JTS

The Journal of Transdisciplinary Studies, 3(1): 93-104, 2019
https://doi.org/10.22685/JTS.2019.3.1.093

A QoE Evaluation of Augmented and Virtual Reality Speech and Language Treatment Applications*

Heejune Park[1], Sinwook An[2], Inho Bae[3], Soonbok Kwon[4]

[1]Dept. of Speech and Hearing Therapy, Catholic University of Pusan, Korea (june@cup.ac.kr), First Author
[2]Dept. of Rehabilitation Medicine, Pusan National University Hospital, Korea
[3]Dept. of Otorhinolaryngology-Head and Neck Surgery, Pusan National University Hospital, Korea
[4]Dept. of Language and Information, Pusan National University, Korea
(sbkwon@pusan.ac.kr), Corresponding Author

ARTICLE INFO

Article history:
Received 25 May 2019
Revised 13 June 2019
Accepted 18 June 2019

Keywords:
Quality of Experience,
Augmented Reality,
Virtual Reality,
Objective Evaluation,
Subjective Evaluation,
Physiological Measures,
Aphasia,
Speech & Language
Assessment

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 Smartwatch 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.

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A5B6075173)

1. Introduction

Virtual Reality (VR) and Augmented Reality (AR) have become to the public. As a result of advanced technology and software body, VR and AR are applied to diverse areas such as health services, hospitality, education, cultural, engineering, gaming and entertainment (Hsieh & Lee, 2018). Although VR technology has been promising and gained traction across many application domains, there still are issues of cost quality content creation, causing motion sickness, and the non-user centric design of hardware (Davis et al., 2015). In parallel to the ongoing work with VR, industry and academia are also evaluating opportunities with AR technologies. While VR involves the creation of only virtual environments, AR focuses on overlaying the real world with virtual content. Traditionally, projection technology was employed to immerse and augment a user's surroundings in AR. AR allows users to interact with virtual objects and be fully aware of their real-world environments (Kesim & Ozarslan, 2012).

Speech and language therapy refers to the rehabilitation of communication impairments in people with speech and/or language difficulties. It is estimated that speech or language difficulties affect more than 12% of people internationally (ASHA, 2016). Speech-Language Pathologists (SLPs) are responsible for the assessment, planning and delivery of treatment of a wide range of speech and language disorders. 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 Korean Western Aphasia Battery (K-WAB) are used in a clinical setting to analyze which language area is affected by TBI(Jang et al., 2016; Kang et al., 2010). Such assessments are administered by SLPs 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.

With regards to user's quality of experience (QoE), defined as the degree of delight or annoyance of a person whose experience involves an application, service, or system (Möller & Raake, 2014) may play a critical role in successful AR and VR applications. QoE is based on the user'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.

QoE is influenced by technical, social, and/or psychological factors. In traditionally quality assessments, technical factors such as the physical device, network, and multimedia file formats were considered the most critical. However, the importance of social and psychological factors should not be overlooked (Ebrahimi, 2009).

According to the literature, the traditional approach to measuring user-perceived quality of experience is based on self-reported measures via post-experience questionnaires. Such questionnaires are used to determine the overall mean opinion score (MOS) based on the feedback from the users. However, several studies have highlighted issues with the MOS rating scale and post-test experience reporting; they are time consuming and cost inefficient (Hobfeld et al., 2011). Because of the disadvantages, researchers have started to investigate methods that can capture user's QoE as the user experiences the content. Such methods include interactive button selection during an experience (Egan et al., 2016), use of EEG (Brouwer et al., 2011), and consumer wearable devices, such as smart watches and Fitbit (Waltl et al., 2010).

A survey of existing quality assessment methodologies, which focused on AR visualizations, is presented in previous study (Puig et al., 2012). Concerns and limitations associated with AR visualizations, such as the perceptual issues and restricted field of view relating to display technology, were identified. Of 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.

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 (Egan et al., 2016).

More closely aligned to the application domain of the work presented here, a subject describes a novel approach to the treatment of aphasia (An et al., 2017). An experiment is presented in which it was hoped to improve the treatment outcomes for aphasic patients by combining the cognitive importance of naming and functional communication for expressive language development with immersive multimedia. For makers, 68 three-dimensional action and 90 background objects were designed in household environments such as the room, living room, bathroom, and kitchen, which were constructed to be presented in the form of a script. Participants were asked to report on the actions that characters were taking on the picture, which were implemented in the form of a script, and the names of the background objects. This was followed by a story telling session involving the participant using an AR 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 picture stories demonstrated an improved linguistic flow resulting in more confidence in conversation and communication.

Considering previous studies, the purpose 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.

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.

2. Method

2.1 Immersive multimedia systems

Treatment using AR and VR forms of media have been created. Among them, AR is a branch of VR, and a compound word of augmentation, a computer graphic technique that makes an object look like they exist in the real world by combining virtual objects, places, and other information in the real environment and reality, referring to an actual world that unfolds before one's eyes (Figure 1). In other words, as it allows their user to share virtual information, which is generated by the computer, in the real world, AR can be combined with many different types of virtual learning contents used in therapies.

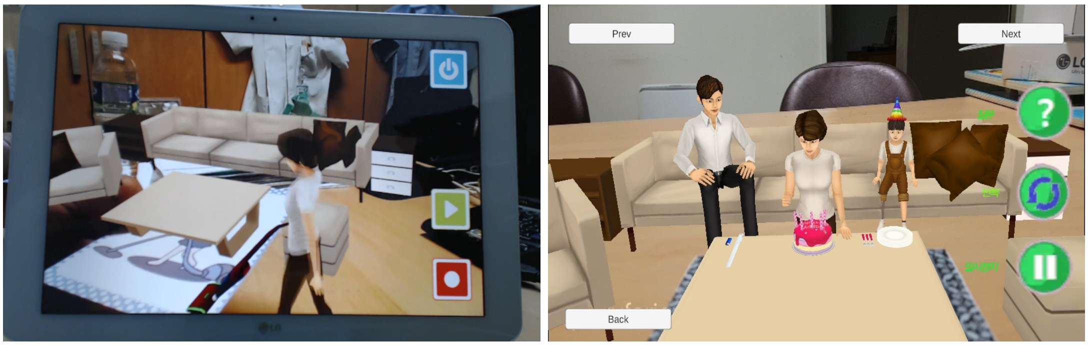

Fig. 1. (a) User interacting with the AR environment, (b) User interacting the VR environment

2.2 Immersive Speech and Language Treatment Application

This study's intervention program was developed to see if activities using an AR-based speech therapy program improved the patients' naming ability and functional communication. This study's system had the AR function that recognized markers through a tablet PC camera and a three-dimensional video on the display, constructed to be presented in the form of a script. Furthermore, the system provided a training mode, practice mode, evaluation mode, and a system configuration interface to support naming therapies for object names or actions. The intervention program used in a session implemented three-dimensional actions and background objects in AR on a desk in a therapy room for the subjects.

In this study, subjects were asked to report on the actions that characters were taking on the screen, which were implemented in the form of a script, and the names of the background objects. When the subjects could not answer a question, they were provided with step-by-step clues to lead them to a target word.

The following were the step-by-step clues: first, clues for the target word's function or meaning; second, clues for the target word's category; third, clues for completing a sentence in which the target word is used; fourth, gesture clues; fifth, clues regarding the first sound of the target word, and; sixth, clues for the first letter of the target word. These clues were presented step-by-step. If the subjects could not find the target word even after the last clue was provided, the researcher articulated the target word and led them to imitate it (Figure 2).

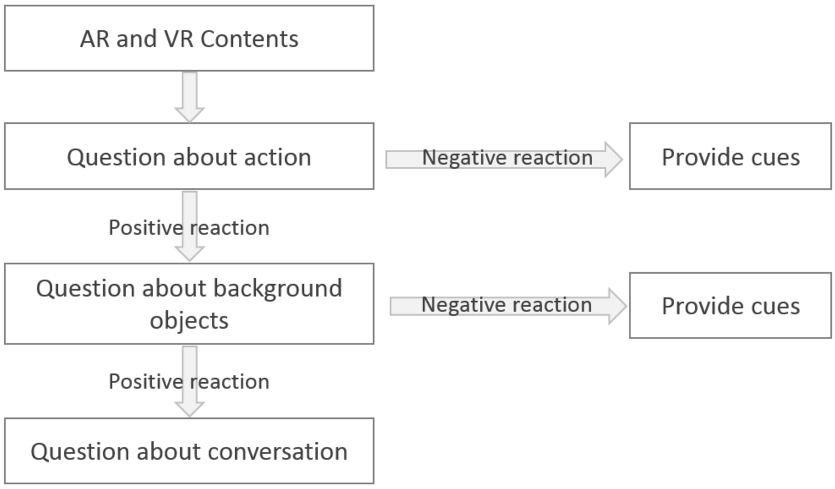

Fig. 2. Process of the AR based speech therapy

In addition, in this study, subjects were asked about what they did in their room, living room, bathroom, or kitchen so that they could speak about the target or other words, which led them to naturally talk about things related to daily life. In this study, the subjects' wrong reactions to the last questions were overlooked. VR followed the treatment procedure using the AR described above with the HMD (head mounted display) worn.

2.3 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 and EDA data was captured using the Galaxy Watch Active (Figure 3). The wireless activity tracker attaches to the participant's wrist, and an optical sensor located under the main interface monitors blood volume changes on a per second basis. Conductivity changes in the skin are monitored in order to analyze a user's stress levels. This change in conductivity is referred to as galvanic skin response (GSR). Galaxy Watch Active 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.

Fig. 3. ‘My BP Lap Watch App’ for Galaxy Watch Active

2.4 Participant Screening

The screening process for participants included visual acuity and color perception. For visual acuity, a Snellen test (Graham, 1965) 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 (Pokorny et al., 1981). Administration of this test required each participant to evaluate a set of thirty-eight colored discs. A maximum of four errors were allowed during the examination. Based on the two screening conditions, two subjects were deemed ineligible to participate in the assessment.

2.5 Laboratory Design

Both the AR and VR evaluations took place in the same lab. A convenience sampling approach was used to recruit participants for the study. A total of ten participants took part in the study with an average age of 23 years. From the ten participants who took part in the AR environment, four had used AR before. Similarly, of the ten participants who took part in the immersive VR experience, eight had experienced VR before.

2.6 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.

The training phase required each participant to view a series of training contents to AR, VR and picture. Heart rate and EDA metrics were gathered throughout this final part of the training phase. During training, the immersive speech and language assessment was conducted. Participants interacted with the assessment in accordance to the S-K-BNT. Once again, heart rate and EDA data was captured throughout the assessment. Participants then completed the questionnaire as outlined in QoE. On average participants completed the test in 30 minutes.

2.7 Questionnaire and Rating Scale

Fourteen questions were developed to evaluate participant interaction, usability, enjoyment, immersion, and discomfort. The QoE question was translated into Korean using the contents of the previous research. The questionnaire is available at ("QoMEX 2017," n.d.). 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. The rating system uses a five-point Likert scale to determine if a user agreed or disagreed with the statements.

3. Results and 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.

3.1 Immersive Speech and Language Assessment

Speech and language assessment performance was recorded throughout the experiment. According to the normative data as provided with the S-K-BNT, it is expected that participants from the healthy population score a mean value of 9.36 with a standard deviation of 3.45, while a mean score of 4.65 with a standard deviation of 3.43 would indicate an Alzheimer's performance (Kang et al., 2000). Since our sample population was classified as healthy, our results in terms of the S-K-BNT, 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 S-K-BNT is considered as future research (e.g dementia, aphasia).

3.2 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 sample Mann-Whitney U test was performed on the data with 95% confidence level using the IBM statistical analysis software package SPSS. As per Table 1, 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 1. 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 hypothesized that the VR experience group would have higher levels of immersion, interaction, and enjoyment than the AR group.

Table 1. Statisitical analysis of self reported measurement

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

3.3 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 (Egan et al., 2016) and information recall (Tortell et al., 2007). 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 Galaxy Smart Watch 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 Galaxy Smart Watch'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 Galaxy Smart Watch software during the screening, training, and testing phases of the assessment. 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 pre-training (4.32), training (4.62), and testing phases (4.81) in the AR group. Similarly, the VR group saw an increase in value from 4.12 during the pre-training phase, to 4.45 during the training phase and 4.75 during the testing phase. In comparison, the average skin conductivity values for VR 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 11, 6.5, and 9 in the AR group and 13, 7, and 15 in the VR group. In terms of stress events, the AR group averaged 8.9, 9.5, and 10 compared to the VR group averages 8.2, 7.8, and 13 during the three respective phases. Based on this data we note an increase of 1.1 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 4.8 stress events on average. This increase was also seen within the relaxed event data which saw the VR data increase by 4.8 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.

Table 2. User EDA metrics during the pre-training, training, and testing phase of the experiment

	Relaxed	Stressed	Constant	Skin Conductance
Pre-training Phase				
VR	13	8.2	10.20	4.12
AR	11	8.9	10.24	4.32
Training Phase				
VR	7	7.8	6.80	4.45
AR	6.5	9.5	6.86	4.62
Testing Phase				
VR	15	13	11.70	4.75
AR	9	10	11.24	4.81

The heart rate data in Figure 4 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 78.5, 80.6 and 81.4 beats per minute (BPM) for each of the pre-training, training, and testing phases respectively. For the AR group, the average heart rate values were 77.9, 78.2 and 77.6 BPM for the phases outlined. Based on these results, it can be seen that for both groups an increase in heart rate occurred as the participants transitioned from the pre-training phase to the training phase. This supports the findings that were reported in Egan et al. (2016) 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.

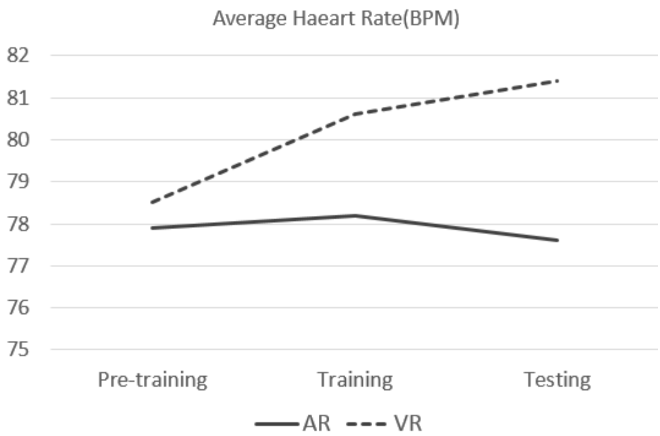

Fig. 4. Heart rate data throughout the screening, training and testing phase

4. 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.

References

An, S. W., Kim, G. H., Park, H. J., & Kwon, S. B. (2017). Effects of the speech therapy program based augmented reality on improving naming and function communication ability of the patients with expressive aphasia. *The Journal of Transdisciplinary Studies, 1*(1), 57-67.

American Speech and Hearing Association (ASHA). (2016). Scope of practice in speech-language pathology.

Brouwer, A.-M., Neerincx, M. A., Kallen, V., Van Der Leer, L., & Ten Brinke, M. (2011). EEG alpha asymmetry, heart rate variability and cortisol in response to Virtual Reality induced stress. *Journal of Cyber Therapy and Rehabilitation, 4*(1), 27-40.

Davis, S., Nesbitt, K., & Nalivaiko, E. (2015). Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters. *Proceedings of the 11th Australasian Conference on Interactive Entertainment (IE 2015), 27*, 30.

Ebrahimi, T. (2009). Quality of multimedia experience: past, present and future. *MM'09: Proceedings of the Seventeen ACM International Conference on Multimedia, (CONF)*, 3-4. ACM.

Egan, D., Brennan, S., Barrett, J., Qiao, Y., Timmerer, C., & Murray, N. (2016). An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments. *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, 1-6. https://doi.org/10.1109/QoMEX.2016.7498964

Graham, C. H. (1965). Vision and visual perception. Wiley.

Hobfeld, T., Schatz, R., & Egger, S. (2011). SOS: The MOS is not enough! *2011 Third International Workshop on Quality of Multimedia Experience*, 131-136. https://doi.org/10.1109/QoMEX.2011.6065690

Hsieh, M. C., & Lee, J. J. (2018). Preliminary Study of VR and AR Applications in Medical and Healthcare Education. *Journal of Nursing and Health Studies, 3*(1), 1. https://doi.org/10.21767/2574-2825.100030

Jang, S. H., Lee, A. Y., & Shin, S. M. (2016). Injury of the Arcuate Fasciculus in the Dominant Hemisphere in Patients With Mild Traumatic Brain Injury: A Retrospective Cross-Sectional Study. *Medicine, 95*(9), e3007. https://doi.org/10.1097/MD.0000000000003007

Kang, E. K., Sohn, H. M., Han, M.-K., Kim, W., Han, T. R., & Paik, N.-J. (2010). Severity of Post-stroke Aphasia According to Aphasia Type and Lesion Location in Koreans. *Journal of Korean Medical Science, 25*(1), 123. https://doi.org/10.3346/jkms.2010.25.1.123

Kang, Y., Kim, H., & Na, D. L. (2000). Parallel short forms for the Korean-Boston Naming Test (K-BNT). *Journal of Korean Neurological Association, 18*(2), 144-150.

Kesim, M., & Ozarslan, Y. (2012). Augmented Reality in Education: Current Technologies and the Potential for Education. *Procedia - Social and Behavioral Sciences, 47*, 297-302. https://doi.org/10.1016/J.SBSPRO.2012.06.654

Möller, S., & Raake, A. (2014). *Quality of experience: advanced concepts, applications and methods*. Springer.

Pokorny, J., Collins, B., Howett, G., Lakowski, R., & Lewis, M. (1981). *Procedures for testing color vision*. National Research Council Washington DC Committee on Vision.

Puig, J., Perkis, A., Lindseth, F., & Ebrahimi, T. (2012). Towards an efficient methodology for evaluation of quality of experience in Augmented Reality. *2012 Fourth International Workshop on Quality of Multimedia Experience*, 188-193. https://doi.org/10.1109/QoMEX.2012.6263864.

Tortell, R., Luigi, D. P., Dozois, A., Bouchard, S., Morie, J. F., & Ilan, D. (2007). The effects of scent and game play experience on memory of a virtual environment. *Virtual Reality, 11*(1), 61-68.

QoMEX 2017. (n.d.). Retrieved from www.niallmurray.info/Research/appendix

Waltl, M., Timmerer, C., & Hellwagner, H. (2010). Improving the Quality of multimedia Experience through sensory effects. *2010 Second International Workshop on Quality of Multimedia Experience (QoMEX)*, 124-129. https://doi.org/10.1109/QOMEX.2010.5517704