Applications

Conor Keighrey[1], Ronan Flynn[1], Siobhan Murray[2], Niall Murray[1]

Department of Electronics & Informatics[1]
Athlone Institute of Technology
Athlone, Co. Westmeath, Ireland

Health Service Executive[2]
Primary Care Centre, Flancare Business Park
Ballyminion, Longford, Co. Longford, Ireland

email: {c.keighrey, nmurray}@research.ait.ie[1], rflynn@ait.ie[1], siobhan.murray1@hse.ie[2]

***Abstract*— The availability of affordable head-mounted displays has resulted in renewed interest in augmented reality (AR) and virtual reality (VR). This in turn has led to research into users' perceptions of quality in AR and VR environments. In this paper, the authors report the results of an experimental study that compared the user quality of experience (QoE) of an interactive and immersive speech and language assessment implemented in both AR and VR. To the best of the authors' knowledge, this is the first work that compares user QoE of VR and AR applications, in particular with a focus on applications in the speech and language domain. Subjective and objective data was captured to assess a user's QoE. Objective data, such as heart rate and electrodermal activity (EDA) was collected using consumer electronic devices such as the Fitbit and the PIP biosensor. Subjective data was captured using a user's self-reported ratings in a post-test questionnaire. The findings of this study demonstrate similar QoE ratings for both the AR and VR environments. The findings also suggest that users acclimatized to the AR environment more quickly than the VR environment.**

Keywords*— *Quality of Experience; Augmented Reality; Virtual Reality; Objective Evaluation; Subjective Evaluation; Physiological Measures; Aphasia; Speech & Language Assessment.

I. INTRODUCTION

In recent times there has been an emergence of head mounted displays (HMD) that claim to provide users with interactive and immersive multimedia experiences. Examples of such virtual reality (VR) HMDs include the Oculus Rift [1] PlayStation VR [2], HTC Vive [3], and mobile ready Samsung Gear VR [4]. While VR technology has proved promising and has gained traction across many application domains, there are issues such as the cost of content creation, the challenge of creating content at an appropriate quality level, the feeling of motion nausea and the non-user centric design of hardware [5].

In parallel to the ongoing work with VR, industry and academia are also evaluating opportunities with augmented reality (AR) technologies. While VR involves the creation of fully simulated environments, AR focuses on overlaying the real world with virtual content. Traditionally, AR used projection technology to immerse and augment a user's surroundings. However, with recent technological advances, AR HMDs, such as Microsoft HoloLens [6] and Epson Moverio BT-300 [7], now allow users to experience AR in a more intuitive way. AR allows users to interact with virtual objects, while allowing users to be fully aware of their real-world environments.

Speech & language therapy refers to the practice of aiding people with speech, language, communication and swallowing difficulties. It is estimated, that speech or language difficulties affect more than 12% of people internationally [8]. Speech & Language Therapists (SLTs) are responsible for the assessment, planning and delivery of treatment of a range of speech and language disorders. Language can be categorized into two broad areas, namely receptive and expressive language. Receptive language is one's ability to understand information, and is dependent on understanding of words, sentences, grammatical forms, concepts and more. Expressive language is the process of combining words into sentences and how one communicates their wants and needs to another. Individuals who suffer from stroke or traumatic brain injury (TBI) often experience symptoms of aphasia as a result of damage to the left frontal lobe. Anomic aphasia is a mild form of aphasia in which patients experience word retrieval problems and semantic memory difficulties. Assessments such as the comprehensive aphasia test (CAT) [9] [10] are used in a clinical settings to analyze which language areas are affected following TBI. Such assessments are administered by SLTs and are typically paper-based. A lack of available objective data in relation to speech and language impairments is a fundamental issue in this domain.

Critical to the success of immersive AR and VR applications (be they within the speech and language domain or not) is the user's quality of experience (QoE). According to [11], QoE can be defined as "the degree of delight or annoyance of a person whose experiencing involves an application, service, or system. It results from the person's evaluation of the fulfilment of his or her expectations and needs with respect to the utility and/or enjoyment in the light of the person's context, personality and current state". As has been reported, user QoE can be influenced by a number of factors, including technical, social, and psychological [12]. Traditionally in quality assessments, technical factors such as the physical device, network, or multimedia file formats were considered the most critical. However, the importance of social and psychological factors should not be overlooked, as illustrated in Fig. 1.

The literature suggests that the accepted approach to measuring a user's perceived quality of his or her experience was based on self-reported measures via post-experience questionnaires. Such questionnaires were used to determine an

QoMEX2017 – Erfurt, Germany; 978-1-5386-4024-1/17/$31.00 © 2017 IEEE

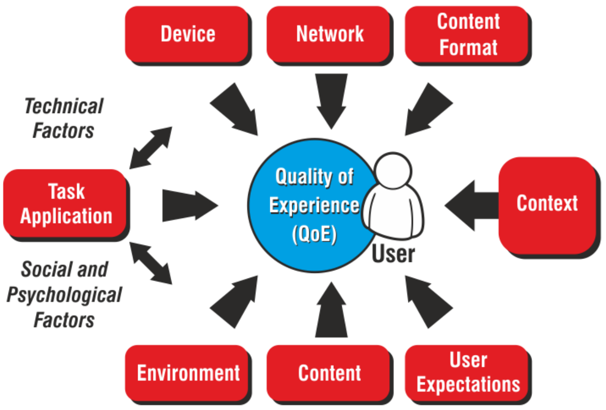

Fig. 1. Factors that influence QoE, adapted from [12].

overall mean opinion score (MOS) based on the feedback from users. However, a number of researchers have highlighted issues with MOS rating scales and post-test experience reporting; they are considered to be time consuming and expensive to implement [13]. For these reasons, the research community has started to investigate methods that capture the user's QoE as the user experiences the content. Such methods have included interactive button selection during an experience [14], the use of EEG [15] and, more recently, consumer wearable devices, such as smart watches and Fitbit [16] [17].

This paper presents the results of a novel evaluation and comparison of user QoE, for a speech and language assessment task. The assessment task was implemented in both AR and VR environments. Analysis and discussion of the comparable (AR vs. VR) objective and subjective data captured is presented. Conclusions are drawn in terms of differences in QoE between the systems but also with respect to the suitability of these systems for the application domain specified.

II. Related Work

A survey of existing quality assessment methodologies which focused on AR visualizations is presented in [18]. Concerns and limitations associated with AR visualizations, such as the perceptual issues and restricted field of view relating to display technology, were identified. Of particular interest were applications associated with neuronavigation. Each participant was asked to perform a precision-based task, which was similar in terms of ergonomics, interaction, and human behavior to a neuronavigation system. As a result of the experiment, the author proposed a quality assessment method that focused on a mixed-methods approach to assess quality in AR neuronavigation fields.

In [17], user QoE levels were compared in an immersive VR and a non-VR environment. A sample size of thirty-three participants was divided randomly into two groups. Both objective and subjective metrics were gathered. The evaluation showed that heart rate and EDA levels were elevated in the immersive VR environments compared with the non-VR environment.

Similarly, [19] investigated the correlation between physiological measures (EDA and heart rate) and subjective data as users experienced a virtual environment in a video game. As part of the study subjects were exposed to three first-person shooters for a twenty-minute time period and asked to complete an in-game questionnaire (iGEQ) every five minutes. As a result of the study a statistically significant correlation was found between heart rate and the iGEQ measured across seven dimensions of gameplay.

More closely aligned to the application domain of the work presented here, [20] describes a novel approach to the treatment of TBI and aphasia. An experiment is presented in which it was hoped to improve the treatment outcomes for aphasic patients by combining the cognitive importance of storytelling for expressive language development with immersive multimedia. Participants were seated in front of a table and provided with a box of memorabilia that was then used to represent characters in a story. The participants were asked to create a story around these objects. This was followed by a story telling session involving the participant using an AR HMD and a set of specially created cards that contained unique identifiers. As the user viewed a card, a virtual object would appear within their visual perspective. Once again they were asked to tell a story using a combination of the AR cards. Results demonstrated that the use of multimedia-assisted storytelling produced an immediate improvement in language skills. An analysis of the communicated AR and non-AR stories demonstrated an improved linguistic flow resulting in more confidence in conversation and communication.

Considering existing literature, the novelty of the work presented in this paper lies in the analysis of users' QoE and the comparison of AR and VR immersive multimedia experiences. Gathering subjective and objective data that is based on the experiences of users within AR and VR environments for receptive language assessments is another key contribution of this paper.

III. Experiment Setup

This section provides an overview of the immersive multimedia systems (AR and VR display technologies) and the devices used to measure physiological responses. The laboratory design and the subjects participating in the experiments are also discussed.

A. *Immersive multimedia systems*

1) *Augmented Reality Display Technology*

The immersive AR equipment used in the experiments was the Microsoft HoloLens [6] as per Fig. 2(a). The HoloLens is a self-contained wireless HMD that operates as a stand-alone computer. The device weighs 579 grams and contains a CPU, GPU, state of the art Holographic Processing Unity (HPU), and batteries, which are evenly distributed around the system for comfort. Four environment-understanding monochromatic cameras and a 120°×120° depth camera situated at the front of the device are used for hand gesture recognition and to map the surroundings of the wearer, allowing the creation of AR spaces. Users interact with holographic content using natural hand gestures. Two HD 16:9 light engines project virtual content onto the holographic lenses located within the user's field of view (FOV). It is estimated that the holographic lenses have a FOV of 30° by 17.5°. External speakers located above each ear provide 3D spatial audio to the wearer.

Fig. 2. (a) User interacting with the AR environment, (b) User interacting the the VR environment.

2) Virtual Reality Display Technology

The Oculus Rift (OR) Development Kit 2 (DK2) HMD [1] was used in this experiment to deliver an immersive VR experience as per Fig. 2(b). An internal OLED screen provides each eye with a display resolution of 960 x 1080 pixels. The OR DK2 has a FOV of 100°, which is similar to a human eye (approximately 120°). A series of sensors and an infrared camera tracks a user's head movements; these movements are then translated into the virtual world. Since the OR does not provide the facility to track natural hand movement as outlined above for the AR experience, a Leap Motion controller [21] was integrated into the VR display system. The Leap Motion controller tracks hand movements and recognizes hand gestures using two monochromatic infrared (IR) cameras and three IR LEDs. To ensure consistency between the AR and VR experiences from an interaction perspective, the Microsoft HoloLens gestures were replicated in the VR environment.

3) Immersive Speech & Language Assessment Application

The immersive speech and language assessment application was developed using the Unity game engine [22] (Version 5.4.0f3). The application replicates a semantic memory assessment as described within CAT [10]. The semantic memory exercise evaluates a person's language processing on a cognitive level. The assessment consists of ten multiple choice scenarios. Each of the scenarios contain five images. One of the five images (placed in the center as per Fig. 3) is the categorical item. As part of the assessment, the participant must correctly identify which one of the four remaining images forms a categorical link with the categorical item. As per CAT, this task was accomplished without using a verbal response. Hence, the focus was on the cognitive process i.e. gestures were used to identify the correct categorical relationship between the images.

B. Objective Metrics

The objective physiological data in the form of heart rate and EDA were stored throughout all stages of the experiment. Two non-invasive consumer devices were used to gather this data. Heart rate information was captured using the Fitbit Charge HR (Fig. 4(a)) [23]. The wireless activity tracker attaches to the participant's wrist, as per Fig. 2, and an optical sensor located under the main interface monitors blood volume changes on a per second basis. The EDA data was recorded using the PIP Biosensor as per Fig. 4(b) [24]. The wireless Bluetooth device is held by a user between the thumb and index finger (as per Fig. 2). Conductivity changes in the skin are monitored in order to analyze a user's stress levels. This change in conductivity is

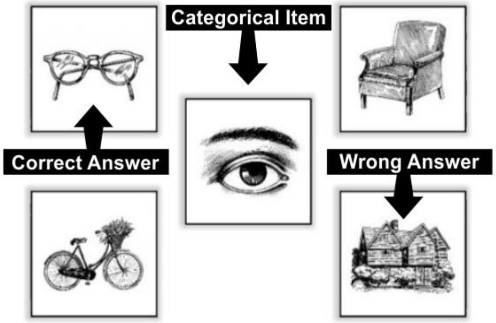

Fig. 3. Typical Semantic Memory Excercise

Fig. 4. (a) Fitbit Charge HR, (b) PIP Biosensor

referred to as galvanic skin response (GSR). The PIP uses an internal GSR analysis system to provide accurate feedback regardless of environmental conditions or skin type. Measurements are taken on average eight times per second.

C. Participant Screening

The screening process for participants included visual acuity and color perception. For visual acuity, a Snellen test [25] was administered. A score of 20/20 was required to pass this portion of the screening process. Red-green color deficiencies were screened using an Ishihara test [26]. Administration of this test required each participant to evaluate a set of thirty-eight colored discs. A maximum of four errors was allowed during the examination [26]. Based on the two screening conditions, two subjects were deemed ineligible to participate in the assessment.

D. Training Videos

Due to the novelty of these AR and VR technologies, a short series of training videos was developed. These lasted approximately 20s-30s each. The video content demonstrated how users should interact with the virtual assessment using a combination of head movement and hand gestures within the AR/VR environments. In addition, the training videos provided an overview of the gesture interaction which participants were required to perform as part of the semantic memory assessment.

E. Laboratory Design

Both the AR and VR evaluations took place in the same lab. This lab was inspired by the guidelines outlined in ISO 8589:2007 [27] and [17]. ISO 8589:2007 provides guidelines for test room design that focus on sensory analysis. In terms of the layout of the simulated VR lab, it was developed on a 1:1 scale with the physical lab in order to ensure consistency between the AR and VR testing conditions.

F. Participants

A convenience sampling approach was used to recruit participants for the study. A total of forty-six participants took part in the study with an average age of 27 years. Due to the screening outlined in Section III C, two participants were omitted. In addition, the results from a further three participants was removed due to incomplete data.

From the twenty-one participants who took part in the AR environment, seven had used AR before. Similarly, of the twenty participants who took part in the immersive VR experience, six had experienced VR before. From all forty-one participants, no individual had used an AR HMD but eleven had been exposed to mobile AR experiences such as Pokémon Go [28].

IV. Assessment Methodology

A. Assessment Protocol

The assessment protocol can be categorized into four key phases: information phase; screening phase; training phase and testing phase. During the information phase, each participant was greeted and thanked for their participation. They were brought to the waiting room and were provided with an information sheet that described the experiment in full. The screening phase assessed a participant's visual acuity and color perception. Upon completion, baseline metrics of heart rate and EDA were captured over a five-minute period using the devices outlined in Section III B.

The training phase required each participant to view a series of training videos as described in Section III D. Training also included participants performing the hand gestures under the guidance of the principal investigator. As per above, this gesture interaction was required as part of the semantic memory assessment. Upon completion, participants were introduced to either an AR or VR HMD. The device was fitted to the participant by the principal investigator and opportunity for adjustment was provided to ensure there was no discomfort.

The final part of the training phase involved the completion of a training exercise while using the HMD. The objective of this was to ensure that each participant understood in full how to interact with the environment. The training exercise was developed to emulate the final semantic memory exercise; however, it required participants to match colored blocks rather than acknowledging categorical relationships between images. Heart rate and EDA metrics were gathered throughout this final part of the training phase. During phase four, the immersive speech and language assessment was conducted. Participants interacted with the assessment in accordance to the CAT. Once again, heart rate and EDA data was captured throughout the assessment. Participants then completed the questionnaire as outlined in Section IV B. On average participants completed the test in 30-35 minutes. Typically, this included an 8 minute informative phase, a 10 minute screening process, a 10 minute training phase and, finally, a 5 minute testing phase.

B. Questionnaire & Rating Scale

Fourteen questions were developed to evaluate participant interaction, usability, enjoyment, immersion, and discomfort. The questions were inspired by the QoE model categories as outlined in Fig. 1. The questionnaire is available at [29]. User interaction was evaluated by questions 3, 7, and 10. These questions aimed to analyze how naturally users interacted with the systems, the learning curves associated with interaction and the ease of interaction. Immersion was measured using questions 5, 6, 11, and 13. These queried the levels of engagement, activity immersion, environment realism, and sense of presence. Questions 1, 4, 8, and 9 aimed to analyze levels of discomfort while using the HMDs. These questions focused directly on device discomfort, levels of annoyance, restriction in movement, and feelings of nausea. Levels of enjoyment were analyzed in questions 2, 12, and 14 via participant expectations, enjoyment levels, and interest in re-experiencing the system. Participants were asked to rate each question using the absolute category rating (ACR) system as outlined in ITU-T P.913 [30]. The rating system uses a five-point Likert scale to determine if a user agreed or disagreed with the statements.

V. Results & Discussion

In this section of the paper, the findings with respect to the objective and subjective data captured during the AR and VR experiments are presented and discussed.

A. Immersive Speech & Language Assessment

Speech and language assessment performance was recorded throughout the experiment. According to the normative data as provided with the CAT, it is expected that participants from the healthy population score a mean value of 9.8 with a standard deviation of 0.4, while a mean score of 6.9 with a standard deviation of 3.03 would indicate an acute aphasic performance [10]. Since our sample population was classified as healthy, our results in terms of the CAT, as expected, fell within the first set of scores mentioned above. As the focus here is on user QoE, the performance of participants in terms of CAT is considered as future work.

B. Self-Reported Questionnaire Results

Table I presents the results of the MOS self-reported measures captured via the post-test questionnaires discussed in Section IV B, along with the statistical analysis. Since the AR and VR participants were part of two independent groups, an independent samples t-test was performed on the data with 95% confidence level using the IBM statistical analysis software package SPSS [31]. Given the findings in related works, such as [17] (in [17], there were numerous statistically significant differences between the VR and non-VR groups), the results here are somewhat surprising. As per Table I, of the fourteen questions asked, only question 12, which asked Participants if they would like to experience this environment again, reported a statistically significant difference between the AR and VR groups with a two-tailed p value of $p=0.019$, $p<0.05$. The AR group reported a MOS rating of 4.48 whereas the VR group reported a MOS rating of 3.24 as per Table I. Question 13, was the only other question that had a MOS difference of greater than 0.5 making it close to being statistically significant. It aimed to discover if participants did not feel a sense of presence while experience the system.

For each of the remaining questions, differences between the AR and VR group were not found to be statistically significant. This was not expected. Prior to the experiments, it was

TABLE I. STATISITCAL ANALYSIS OF SELF REPORTED MEASURES WITH 95% CONFIDENCE LEVEL

	AR		VR				
	MOS	***SD***	***MOS***	***SD***	***F***	***df***	***Sig. (2-tailed)***
Q1	3.90	0.995	3.55	1.468	5.341	39	0.369
Q2	4.57	0.507	4.60	0.503	0.129	39	0.857
Q3	3.71	1.007	4.05	0.945	0.402	39	0.278
Q4	2.43	1.028	2.00	0.858	2.436	39	0.156
Q5	4.38	0.590	4.35	0.489	1.576	39	0.856
Q6	4.71	0.463	4.60	0.503	2.106	39	0.453
Q7	3.24	1.700	3.35	1.599	0.359	39	0.829
Q8	2.48	0.928	2.40	0.995	0.283	39	0.801
Q9	1.14	0.359	1.30	0.470	6.174	39	0.235
Q10	4.57	0.598	4.50	0.607	0.069	39	0.706
Q11	3.48	1.123	3.90	0.641	9.107	39	0.149
Q12	4.48	0.512	4.05	0.605	1.386	39	0.019
Q13	2.71	0.902	3.30	1.174	2.043	39	0.080
Q14	2.29	0.845	1.85	0.813	0.009	39	0.101

hypothesized that the VR experience group would have higher levels of immersion, interaction, enjoyment than the AR group.

C. Objective Metrics

As outlined previously, the physiological metrics considered as part of this work were heart rate and electrodermal activity (EDA). Previous works indicated correlations between these metrics and levels of immersion [17] and information recall [32]. The results are presented in Table II and Fig. 5 for EDA and heart rate respectively. With respect to Table II, the data presented from the PIP biosensor includes relaxed, stressed, constant, skin conductivity and accumulated trend, for both the AR and VR groups, and for each of the screening, training and testing phases as described earlier.

The PIP's event classification algorithm detects changes in EDA, which can be categorized as either a stressed event, a relaxed event, or a constant event. Each occurrence of any of these three possible events was recorded by the PIP software during the screening, training, and testing phases of the assessment. As per [33], stress events indicate a short-term increase in EDA, relax events indicate a short-term reduction in EDA, and constant events refer to no significant change in EDA over that time period. The skin conductance column (measured in micro-Siemens) corresponds to the raw skin conductance data measured from the participant's skin during each of the phases for the AR and VR groups. The results show an increase in EDA for both AR and VR as participants transitioned from the screening phase to training to testing phase.

In Table II we note that the average skin conductivity values increased across the screening (3.26), training (3.36), and testing phases (3.54) in the AR group. Similarly, the VR group saw an increase in value from 3.02 during the screening phase, to 3.27 during the training phase and 3.39 during the testing phase. In comparison, the average skin conductivity values for VR

TABLE II. USER EDA METRICS DURING THE SCREENING, TRAINING, AND TESTING PHASE OF THE EXPERIMENT.

	Relaxed	**Stressed**	**Constant**	**Skin Conductance**	**Acc. Trend**
Screening Phase					
VR	11.00	8.30	10.20	3.02	195.66
AR	9.90	8.76	10.24	3.26	88.80
Training Phase					
VR	6.85	8.00	6.80	3.27	10.49
AR	6.76	9.19	6.86	3.36	-55.62
Testing Phase					
VR	14.55	12.80	11.70	3.39	66.74
AR	7.33	9.81	11.24	3.54	-27.33

had a higher increase as they transitioned from the baseline data recorded in the screening phase to the training phase. Similarly, larger differences in skin conductance is observed between the baseline and test phase for VR compared to AR. This suggests that a VR application solicits higher levels of arousal in the participants compared with the AR group. Demonstrating a correlation between this objective dataset and the self-reported measures of Q13.

The average number of relaxed events recorded throughout the three phases was 9.9, 6.76, and 7.33 in the AR group and 11, 6.85, and 14.55 in the VR group. In terms of stress events, the AR group averaged 8.76, 9.19, and 9.81 compared to the VR group averages 8.3, 8, and 12.8 during the three respective phases. Based on this data we note an increase of 1.05 on average in stress events for the AR group, as we progressed from the screening phase to the testing phase. The VR group experienced a much higher increase of 4.5 stress events on average. This increase was also seen within the relaxed event data which saw the VR data increase by 4.55 while the AR data decreased in the average relaxed events record. This data suggests that the VR group experienced more stress throughout the training and testing phases of the experiment than the AR group.

The heart rate data in Fig. 5 shows the readings for the AR and VR groups during the screening, training and testing phases. For the VR group, the average heart rate values were 78.86, 79.59 and 80.16 beats per minute (BPM) for each of the screening, training and testing phases respectively. The max and min heart rate values during each of these phases was 81.8/76.05, 81.25/77.9 and 83/77.05. For the AR group, the average heart rate values were 77.3, 78.31 and 77.65 BPM for the phases outlined with the max and min AR values during each of these phases was 80/74.667, 80.28/75.76 and 80.57/74.8. Based on these results, it can be seen that for both groups an increase in heart rate occurred as the participants transitioned from the screening phase to the training phase. This supports the findings that were reported in [17] between the non-VR and VR readings. However, a key difference here between the AR and VR dataset exists when the participants transitioned from the training phase to the testing phase. The VR group's heart rate continued to rise (at a low rate), but the heart rate for the AR group dropped, which was unexpected. This potentially suggests that the AR group acclimatized to their immersive experience more easily than the VR group.

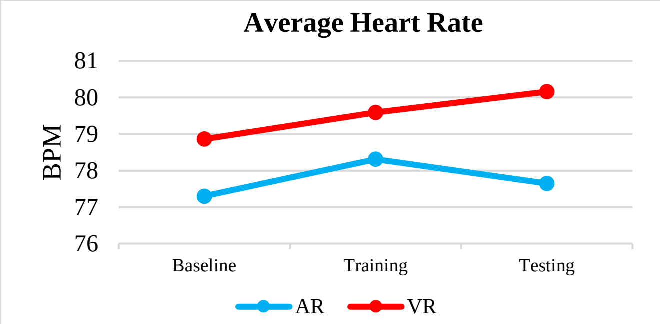

Fig. 5. Heart rate data throughout the screening, training and testing phase.

VI. Conclusion

This paper presents a comparison, in terms of subjective ratings and objective metrics, of user QoE of immersive AR and VR speech and language therapy applications. Based on the results, both novel approaches have significant potential as methods of administration of semantic memory assessments. The analysis in terms of HR, suggests that AR users became more accustomed to their environment when compared with VR users. In addition, this work highlights the potential and benefits of using objective metrics such as EDA and HR as indicators of user QoE for these immersive experiences. Future work will validate the assumption that higher levels of immersion and engagement facilitates better assessment and treatment outcomes. Another avenue for future work is the potential addition of multisensory components to these immersive applications.

Acknowledgment

This research was supported by the Irish Research Council (grant number: GOIPG/2016/1493).

References

[1] "Oculus Rift Development Kit 2," Oculus, [Online]. Available: https://www.oculus.com/en-us/dk2/. [Accessed 29 04 2016].

[2] "PlayStation VR | PlayStation.com," PlayStation, [Online]. Available: https://www.playstation.com/en-ie/explore/playstation-vr/. [Accessed 29 04 2017].

[3] "VIVE™ | Discover Virtual Reality Beyond Imagination," HTC, [Online]. Available: https://www.vive.com/eu/. [Accessed 29 04 2017].

[4] "Samsung Gear VR," Samsung, [Online]. Available: http://www.samsung.com/global/galaxy/gear-vr/. [Accessed 29 04 2017].

[5] D. Simon, K. Nesbitt and E. Nalivaiko, "Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters," *Proceedings of the 11th Australasian Conference on Interactive Entertainment,* vol. 27, 2015.

[6] "Microsoft HoloLens," Microsoft, [Online]. Available: https://www.microsoft.com/microsoft-hololens/en-us. [Accessed 29 04 2017].

[7] "Epson Moverio Augmented Reality Smart Glasses," Epson, [Online]. Available: https://epson.com/moverio-augmented-reality. [Accessed 29 04 2017].

[8] "Scope of Practice in Speech-Language Pathology," 2016. [Online]. Available: http://www.asha.org/uploadedFiles/SP2016-00343.pdf. [Accessed 29 04 2017].

[9] J. Reilly, "Semantic Memory and Language Processing in Aphasia and Dementia," *Seminars in Speech and Language,* vol. 29, no. 1, pp. 3-4, 2008.

[10] K. Swinburn, G. Porter and D. Howard, Comprehensive Aphasia Test, Psychology Press, 2004.

[11] S. Möller and A. Raake , "Quality and Quality of Experiance," in *Quality of Experience: Advanced Concepts, Applications and Methods*, Springer, 2014, p. 19.

[12] T. Ebrahimi, "Quality of Multimedia Experience: Past, Present and Future," in *Proceedings of the 17th ACM international conference on Multimedia*, 2009.

[13] T. Hoßfeld, R. Schatz and S. Egger, "SOS: The Mos Is Not Enough!," in *Quality of Multimedia Experience (QoMEX),* 2011.

[14] M. Waltl, C. Timmerer and H. Hellwagner, "Improving the Quality of multimedia Experience through sensory effects," *Quality of Multimedia Experience (QoMEX),* pp. 124-129, 2010.

[15] A.-M. Brouwer, M. Neerincx, V. L. Kallen and M. t. Brinke, "EEG alpha asymmetry, heart rate variability and cortisol in response to Virtual Reality induced stress," *Journal of CyberTherapy & Rehabilitation,* vol. 4, no. 1, pp. 27-40, 2011.

[16] C. Timmerer, T. Ebrahimi and F. Pereira, "Toward a New Assessment of Quality," *Computer,* vol. 48, no. 3, pp. 108-110, 015.

[17] 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.

[18] J. Puig, A. Perkis, F. Lindseth and T. Ebrahimi, "Towards an Efficient Methodology for Evaluation of Quality of Experience in Augmented Reality," in *Quality of Multimedia Experience (QoMEX)*, 2012.

[19] A. Drachen, L. E. Nacke, G. Yannakakis and A. L. Pedersen, "Correlation between Heart Rate, Electrodermal Activity and Player," in *SIGGRAPH Symposium on Video Games* , 2010.

[20] C. Stapleton, J. Whiteside, J. Davies, D. Mott and J. Vick, "Transforming Lives Through Story Immersion: Innovation of Aphasia Rehabilitation Therapy through Storytelling Learning Landscapes," in *ACM International Workshop on Immersive Media Experiences*, 2014.

[21] "Leap Motion," Leap Motion, [Online]. Available: https://www.leapmotion.com/. [Accessed 29 04 2017].

[22] "Unity - Game Engine," Unity3D, [Online]. Available: https://unity3d.com/. [Accessed 29 04 2017].

[23] "Fitbit Charge HR," Fitbit, [Online]. Available: https://www.fitbit.com/chargehr. [Accessed 29 04 2017].

[24] "PIP Biosensor," PIP, [Online]. Available: https://thepip.com/en-eu/. [Accessed 29 04 2017].

[25] C. H. Graham, N. R. Bartlett, J. L. Brown, C. G. Mueller, Y. Hsia and L. A. Riggs, Vision and Visual Perception, John Wiley & Sons Inc., 1965.

[26] Committee on Vision, Assembly of Behavioral and Social Sciences, National Research Council, "Procedures for Testing Color Vision: Report of," NATIONAL ACADEMY PRESS, 1981.

[27] "ISO 8589:2007 Sensory analysis — General guidance for the design of test rooms," International Standards Organisation, [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:8589:ed-2:v1:en. [Accessed 29 04 2017].

[28] "Pokémon Go," [Online]. Available: http://www.pokemongo.com/. [Accessed 29 04 2017].

[29] "QoMEX 2017 - Post-test Questionnaire," [Online]. Available: www.niallmurray.info/Research/appendix. [Accessed 29 04 2017].

[30] "ITU-T 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," ITU-T, [Online]. Available: https://www.itu.int/rec/T-REC-P.913/en. [Accessed 29 04 2017].

[31] "IBM SPSS - IBM Analytics," IBM, [Online]. Available: https://www.ibm.com/analytics/us/en/technology/spss/. [Accessed 29 04 2017].

[32] R. Tortell, D. P. Luigi, A. Dozois, S. Bouchard, J. F. Morie and D. Ilan, "The effects of scent and game play experience on memory of a virtual environment," *Virtual Reality,* vol. 11, no. 1, pp. 61-68, 2007.

[33] PIP Data Recorder User Guide, Galvanic Ltd., 2015.