

# Evaluation of Animation and Lip-Sync of Avatars, and User Interaction in Immersive Virtual Reality Learning Environments

Bruno Peixoto<sup>\*†</sup>, Miguel Melo<sup>†</sup>, Luciana Cabral<sup>‡§</sup> and Maximino Bessa<sup>\*†</sup>

<sup>\*</sup>University of Trás-os-Montes and Alto Douro, 5001-801 Vila Real, Portugal

<sup>†</sup>Institute for Systems and Computer Engineering, Technology and Science, 4200-465 Porto, Portugal

<sup>‡</sup>Transdisciplinary Research Centre Culture, Space and Memory (CITCEM),

Faculty of Letters, University of Porto, 4099-002 Porto, Portugal

<sup>§</sup>Polytechnic Institute of Bragança, 5300-253 Bragança, Portugal

**Abstract**—Virtual Reality (VR) has been showing potential in new and diverse areas, notably in education. However, there is a lack of studies in the Foreign Language Teaching and Learning field, particularly in listening comprehension. Therefore, this study investigated the effects of avatar animations and lip synchronization, and user interaction; features deemed relevant in this broader area. A sociodemographic, a quick CEFR - Common European Framework of Reference for Languages - 15-minute English test, and questionnaire were used to evaluate the participants' Presence, Quality of Experience, Cybersickness and Knowledge Retention. Results show that, overall, the use of avatars with realistic animations and movements, and featuring lip synchronization have a positive influence on the users' sense of presence, knowledge retention and a more enjoyable overall quality of experience. The same can be said for the use of object interaction and navigation in the cultural representative environment, which had an overall positive impact.

**Keywords** - Virtual Reality, Education, Foreign Language Learning, EFL;

## I. INTRODUCTION

Virtual Reality (VR) systems have existed for decades, though often hidden in state-of-the-art industrial and military installations [1]. In the last few years, VR has been showing its potential in new and diverse areas such as the medical field [2], [3], training [4], [5], entertainment [6] and, most relevant to this article, education [7], [8]. At the same time, the costs of these systems have been decreasing, becoming cheaper and more accessible to the general public. According to Fuchs [9], VR can be simply defined as “providing a cognitive activity in an artificial digitally-created world”. This 3D simulated world copies certain attributes from the real world and, with the help of immersive VR systems such as head-mounted displays (HMD), allows users to feel a state of consciousness called Presence, “the (psychological) sense of being in the virtual environment” [10] leading them to act as if they were in the real world. This sense of presence in VR can contribute to better learning as with a higher level of presence [11], the user experience gets more focused

on the activity that occurs in the virtual environment (VE). Furthermore, several authors confirm that presence not only reduces the social distance between students but also improves knowledge transfer [12], [10]. In this vein, VR has been proven to be valuable for educational purposes [11], [13]. However, it is necessary to take notice of cybersickness, and this is, symptoms such as nausea, disorientation, headaches, and so on [14], which are mostly associated with sensory conflicts and that damage the user experience in VR, overcoming the advantages of this technology [15]. Some studies also [16], [17], [18] found that students felt a greater sense of presence with VR. However, there was less learning compared to the low immersion version. The authors theorize that these lower results are justified by the enthusiasm of using VR for the first time, as well as the overwhelming spectre of more significant cognitive load inherent in the use of HMD.

In 2017, a study [8] explored differences in learning between text documents, educational videos and virtual environments, all with the same content relating to historical facts of one of humanity's first cities, Uruk. In this experience, it was possible to verify the virtual environments group's better performance than the text and video groups. Thus, they concluded that learning in virtual environments results in better performance.

In the early 2000s, Schwienhorst stated that the development of virtual reality in the field of foreign language teaching was still largely unexplored [19] when compared to other areas. However, it has been recently growing as it has great potential in helping students by allowing them to visit realistic simulations that would not even exist in the world, as well as the opportunity to recreate episodes and places of great cultural importance where the user can be immersed, without the need to travel thousands of kilometres, waste time and money and even avoiding possible dangers. This aspect of culture is an essential feature in learning a foreign language, but it is still not being used to its full potential [20], [21].

In 2007, Cheng and his team [22] researched learning Japanese as a foreign language where, in addition to cultural characteristics recreated in the virtual environment, explored

if they could use the virtual reality technology to teach embodied cultural interaction. For this, they developed a physical interaction where users had to bow as a greeting with an avatar (non-player character or NPC), a typical cultural characteristic of Japanese greetings.

This use of avatars is reported to allow an increase in the level of presence, as the user navigates and interacts with it [23], [24]. In studies with Liang-Yi Chung [25], and Alan Cheng [22], users interact with avatars in a conversation to learn a new vocabulary. This integration of avatars into learning materials can improve users' effectiveness in the language they are learning [25].

In addition to using avatars as a VR learning/teaching method, the use of objects is also quite common. Some studies show that the method of searching and finding objects as a learning technique shows several benefits [26], [27]. In addition to searching, sometimes objects are also linked to their name in text and/or audio format [28], for better learning of their meaning in a foreign language.

Generally speaking, there is a consensus that VR impacts positively foreign language teaching, improving student learning, the level of fun and positive results not only in the student's language skills but also in their cognitive abilities [28], [29]. The increase in motivation and willingness to learn [25], [30] and relish from the sense of presence and immersion provided by VR is also a positive outcome of this technology [23]. From a perspective of comparison with traditional teaching methods, studies [25], [31] prove that the use of this technology is significantly better than conventional teaching practices. Even in experiments like Ebert's [28], where there was a record of higher initial test scores using the traditional method when it was tested to what level users can remember the words a week later, the VR method was significantly higher.

Looking in the subfields of this area, such as foreign language listening comprehension - a rather common method of traditional foreign language teaching to improve the understanding of the language and pronunciation - there is even more limited VR content. Earlier study established that it is possible to use VR-based learning tools as an alternative to audio-only listening sessions to learn a foreign language [17], [32]. Just very recently, Tai and Chen [33] studied the impact of VR on EFL learners' listening comprehension. The VR group used mobile-rendered head-mounted displays while the control group watched a video version on PC. Their results revealed that the VR players' listening comprehension and retention were significantly higher than the video watchers'. Not only that, peculiarities like simulated, interactive, and immersive VE to perform authentic learning exercises help the users prevent cognitive overload, reduce anxiety, and thus aided comprehension. Most of the users also stated VR-assisted EFL listening to be appealing and beneficial.

In 2010, a study by Hirata [34] examined whether lip movement and hand gesture help to improve native English speakers' ability to comprehend Japanese vowel length contrasts. Although, the users in the audio-mouth condition im-

proved more than those in the audio-only condition, conditions concerning the hand gestures did not. After that, in 2019, Buechel [35] alleged lip-sync activities encourage foreign language development, as it supports pronunciation practice and students from her English language class got excited to understand a listening text.

There is also a lack of studies that demonstrate the effects of avatars animations, avatars lip synchronization and/or interaction/navigation in this listening comprehension subfield. With this in mind, this paper aims to study such conditions to allow those who seek to develop apps related to this to develop better apps and tools with an extra positive impact. For this, two different studies were carried out where; in the first study, we investigated the effect of avatars animations and lip synchronization. In the second study, we examined the outcome of user navigation and interaction with objects being referenced in the dialogue.

## II. STUDY I: EVALUATING THE EFFECT OF BODY ANIMATIONS AND LYP-SYNC

### A. Methods

This first experiment was developed to evaluate the effect of body animations and lip-sync of the avatars. Here, the users start in a tutorial environment before transiting to a formal language dialogue environment, consisting of an office, Fig. 1. Once the dialogue is over, the users appear in a VE with an interactive knowledge retention test of what they observe.



Fig. 1. VE with the avatars (A) and the user's view (B) - study 1.

1) *Sample*: The sample consists of 48 participants (41 males and 7 females) aged between 20 and 29 ( $M=21.5$ ,  $SD=1.989$ ). These were divided into the 4 conditions, 12 participants each. Due to the current situation of COVID-19 and consequently the impossibility of carrying out experiments with students from the English Centre in Vila Real as initially planned, the participants were recruited in classes of Computer Engineering, Electrical and Computer Engineering and/or Multimedia.

2) *Materials*: The stand alone HMD Oculus Quest was used for the experiments. This HMD features 6DOF, resolution of 1600x1440 pixels per eye and a refresh rate of 72 Hz. The audio stimulus was provided by noise canceling headphones (Bose QuietComfort 25) connected to the HMD. The HMD's built-in speakers were not used due to the lack of noise can-

cellation and therefore possible distractions. The controllers of this HMD system were used for interaction with the VE.

3) *Variables:* The independent variable is the learning mode. This variable has four different conditions, described as follows:

- WAWL condition: With animations and lip sync;
- WANL condition: With animations, no lip sync;
- NAWL condition: No animations, with lip sync;
- NANL condition: No animations and no lip sync.

The dependent variables are Presence, Quality of Experience, Cybersickness and Knowledge Retention. These were measured by multiple-choice post-test questionnaires, except for Knowledge Retention, which was carried out in an interactive virtual environment.

4) *Instruments:* The adopted questionnaires were following: (1) a simple sociodemographic questionnaire to describe the sample; (2) a 14 item 5-point Likert scale questionnaire based in the Portuguese version of the Igroup Presence Questionnaire (IPQp) to assess the sense of presence [36]; (3) a 15 item 5-point Likert scale to assess the quality of experience (QoE) adopted from similar studies [37], [38], [39]; (4) a 16 item questionnaire based in Simulator Sickness Questionnaire (SSQ) to assess the Nausea, Oculomotor Discomfort, Disorientation and Cybersickness. The questionnaires were presented in printed format in the participants' first language, Portuguese. A translation of the QoE is presented below:

- Assess the overall quality of the experience you just had.
- The visual quality of the experience was adequate.
- I liked the virtual reality experience.
- The avatar animations made the experience more realistic.
- The avatar animations are a distraction.
- The avatar animations are annoying.
- I liked the experience with avatar animations.
- The interaction with avatars and environment made the experience more realistic.
- Interaction with avatars and environment is a distraction.
- The interaction with the avatars and environment was annoying.
- I enjoyed the interaction experience.
- The lip sync of the avatars made the experience more realistic.
- The lip sync of avatars is a distraction.
- Lip sync is bothersome.
- I liked the experience with lip synchronization for the avatars

Regarding knowledge retention assessment, a 10 questions virtual test (four multiple choice) was created by an English teacher, based on the scenario dialogue, designed to assess the students' knowledge retention after the experience. The test has a total value of 100%, where each answer has a value of 10%.

5) *Procedure:* All the equipment is disinfected before the experience itself, from the VR equipment to the pen to fill in the questionnaires. Participants were tested individually in a room with a controlled environment. A sociodemographic

questionnaire was given before the beginning of the experience. Due to the impossibility of conducting the experiment with the English Centre students, and suggested by an English teacher, the participant was directed to a table with a computer to perform a quick 15-minute English test, with results according to the CEFR levels. If the participant had a result lower than "Intermediate", the result of the final experiment was discarded.

After completing this quick English test, the participant was randomly selected for one of the four learning modes to be evaluated. The participant was then directed to the centre of the room to sit on a chair, where a brief explanation of how to work with the Oculus Quest commands was given. Prior to the experiment itself, participants enter a tutorial scene, without the NPCs, Fig.2. The aim was to allow the subjects, who could not be familiar with virtual environments, to become accustomed to manoeuvring the HMD and learn how to use the controllers to interact.



Fig. 2. Tutorial Environment teaching the user how to use interact with the scene.

At the end of this VE, the participant starts the dialogue scenario (1) lasting about 2:35 minutes where, once it reaches the end, it automatically transitions to the interactive knowledge retention test VE. At the end of the assessment, the participant is helped to remove the equipment and receives instructions regarding the completion of the presence, quality of experience, and cybersickness questionnaires. The whole exercise takes around 25 minutes.

### B. Statistical Procedures

To verify if there were statistically significant differences between the 4 conditions (WAWL, WANL, NAWL, NANL) with regard to Presence, Quality of Experience and Knowledge Retention, an analysis was performed using One-Way ANOVA. Due to outliers in the Cybersickness questionnaire, it was analyzed using a non-parametric Kruskal-Wallis test. Later, a Spearman's correlation was performed between knowledge retention and the other variables, including the initial quick English test.

### C. Results

1) *Presence*: Regarding the IPQp, global presence score and individual scores for the subscales of spatial presence, involvement, and experienced realism were analysed. The scores were normally distributed, as assessed by Shapiro-Wilk's test ( $p > 0,05$ ). Variances were homogeneous, as assessed by Levene's test for equality of variances ( $p > 0,05$ ). A one-way ANOVA was conducted to determine if presence was different for groups with different conditions, but the differences were not statistically significant.

2) *Quality of Experience*: A one-way ANOVA showed statistically significant differences between the quality of experience and the groups with different conditions. Data was normally distributed for each group, as assessed by Shapiro-Wilk test ( $p > 0,05$ ); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ( $p = 0,753$ ). Data is presented as mean  $\pm$  standard deviation. A post hoc test with a Bonferroni correction revealed that the quality of experience reported by the NANL condition group ( $3,76 \pm 0,485$ ) was significantly lower than the WAWL condition group ( $4,48 \pm 0,439$ ) and that the WANL condition group ( $4,29 \pm 0,479$ ).

3) *Cybersickness*: Data from the SSQ questionnaire was analyzed for global cybersickness values and individually for the subscales: nausea, oculomotor discomfort, and disorientation. As the data did not follow a normal distribution, cybersickness values under different conditions were compared using a Kruskal-Wallis test. The results showed no significant differences between cybersickness and the different group conditions.

4) *Knowledge Retention*: A one-way ANOVA was conducted to determine if knowledge retention was different for the different conditions groups. Data was normally distributed for each group, as assessed by Shapiro-Wilk test ( $p > 0,05$ ); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ( $p = 0,660$ ). The examined scores were not statistically significant.

5) *Correlations*: The Spearman's correlation between the initial quick English test and the knowledge test showed the existence of statistical significance, with  $r_s(46) = 0,482$ ,  $p = 0,001$ . The analysis also showed significant correlation with the subscales of the IPQp questionnaire, highlighted in Table I.

### D. Discussion

After reviewing the questionnaires answered by the participants, some interesting results were collected. Regarding presence, and despite a lower average in the condition NANL, it was not possible to verify statistically significant differences between the different conditions. The variations among conditions may be too small, as they are all in the same VE, and only avatar related details change. However, some statistically significant results were found regarding the quality of experience. The NANL condition group scored significantly lower than the WAWL condition group and the NAWL condition group. The participants didn't like the experience as much when there was

no animations or lip synchronization of the avatars, compared to the presence of both. Participants also enjoyed the experience more when there was lip synchronization, even without animations, compared with the condition without either. In terms of cybersickness, no significant differences were found between conditions, suggesting that different animation types (or the lack of) did not influence this dependent variable. This is expected since the user engaged in the experience in the sitting position, without the need to move around physically and no variables that could cause symptoms of cybersickness were changed between the conditions [15]. The knowledge retention data showed no significant differences between conditions; however, analyzing the averages of the results, it is possible to see a higher average in the final test score in the WAWL condition and the NANL condition. This may symbolize that complete animation of avatars (body and lip sync) is more beneficial to learning than just one of the conditions; not only that, there is the feasibility of animations as possible distractions in the WANL and NAWL conditions without the advantages of WAWL, allowing users in the NANL condition to concentrate on the dialogue itself.

Correlation tests were performed between the questionnaire subscales (including the initial rapid test) and knowledge retention to understand better which elements can influence learning. Since Spearman's correlation test was performed, it should be noted that the correlations found may not be due to linear relationships. Only significant correlations will be analyzed. For example, significant correlations were found between the quick 15-minute English test scores and the final score in the knowledge retention questionnaire. Users who have originally a higher level in this language end up having a higher score in the final test is something to be expected. This study also demonstrates a negative correlation between the "Experienced Realism" and Knowledge Retention subscale. It is suspected that greater realism ends up drawing the users to the VE, distracting them from the avatars dialogue, and, as a consequence, worst results in the final test [16].

## III. STUDY II: EVALUATING THE EFFECT OF INTERACTION

### A. Methods

This second experiment was developed to investigate the effects of user interaction with specific objects of the VE. The users start in a tutorial environment before transiting to an environment of an informal language dialogue, consisting of a mall shop, where the user could teleport to select some particular object being talked about in the dialogue, Fig. 3. Once the dialogue is over, the users appear in a VE with an interactive knowledge retention test of what they witnessed.

1) *Sample*: The sample consists of 24 participants (21 males and 3 females) aged between 20 and 29 ( $M=21,4$ ,  $SD=2,205$ ). These were divided by the 2 conditions, 12 participants each. Seeing as the sample of this study was overlapping with the sample from Study I, the ordering of the experiment was randomised. Participants were also recruited in classes of Computer Engineering, Electrical and Computer Engineering and/or Multimedia.

TABLE I  
SPEARMAN'S CORRELATION BETWEEN PRESENCE AND KNOWLEDGE RETENTION

Knowledge Retention	Spatial Presence		Involvement		Experienced Realism		Presence	
	$r_s$	$p$	$r_s$	$p$	$r_s$	$p$	$r_s$	$p$
Study I	-0,138	0,349	-0,084	0,570	<b>-0,325</b>	<b>0,024</b>	-0,251	0,085
Study II	-0,353	0,091	-0,022	0,920	<b>-0,418</b>	<b>0,042</b>	<b>-0,442</b>	<b>0,031</b>



Fig. 3. VE of the second study and the user's view.

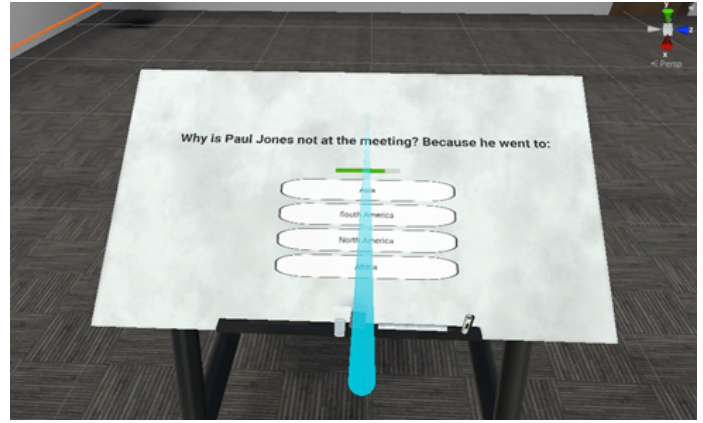


Fig. 4. Quiz Environment.

2) *Materials*: Similarly to the prior experiment, the standalone HMD Oculus Quest together with noise canceling headphones were used for the experiments.

3) *Variables*: The independent variable is the learning mode. This variable has two different conditions, described as follows:

- WI condition: With interaction;
- NI condition: No interaction.

The dependent variables are Presence, Quality of Experience, Cybersickness and Knowledge Retention. These were measured by multiple-choice post-test questionnaires, with the exception of Knowledge Retention, which was carried out in an interactive virtual environment.

4) *Instruments*: The questionnaires used in this experiment were the same used in Study I, i.e a sociodemographic questionnaire, IPQp, QOE, and SSQ answered in a pen-and-paper based format. Also, regarding the knowledge retention, a virtual test (fig. 4) was presented after the dialogue's VE.

5) *Procedure*: The procedure in this second experiment was similar to the first. The two main differences are: (1) The participant was then directed to the centre, but he would stand in the room instead of sitting on the chair; (2) the tutorial environment showed the participants not only how they can use the controllers to interact, but also to teleport. The whole exercise takes around 20 minutes.

### B. Statistical Procedures

To verify the existence of statistically significant differences between the 2 conditions (WI vs NI) in Knowledge Retention, an analysis was performed using an independent-samples t-

test. As for Presence, Quality of Experience and Cybersickness, either due to outliers or not passing the normality tests, an analysis was performed using non-parametric methods, Mann-Whitney U. A Spearman's correlation was performed between knowledge retention and the other variables, including the initial quick English test.

### C. Results

1) *Presence*: Scores from the IPQp questionnaire were analyzed. The results, using the non-parametric method Mann-Whitney U, showed no statistically significant differences, ( $U = 66.5, z = -0.318, p = 0.755$ ). The same situation occurring in all Presence subscales: the Spatial Presence ( $U = 65.5, z = -0.379, p = 0.713$ ), the Involvement ( $U = 42, z = -1.741, p = 0.089$ ) and the Experienced Realism ( $U = 65.5, z = -0.378, p = 0.713$ ).

2) *Quality of Experience*: Regarding the QoE, despite a higher mean in the WI scenario ( $4.450 \pm 0.467$ ) than in the NI scenario ( $4.261 \pm 0.570$ ), according to the Mann-Whitney U analysis, the results demonstrate that they are no statistically significant differences, ( $U = 62.5, z = -0.550, p = 0.590$ ).

3) *Cybersickness*: Scores from the SSQ questionnaire were analyzed for both global cybersickness values, as well as individually for the subscales. There is statistically significant differences in global cybersickness ( $U = 113, z = 2.254, p = 0.017$ ), as well as in the nausea subscale ( $U = 110, z = 2.508, p = 0.028$ ). These data can be found in table II. The results show that the WI scenario introduces significantly less nausea to users than the NI. The same goes for the global values of cybersickness.

TABLE II  
MANN-WHITNEY U BETWEEN CYBERSICKNESS AND THE DIFFERENT CONDITIONS.

Cybersickness Scale	With Interaction (WI)		No Interaction (NO)		$U$	$z$	$p$
	$\bar{X}$	SD	$\bar{X}$	SD			
Nausea	1,590	3,713	7,950	6,847	<b>110</b>	<b>2,508</b>	<b>0,028</b>
Oculomotor Discomfort	0	0	6,317	13,260	96	2,135	0,178
Disorientation	5,800	11,038	12,760	20,091	88	1,069	0,378
Cybersickness	27,639	54,466	101,080	114,960	<b>113</b>	<b>2,254</b>	<b>0,017</b>

4) *Knowledge Retention*: An Independent Samples t-test was conducted to determine if knowledge retention was different for the different conditions groups. There was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ( $p = 0,715$ ). Despite a higher mean in the WI condition ( $7,83 \pm 1,697$ ) than in the NI condition ( $7,75 \pm 1,685$ ), the results demonstrate that they are no statistically significant differences, ( $T(22) = 0,122$ ,  $p = 0,590$ ).

5) *Correlations*: Only the Spearman's correlation with the subscales of the IPQp questionnaire showed the existence of statistical significance, highlighted in Table I.

#### D. Discussion

Starting with the IPQp and its subscales, it was again impossible to verify between the different conditions. We speculate this happens for the same reasons as in the first experiment. No statistically significant differences were found regarding the quality of experience, it is possible that the variations among conditions are too small, as they are all in the same environment with the same avatars, and only interaction/teleport may not be enough for significant changes. The same can be applied to knowledge retention, where data revealed no significant differences between conditions. Regarding cybersickness, outcomes show that the WI condition introduces significantly less nausea to users than the NI. The same goes for the global values of cybersickness. The team presumes that this happens because, in the NI scenario, the user is stationary but is standing, focusing on the moving avatars and in dialogue in front of him, without being able to locate himself better in the environment; On the WI condition, the user can teleport and interact with dialogue-related objects, giving the user a better context and sense of space in the VE, decreasing cybersickness.

Correlation tests were performed between the questionnaire subscales (including the initial rapid test) and knowledge retention to understand better which elements can influence learning. Since Spearman's correlation test was performed, it should be noted that the correlations found may not be due to linear relationships. Only significant correlations will be analyzed. Significant (negative) correlations were found between the "Experienced Realism" and knowledge retention subscale. Similarly to the last experiment, we suspected that greater realism distracts users and, consequently, worst performances in the final test [16].

#### IV. CONCLUSION

The main objective of this study was to study the effect of avatars animations and lip synchronization as well as user navigation and interaction with objects in a virtual English listening sessions environment for English learning so that those who seek to develop apps related to this to create better apps and tools with an extra positive impact. To understand this, two experiments were carried out with multiple conditions, from avatar animations and lip-synced to object interactions and teleportation in the VE, which allowed the user to be immersed in the virtual world representing formal and informal scenarios. After testing it with a diverse number of users, and even with some unexpected results, the use of avatars with realistic animations and movements and featuring lip synchronization positively influences the users' sense of presence, knowledge retention, and a more enjoyable overall quality of experience. The same can be said for the results of the second experiment; this is, the use of object interaction and navigation on the representative cultural environment has an overall positive impact.

The study's main limitation was, due to COVID-19, the impossibility of carrying out the experiments with level-specific students from the English Centre in Vila Real as initially planned and the sample size. Future work intends to broaden the scope of the study to get a larger sample, experiences with said level-specific students, and extend this VR-based teaching tool to incorporate more learning activities.

#### ACKNOWLEDGMENT

This work is co-financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 under the PORTUGAL 2020 Partnership Agreement, and through the Portuguese National Innovation Agency (ANI) as a part of project "SMARTCUT - Diagnóstico e Manutenção Remota e Simuladores para Formação de operação e manutenção de Máquinas Florestais: POCI-01-0247-FEDER-048183".

#### REFERENCES

- [1] M. A. Gigante, "Virtual reality: definitions, history and applications," in *Virtual reality systems*. Elsevier, 1993, pp. 3–14.
- [2] N. E. Seymour, "Vr to or: a review of the evidence that virtual reality simulation improves operating room performance," *World journal of surgery*, vol. 32, no. 2, pp. 182–188, 2008.
- [3] F. Mantovani, G. Castelnuovo, A. Gaggioli, and G. Riva, "Virtual reality training for health-care professionals," *CyberPsychology & Behavior*, vol. 6, no. 4, pp. 389–395, 2003.

- [4] D. Narciso, M. Melo, J. V. Raposo, J. Cunha, and M. Bessa, "Virtual reality in training: an experimental study with firefighters," *Multimedia Tools and Applications*, vol. 79, no. 9, pp. 6227–6245, 2020.
- [5] M. Koźlak, A. Kurzeja, and A. Nawrat, "Virtual reality technology for military and industry training programs," in *Vision Based Systems for UAV Applications*. Springer, 2013, pp. 327–334.
- [6] M. Zyda, "From visual simulation to virtual reality to games," *Computer*, vol. 38, no. 9, pp. 25–32, 2005.
- [7] L. Freina and M. Ott, "A literature review on immersive virtual reality in education: state of the art and perspectives," in *The international scientific conference elearning and software for education*, vol. 1, no. 133, 2015, pp. 10–1007.
- [8] K. Ijaz, A. Bogdanovych, and T. Trescak, "Virtual worlds vs books and videos in history education," *Interactive Learning Environments*, vol. 25, no. 7, pp. 904–929, 2017.
- [9] P. Fuchs, G. Moreau, and P. Guitton, *Virtual reality: concepts and technologies*. CRC Press, 2011.
- [10] M. Slater and S. Wilbur, "A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 6, no. 6, pp. 603–616, 1997.
- [11] M. M. North and S. M. North, "The sense of presence exploration in virtual reality therapy," *J. Univers. Comput. Sci.*, vol. 24, no. 2, pp. 72–84, 2018.
- [12] C. Dede, "Immersive interfaces for engagement and learning," *science*, vol. 323, no. 5910, pp. 66–69, 2009.
- [13] T. A. Mikropoulos and J. Bellou, "The unique features of educational virtual environments," in *Teaching and learning with technology*. Routledge, 2010, pp. 269–278.
- [14] E. C. Regan and K. R. Price, "The frequency of occurrence and severity of side-effects of immersion virtual reality," *Aviation, space, and environmental medicine*, 1994.
- [15] J. J. LaViola Jr, "A discussion of cybersickness in virtual environments," *ACM Sigchi Bulletin*, vol. 32, no. 1, pp. 47–56, 2000.
- [16] G. Makransky, T. S. Terkildsen, and R. E. Mayer, "Adding immersive virtual reality to a science lab simulation causes more presence but less learning," *Learning and Instruction*, vol. 60, pp. 225–236, 2019.
- [17] D. Pinto, B. Peixoto, A. Krassmann, M. Melo, L. Cabral, and M. Bessa, "Virtual reality in education: Learning a foreign language," in *World Conference on Information Systems and Technologies*. Springer, 2019, pp. 589–597.
- [18] B. Peixoto, R. Pinto, M. Melo, L. Cabral, and M. Bessa, "Immersive virtual reality for foreign language education: A prisma systematic review," *IEEE Access*, vol. 9, pp. 48 952–48 962, 2021.
- [19] K. Schwienhorst, "The state of vr: A meta-analysis of virtual reality tools in second language acquisition," *Computer Assisted Language Learning*, vol. 15, no. 3, pp. 221–239, 2002.
- [20] D. Thanasoulas, "The importance of teaching culture in the foreign language classroom," *Radical pedagogy*, vol. 3, no. 3, pp. 1–25, 2001.
- [21] P. D. Ware and C. Kramsch, "Toward an intercultural stance: Teaching german and english through telecollaboration," *The Modern Language Journal*, vol. 89, no. 2, pp. 190–205, 2005.
- [22] A. Cheng, L. Yang, and E. Andersen, "Teaching language and culture with a virtual reality game," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 541–549.
- [23] A. Pack, A. Barrett, H.-N. Liang, and D. V. Monteiro, "University eap students' perceptions of using a prototype virtual reality learning environment to learn writing structure," *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, vol. 10, no. 1, pp. 27–46, 2020.
- [24] M. Usuh, E. Catena, S. Arman, and M. Slater, "Using presence questionnaires in reality," *Presence*, vol. 9, no. 5, pp. 497–503, 2000.
- [25] L.-Y. Chung, "Using avatars to enhance active learning: Integration of virtual reality tools into college english curriculum," in *The 16th North-East Asia Symposium on Nano, Information Technology and Reliability*. IEEE, 2011, pp. 29–33.
- [26] J. A. Jones, S. Smith, and M. Royster, "The scavenger hunt as an active learning technique," *NACTA Journal*, vol. 61, no. 1, p. 94, 2017.
- [27] S. Garcia, R. Kauer, D. Laesker, J. Nguyen, and M. Andujar, "A virtual reality experience for learning languages," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–4.
- [28] *ence on pervasive technologies related to assistive environments*, 2016, pp. 1–5.
- [29] Y.-L. Chen, "The effects of virtual reality learning environment on student cognitive and linguistic development," *The Asia-Pacific Education Researcher*, vol. 25, no. 4, pp. 637–646, 2016.
- [30] R. R. Divekar, J. Drozdal, Y. Zhou, Z. Song, D. Allen, R. Rouhani, R. Zhao, S. Zheng, L. Balagoyzian, and H. Su, "Interaction challenges in ai equipped environments built to teach foreign languages through dialogue and task-completion," in *Proceedings of the 2018 Designing Interactive Systems Conference*, 2018, pp. 597–609.
- [31] M. Zhang, "Virtual situated learning of spoken english based on computer simulation technology," *International Journal of Emerging Technologies in Learning*, vol. 15, no. 4, 2020.
- [32] B. Peixoto, D. Pinto, A. Krassmann, M. Melo, L. Cabral, and M. Bessa, "Using virtual reality tools for teaching foreign languages," in *World Conference on Information Systems and Technologies*. Springer, 2019, pp. 581–588.
- [33] T.-Y. Tai and H. H.-J. Chen, "The impact of immersive virtual reality on efl learners' listening comprehension," *Journal of Educational Computing Research*, p. 073563312199429, 2021.
- [34] Y. Hirata and S. D. Kelly, "Effects of lips and hands on auditory learning of second-language speech sounds," 2010.
- [35] L. L. Buechel, "Lip syncs: Speaking... with a twist," in *English Teaching Forum*, vol. 57, no. 4. ERIC, 2019, pp. 46–52.
- [36] J. Vasconcelos-Raposo, M. Bessa, M. Melo, L. Barbosa, R. Rodrigues, C. M. Teixeira, L. Cabral, and A. A. Sousa, "Adaptation and validation of the igroup presence questionnaire (ipq) in a portuguese sample," *Presence*, vol. 25, no. 3, pp. 191–203, 2016.
- [37] A. Covaci, R. Trestian, E. B. Saleme, I.-S. Comsa, G. Assres, C. A. Santos, and G. Ghinea, "360° mulsemmedia: A way to improve subjective qoe in 360° videos," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 2378–2386.
- [38] G. Mesfin, E. B. Saleme, O. A. Ademoye, E. Kani-Zabihi, C. A. Santos, and G. Ghinea, "Less is (just as good as) more-an investigation of odor intensity and hedonic valence in mulsemmedia qoe using heart rate and eye tracking," *IEEE Transactions on Multimedia*, vol. 23, pp. 1095–1105, 2020.
- [39] N. Murray, B. Lee, Y. Qiao, and G. Miro-Muntean, "The impact of scent type on olfaction-enhanced multimedia quality of experience," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 9, pp. 2503–2515, 2016.