

VIRTUAL REALITY ASSISTED PRONUNCIATION TRAINING (VRAPT) FOR YOUNG EFL LEARNERS

by **Minoo Alemi**

Islamic Azad University, West Tehran Branch, Tehran, Iran
Social & Cognitive Robotics Laboratory(CEDRA), Sharif University of Technology,
Tehran, Iran

minooalemi2000 @ yahoo.com

and **Shiva Khatoony**

Islamic Azad University, Central Tehran Branch, Tehran, Iran
shi.khatoony.lng @ iauctb.ac.ir

Abstract

Technology integration has become increasingly prevalent in language education and technology teaching tools are currently used for English instruction. In line with this trend, it is also believed to be influential in teaching and learning pronunciation. To cast an empirical light on this issue, this study aimed to analyse the influence of Virtual Reality Assisted Pronunciation Training (VRAPT) on pronunciation of young Iranian EFL learners. With respect to this, 18 low-intermediate English language learners (6-12 years old) were invited to take part in the study held in the lab at Sharif University of Technology in Iran. First, their English language pronunciation was measured by a pronunciation pre-test through *Speech ace Browser*. Next, in a virtual environment, during 10 sessions (each session 90 minutes), the participants received instructions on all of the pronunciation items measured in the test (consists of English vowels like /ɪ/ and /i/ and also vowels of /ʊ/ and /u/) through the Virtual Reality (VR) game administered by a humanoid robot. The results of a paired samples t-test revealed that there was a significant difference between learners' performance before and after the training sessions. This paper recommends that simulated games within virtual reality could be applied in enhancing pronunciation skills of EFL learners and utilized in helping them to develop native-like pronunciation. It also generates new ways of thinking about VRAPT as an innovative teaching approach and creating an efficient and pleasurable English learning environment.

Keywords: Virtual Reality Assisted pronunciation Training (VRAPT); pronunciation practice; young EFL learners

1. Introduction

As the global features of English language is enhancing every day (Cook, 2016), English

language teaching and learning have been accompanied by many problems, both for native speakers as well as English as a Foreign Language (EFL) learners and teachers (Far & Murray, 2016). As these learners acquire languages in an environment beside the original language situation, they are always faced with many obstacles (Cohen, 2014). Reed and Levis (2019) highlighted that one of the sources of these challenges for a large number of foreign language learners has been proven to be the formidable task of vocabulary learning, especially mastering correct and error-free pronunciation of the words which are pronounced like learners' mother tongue.

It is widely argued that non-native speakers tend to pronounce the sounds of a second language in a way which is much similar to the sounds of their own mother tongue. By the same token, EFL learners are likely to assimilate the pronunciation of the English sounds into the most similar ones in their own mother tongue, which might lead to some communication breakdown or confusion (Al-Zayed, 2017). Moreover, Morley (1994) asserts that when non-native speakers' pronunciation is heavily affected by their accents, others may find understanding them too difficult or rather impossible; hence, intelligibility in English pronunciation should be the prime goal of every speaker in conversations (Scovel, 1988). Based on this, accurate pronunciation in English conversations is an inseparable part of language learning for EFL learners in non-native contexts.

According to a wealth of recent studies, a high level of competence in pronunciation can be attained when teachers provide the learners with the materials that help them upgrade their level of pronunciation (Fraser, 2000) and set the pronunciation instruction at the very top of the agenda of teaching speaking skills (Morley, 1991). Furthermore, many scholars are of the opinion that learners who do not care about the pronunciation issue are more likely to encounter communication problems (Waniek-Klimczak, Rojczyk, & Porzuszek, 2015). That is, they cannot communicate effectively beyond classroom walls, where they have to interact with others in authentic contexts (Hismanuglu, 2009).

By the emergence of technological innovations such as virtual reality (VR) technologies, it seems possible to create educational conditions for learners who are away from native environments (Lin & Lan, 2015). Virtual reality is considered a new technology with remarkable features that can provide situations where human beings can experience inaccessible environments (Minocha & Tudor, 2017). Studies have already been conducted by using computers (Nami, Marandi, & Sotoudehnama, 2016) and mobile technologies (Alemi, Sarab, & Lari, 2012), many of which have shown that computers could be employed to enhance

language teaching (Yates, 2017) or decrease the learners' problems in language learning.

2. Literature review

In this section of the paper, the authors review previous studies about pronunciation in EFL contexts, and language learning in virtual world.

2.1. Pronunciation in language learning

There are many theoretical and operational definitions for pronunciation, each of which can cover distinct aspects of this component. According to Burgess and Spencer (2000), pronunciation can be defined as a system in which speakers of language use a variety of phonological structures. Based on this, phonological features can make a difference in a target language discourse. In other words, pronunciation is one of the most challengeable issues among language teachers and learners. This important component, which is generally ignored due to lack of access to native teachers or lack of adequate training method, is a serious challenge and in this way many language teachers have not yet been able to use the correct method to improve pronunciation (Freeman, Katz, Gomez, & Burns, 2015). However, many have created new ways to teach this essential part of language with the application of technology (Li & Somlak, 2017).

Moreover, Reed and Levis (2019) explained many aspects of pronunciation for language learners in non-native English environments and found that some learners have many mistakes and errors in their speech production. This study mainly focused on the influence of teachers' first language on teaching pronunciation. To this end, students were divided into two groups (one with native and one with non-native teacher). Both teachers taught with similar pronunciation methods, finally similar results were obtained from both groups, which indicated that the exercises or rules that the teacher considers to improve pronunciation are important and can produce significant results.

Scientific research about the audio-visual aids can be useful for both speech recognition and acceptable pronunciation; therefore, many laboratory studies revealed that analysis of the problematic situations represent the proper use of audio-visual aids. The exact example of this situation can be found in the research of Li (2016), who investigated the behaviour of the participants in difficult situations by using read-aloud and audio-recording without the images of speaker's face in a very ethical situation. Results of the study showed significant differences between two groups and the superior group could recognize the problematic situation and find a

way to overcome those problems. It can be concluded that identifying the problematic situation in pronunciation aspect of the language can be useful for increasing the pronunciation level so that a good teacher should predict the complexity or even problematic situation and help the learner improve his or her pronunciation level.

2.2. Teaching pronunciation through technologies

Pronunciation is construed as an essential component in ESL or EFL curriculum, and pronunciation teaching and learning is perceived as a high-priority objective in the realm of second language education. Nevertheless, generally due to lack of access to native teachers or lack of efficient training methods, it is difficult to obtain performance enhancement. Therefore, many language teachers may not have been able to use the appropriate method to improve pronunciation among learners (Freeman et al, 2015). On the other hand, the past decade has seen rapid advancement of technology in various areas of language learning, together with teaching and pronunciation as one of the most active areas in language research is no exception. Despite the effectiveness of the conventional approach, technology-based methods have tried to improve this component of the language like other topics; therefore, computer programs have succeeded in replacing language labs (Yoshida, 2003). More recently, Dixon and Foster (2018), also emphasize that despite the effectiveness of the traditional methods used to teach and correct pronunciation for teachers and students, the development of technology and the application of modern methods has made significant progress in achieving effective pronunciation teaching and student learning. Subsequently, a novel research agenda has been opened up and, as Li and Somlak (2017) point out, an increasing number of studies have relied on innovative ways to teach this essential part of language with the application of technology.

The use of software programs for teaching and improving pronunciation can be observed in a current study by Cruttenden (2014), which mainly focused on using computational methods to give feedback on the pronunciation errors of learners with the aim of bringing pronunciation of the learners to the native model. It has conclusively been shown that the technique can reduce the foreign accents without significantly altering the voice quality characteristics of the foreign speaker. Furthermore, similar attempts have been made to teach pronunciation and phonetics through computer programs to learners. For example, in a study by Li et al. (2015), the Automatic Speech Recognition Tool (ACER) was used to calculate the language of the learners manually and without interference and different feedback was used to correct incoming speech as well as correct spelling mistakes in language learners. Although

such programs are made by human hands, they are so powerful that they can evaluate the variety of dialects and accents and encourage students and teachers to use educational programs (Yoshida, 2003).

Along with the development of computer science as well as software programs in language contexts, linguists and researchers turned their attention towards studying linguistic characteristics, especially the level of the pronunciation of learners, in an entirely separate domain, named Computer Assisted Language learning (CALL) (Nami, Marandi, & Sotodehnama, 2016). For instance, the use of computer tools to detect and correct pronunciation can be seen in the work of Mehrpour, Alavi Shoushtari, and Haghghat Nezhad (2016). In their study, researchers used educational software programs for advanced-level English language learners to help them reduce their incorrect pronunciation affecting English dialects by creating a quasi-experimental design.

In addition to computers and software available today, there are applications on small gadgets like mobile phones and tablets, some of which have tried to contribute to learning and teaching a second language (Yoshida, 2003). As Franco et al. (2010) highlighted, many people are keen to use these programs to improve their second language in their spare time. Hence, researchers and linguists and even teachers use these inclusive mobile-based applications to teach the second language in a fast-growing field entitled "MALL" or Mobile Assisted Language Learning (Alemi, Sarab, & Lari, 2012).

After the advent of computer software, as well as mobile applications for language teaching and their use to evaluate various functions, researchers and linguists have also succeeded in incorporating humanoid robots into the realm of teaching, in particular the teaching of English as a second language called RALL or Robot-Assisted Language Learning (Aidinlou, Alemi, & Farjami, 2014). In a study conducted by Alemi, Meghdari and Ghazisaedy (2015), the authors used a humanoid robot as a new field toward teaching language to assess the level of anxiety and attitude of about 46 female students in English language classes. Thus, using RALL indicated that the use of a robot as a teacher assistant in the classroom can be very entertaining and enhance learning more effectively. Therefore, in the modern era of technology, as described by Cook (2016), many new techniques which are mainly based on computational methods have been capable of analysing the correctness of pronunciation and making a significant difference in language research.

2.3. Language learning in virtual worlds

The advancement of technology, in particular virtual reality (or virtual reality games), in second language learning and teaching has been observable in the recent years. Concerning this, linguists and language programmers have decided to investigate the role of new technologies like virtual reality to train EFL learners (both children and adult learners) to observe their effect on the language learning process (Fowler et al., 2015). Due to the increasingly rapid advances in virtual reality games, this field has attracted much attention as an effective way of instruction by many scholars such as Chik (2014) who points out that these games have been able to raise the level of learning in the audience for the purpose of teaching English as a second language. Similarly, Tai, Chen, and Tod (2020) proposes that the application of virtual reality-based 3D games increased the ability of learners to learn the second language. More recently, Zhang, Zhang, & Chang (2018) also note the technology of virtual reality exists in simulated games which can be efficient in teaching languages, especially in non-native environments.

In several studies, the importance of virtual reality systems has been described, which can be helpful and illustrate the previous path of research in this field. As an example, in a study conducted by Kröger, Birkholz, Hoffman, and Meng (2010), the learners' facet system was utilized through virtual reality-based computer programs to examine its effectiveness in language acquisition. Based on the results, audiovisual programs using virtual reality technology prove to be effective in teaching a second language and can be a good tool for pronunciation instruction and the improvement of the functional language system of different learners. In addition, many of the past studies have indicated an increase in the level of motivation among participants who use virtual reality and virtual characters and robots (Shabani, & Alipoor, 2017; Tai, Chen, & Tod, 2020). In these studies, the necessity of virtual reality has been estimated in the level of motivation and language learning of individuals in learning and it has been proven to play a key role in interactions and communications among participants.

Furthermore, virtual reality technology is not limited to using computer programs or large-scale simulations only. Many virtual reality situations have been exploited to a lesser extent, especially in the size and performance of mobile phones, from which surprising results have been reported. One of these examples was the study of Chiu (2017), which depicted the influence of virtual reality in the dimensions of mobile phones on language teaching. Through game-based learning, the author attempted to investigate the influence of virtual reality-based

games on language teaching. The results confirmed that educational-based games with the use of virtual reality could increase the level of education in the second language.

Virtual reality has also been analyzed in a completely Iranian classroom environment; for example, Rahimi, Golshan, and Mohebi (2014) attempted to examine the role of virtual reality in linguistic aspects, such as technical and personal as well as pedagogical language. The findings provided evidence that the use of virtual reality technology in Iranian environments had a positive effect and many students were willing to study and learn through these technologies.

As evidenced from past research on the main variables of the present study, namely pronunciation, games and virtual reality, numerous investigations have been conducted to address various aspects of language teaching in a virtual world. Yet, few reports have been found so far on the role of educational applications of virtual reality on L2 pronunciation as an effective competence. Hence, the main aim of the current study is to focus on untapped areas to explore the potential of virtual reality game and its suitability as a medium for L2 pronunciation training in a non-native context of language learning.

3. Methodology

3.1. The aim of the study

Based on the information provided in the previous section, further research in the area of technology application, especially virtual reality technologies in language teaching and learning, is of prime importance to understand if virtual reality can facilitate learning and whether it has any effect on learners' pronunciation to help them become further aware of correct pronunciation of words and sentences in verbal communication in non-native as well as native contexts. To address this issue and propel research in this area, the present study focused on the following research question: Does Virtual Reality Assisted Pronunciation Training (VRAPT) have any significant impact on the pronunciation of young EFL learners (from pre-test to post-test)?

3.2. Participants and the research context

Eighteen EFL learners (boys and girls equally) aged from 6 to 12 years through convenience sampling were invited to participate in this study with the help of the Social Robotic

Laboratory at Sharif University of Technology (SUT) in Tehran, Iran through a public and promotional call as an advertisement poster.

3.3. Design and procedure

This research is a single-group study that adopted the WOZ design (Wizard of OZ conducted by Kelley, 1980) too wherein a researcher (a wizard), in a laboratory environment, simulates the behaviours of a theoretical computer software and intercepts all actions and behaviours between participants and system. According to Bernsen, Dybjaer, and Dybjaer (1994), WOZ is believed to have the capacity to make complete specification of the systems' behaviours. In this study, this is done with the participants' prior knowledge and with a low-level deceit to manage the participants' expectations and encourage natural behaviours. In order to do this, the teacher estimated all the students' responses and stimulated these behaviours in the game scenarios. In addition to this, both teacher and computer operator controlled all the actions and behaviours of the learners in classroom. Therefore, human-computer interaction, as a pre-planned program, was believed to be autonomous but operated partially by an unseen human being operator. However, it should be noted that pronunciation of the group of participants was assessed prior to and after the classroom training to make it possible for the researcher to compare learners' command of pronunciation before and after the treatment sessions.

3.4. Data collection tools and procedures

Speech Ace (<https://www.speechace.com/>) and *Pronunciation Check Browser* were the tools used in the study. In order to test the impact of virtual reality games, the information received on the students' pronunciations of words and sentences was gathered through the *Speech ace* pronunciation check in browser mode (*Speech ace* LLC, Seattle, USA, based on American English), because this browser can evaluate the level of one's pronunciation and provide feedback to improve this component among language learners. It should be mentioned that *Speech Ace* is used by a variety of users such as learners, teachers, universities and many language specialists. In order to minimize the measurement errors and achieve reliable data from participants, *Speech Ace* browser was selected as the most appropriate instrument to measure the participants' pronunciation level in a computerized manner. It should be noted that this instrument can evaluate the non-native pronunciation (in simple or complicated structures) in comparison with native speaker's pronunciation. Language learners can repeat and record his/her voice in this browser to check the native-like pronunciation. After that, this uploaded

voice is checked through the computerized process of this browser. At the final stage, the percentage of native-like pronunciation is shown for the learner.

On the other hand, the items of the tests were two categories of short vowels (see Appendix A). The reason of choosing two categories of vowels /ɪ/ and /i/ and vowels /ʊ/ and /u/ was that those sounds are the typical cases of a fossilized pronunciation error among Iranian EFL learners. In order to get the information on pronunciation of the participants, the percentage of native-like pronunciation was shown for the learner gained for all items. The maximum point for these items was 10 for both pre-test and post-test. For example, students' scores were calculated according to the percentage of accuracy of their efforts (see Appendix B for average maximum scores and average minimum scores). One full point was given for each item upon the performance of a percentage of accuracy up to 100%. Therefore, for each item, low performance (0 to 30%) was calculated from 0.00 to 0.30; average performance (30 to 50%) from 0.30 to 0.50, medium-high from 0.50 to 0.70 (50 to 70%) and very successful achievements from 0.70 to one full score (70 to 100%). All items could be repeated just one time.

A simulated game implementing the virtual reality technology was developed for use during classroom training. The game was created and animated in 10 episodes with storylines suitable for young learners whose age ranged from 6 to 12 years old. The stories offered in this section were designed in a way to be compatible with VR devices and help the learners feel comfortable about navigating around the virtual environment. To achieve more unity, the scenarios and stories of the games were designed and written by the researchers based on the mentioned vowels in simple sentences using VR technology (see Appendix D). It should be noted that the Robotics team of Sharif University of Technology was responsible for its graphic design.

During the 10 episodes, the learner got engaged through virtual glasses in various conversations in a variety of situations such as food ordering, shopping for items of clothing and familiarity with different occupations which included vowels /ɪ/ and /i/ and vowels /ʊ/ and /u/ in simple sentences. After each teaching situation, the robot asked questions and knowledge responses are anticipated. After playing the game, the learner came to the stream and could receive proper training. The whole process was monitored by the teacher and a game design specialist. All learners' behaviours and performances were recorded by virtual reality cameras.

The researchers of the study conducted the following steps in the process of data collection:

Pre-test

At the beginning stage of the study, the pronunciation of vowels /ɪ/ and /i/ and vowels /ʊ/ and /u/ in 28 simple words (see Appendix A) were checked among participants. Based on the percentages of native-like pronunciation, the level of participants was determined in the software; this measurement could help the researchers to design an appropriate game for the participants of the study. Therefore, the proficiency level of learners was considered in game design.

VRAPT Class

Throughout the treatment phase, in 10 sessions of ninety-minute-long, students were instructed by the use of a game which was designed for the purpose of training. At the beginning of each session, new topics were taught by the teacher. Then students used virtual reality glasses to play virtual reality games. Within 10 sessions, learners received treatment on a diverse range of subjects such as hobbies, colours, fruits, jobs, sports, body parts, toys, animals, foods, and birthday parties. All of these focused on pronunciation aspects of the English language and each scenario was specifically designed and created in a fully 3D environment (see Appendix C).

Post-test

At the end of the sessions, the level of participants' pronunciation was checked again as a post-test through the *Speech Ace* browser to see the improvement of the learners.

4. Findings and discussion

To investigate whether VRAPT has any significant impact on the English language pronunciation of young EFL learners from pre-test to post-test, 10 sessions of virtual reality trainings were administered. At the end, descriptive statistics were calculated. As shown in Table 1, students in the post-test ($M = 6.01$, $SD = 1.08$) performed better compared to their performance in the pre-test ($M = 3.64$, $SD = .81$).

Table 1. Descriptive statistics of two testing times of VRAPT (N=18)

		Mean	N	Std. Deviation	Std. Error Mean
VRAPT	Pre-test	3.64	18.00	0.81	0.19
	post-test	6.01	18.00	1.08	0.25

Moreover, the paired samples t-tests was run (Table 2) to compare pronunciation scores of young EFL learners before and after the virtual reality treatment sessions. It was demonstrated that there was a statistically significant difference between students' performances in the pre-test and post-test, $t(17) = -13.41$, $p = .00$, with the Cohen's effect size value being 2.47, which can be considered as a very high effect size. Hence, the results of the current study shed light on the efficiency of VRALL in the enhancement of pronunciation skills in young EFL learners, particularly in the case of the vowels /i/ and /i/ and vowels /o/ and /u/).

Table 2. Paired samples t-test

	Paired Differences				t	df	Sig. (2-tailed)
	Mean	Std. Deviation	95% Confidence Interval of the Difference				
			Lower	Upper			
Pre-test and post-test	-2.37	.75	-2.75	-2.00	-13.41	17.00	.00

The analysis of the results suggested that the utilization of simulated games has a statistically significant effect on the pronunciation skills of language learners. In other words, considering the necessity of verbal interaction for improving the pronunciation level of learners, virtual reality was utilized as a suitable tool to enhance English pronunciation of young Iranian EFL students. In the following section, these results have been discussed in light of the question posed earlier.

By comparing the outcome of the current research with the results of the reports mentioned earlier, it becomes evident that the finding is consistent with certain previously conducted research. For instance, it corresponds to the results of the previous report by Alemi, Sarab and Lari (2012), showing that the improvement in technology-based group treated through MALL was much higher than in the traditional group; the result of the delayed post-test also demonstrated that the experimental group performed better on vocabulary retention compared to the one taught with the dictionary. Hence, it can be mentioned that VRAPT classroom's performance can be in line with MALL group's performance in the case of the implementation of technology tools in the EFL context.

In addition, the findings of this study are consistent with those of previous reports by Aidinlou, Alemi, Farjami and Makhdoumi (2014), which proposed complete viewpoints of the theoretical rationale and frameworks of application of robots, instructional inter-faces like CALL, MALL, m-learning, r-learning along with their instructional roles, educational activities and related research and findings. In line with the reports of the mentioned study, VRAPT can also be counted as an instructional tool to improve the language proficiency of EFL learners, especially in pronunciation skills.

Moreover, the outcome of the analysis performed by Alemi, Meghdari, and Ghazisaedy (2015) ensured that the students in the RALL group had great fun in the learning process which helped them boost their motivation and consequently facilitated the process of learning in the long run. Similarly, games in virtual reality had the potency to inspire and motivate the participants and keep them engaged throughout the training sessions. Tseng and Schmitt (2008) also maintained that motivated learners can have better performance or better pronunciation in their language learning.

Furthermore, Cruz-Neira et al. (1993) highlight that some elements of virtual games like interactivity, rules, goals, challenge, risk, fantasy, curiosity, and control can serve as a strong impetus to spark up motivation and keep the learners attentive and motivated. By the same token, it could be argued that, as evidenced by the current study, working in a virtual world by means of VR game-based activities raised and maintained the students' motivation and contributed to investing more attention in mastering pronunciation.

The present finding also supports Alemi's (2010) study which made an attempt to investigate the role of word games in developing learners' vocabulary. To this end, an experiment using a variety of vocabulary games was conducted and its results confirmed that word games make a significant contribution to expanding learners' vocabulary. According to this study, VRAPT can be mentioned as an educational game which impacts the pronunciation of young EFL learners and enhances the language knowledge of learners.

Furthermore, the finding from the present study agrees relatively well with previous studies on the effect of CALL on the improvement in pronunciation of ESL students which can have similar results in the case of VRAPT (Mikropoulos & Natsis, 2011; Zhang, Zhang, & Chang, 2018; Zhao, 2003). It should be mentioned that the participants of VRAPT had better performance from pre-test to post- test. Thus, it can be argued that VRAPT had a similar effect as CALL on the pronunciation skills of EFL learners.

Other studies (e.g., Eskenazi, 1999) focused on various CALL-based technologies that provide ESL students with immediate feedback on their pronunciation, which can also be considered an advantage in VRAPT. Eskenazi (1999) acknowledges that pronunciation teaching is considered as a time-consuming task for EFL/ESL teachers and instructors. Consequently, EFL/ESL students may miss valuable feedback on pronunciation skills from the instructor. Therefore, it can be concluded that using new VR technologies facilitates receiving feedback for EFL learners and improves their pronunciation, which is in agreement with the results of the present study.

In light of the findings of the study by Ong'onda, Anashia, and Muindi (2016), it can be suggested that computerized instruction is more functional in teaching and learning of English vowels and consonants. The result of the present study can be consistent with the study of CAPT (Computer Assisted Pronunciation Training) in which the test group (the ones using the software) scored significantly higher on pronunciation and fluency ratings. Thus, embedding technologies such as CAPT program and VR games in classrooms effectively serve the learners with opportunities to practice a difficult sound on their own, monitor their pronunciation and familiarize themselves with new sentence patterns (Khoshsima, Saed, & Moradi, 2017).

Moreover, the efficiency of VR games could be attributed to such factors as 'authenticity' and 'repetition'. That is, the games provide a context in which the participants hear the correct pronunciation of linguistic forms in genuine contexts. Hence, they find associating form and meaning easier. As noted by Vahdat and Behbahani (2013) as well as Lawrence and Ahmed (2020), virtual games can establish an authentic situation in which pictures, sounds and graphics serve as strong stimuli for learners to hear and feel the sounds in a meaningful context. Besides, the games expose the learners to the same piece of language on different occasions which has proved to be conducive to mastering a language. This repetitive exposure, in fact, leaves a very strong trace on one's memory and makes it more likely for learners to retain and recall different aspects of words like pronunciation (Cruttenden, 2014).

Yet, another possible reason for substantial positive changes in pronunciation performance of learners is a result of the fact that the VR games can tap into two coding systems, as proposed by 'dual coding' theory (Paivio & Clark, 2006). This theory is premised on the tenet that cognition comprises two coding subsystems, namely 'verbal' and 'imagery' or non-verbal coding. Recall and retention are enhanced when both coding sub-systems work hand in hand. Accordingly, it can be purported that the VR games can create a context in which sound and meaning of the words are simultaneously presented and hence makes it for learners

easier to learn and remember meaning associations. This function of VR games has been also underscored by other researchers (e.g. Yueting, Varasrin, and Prabjandee, 2016).

5. Conclusion

The present study aimed to investigate and explore the possible effect of technological advances, especially the use of virtual reality tools, on developing the pronunciation skills of Iranian young EFL learners. When regarding the research question; *Does VRAPT have any significant impact on the pronunciation of young EFL learners (from pre-test to post-test)*, the results from the two testing times revealed that there was a significant difference between the performance of learners before and after the VRAPT training sessions. In other words, students made significant progress from the pre-test to the post-test in terms of their pronunciation.

This efficiency could be attributed to the games in virtual reality which have the potency to inspire and motivate the participants, keep them engaged and provide them with feedback. For these reasons, the participants could have been engaged more so as to invest more in playing more attentively, which means further exposure to samples of language and hence mastering pronunciation.

It should be noted that stakeholders in the language teaching field such as teachers, language institute administrators, and teacher trainers can benefit from this study. This study proposes that using VR games can be a viable strategy to help language learners overcome pronunciation difficulties and develop an acceptable competence in pronunciation. In other words, the results of the study showed that technology devices such as virtual reality can help learners to improve their language learning at a particular pronunciation level.

In addition to all the above, materials developers can use results of this study to design their educational materials based on the new technologies and new games in this area which is more beneficial to EFL learners.

This research has also offered certain recommendations for how research and practice in the field of VRAPT may be carried out in order to be of optimum benefit to language teachers and instructors who desire to implement virtual reality technology tools successfully within their EFL classrooms.

For further work, the extension of the study reported here can be proposed in two directions. First, the participants in the current study were a cohort of young language learners, so other studies should be conducted with learners at different age and with different background knowledge in language learning. Next, as the present study analysed students'

progress in the pronunciation of vowels /ɪ/ and /i/ and vowels /ʊ/ and /u/ in simple sentences and short conversations by using VR technologies, further research can be done on other phonological areas (vowels, consonants, diphthongs and trip thongs and also consonant clusters of the English language) in more complex and difficult words and sentences.

Acknowledgement

Our profound gratitude goes to “Centre of Excellence in Design, Robotics, and Automation (CEDRA)” and “Social & Cognitive Robotics Laboratory of the Sharif University of Technology, Prof. Ali Meghdari & Dr Alireza Taheri for their support and encouragements. We also appreciate the Iranian National Science Foundation (INSF) for their research grant award in support of the Social & Cognitive Robotics Laboratory at Sharif University of Technology (<http://en.insf.org/>).

References

- Aidinlou, N. A., Alemi, M., Farjami, F., & Makhdoumi, M. (2014). Applications of robot assisted language learning (RALL) in language learning and teaching. *Teaching and Learning (Models and Beliefs)*, 2(1), 12-20.
- Al-Zayed, N. (2017). Non-native pronunciation of English: Problems and solutions. *Amerikaya International Journal of Contemporary Research*, 7(3), 34-52.
- Alemi, M. (2010). Educational games as a vehicle to teaching vocabulary. *The Modern Journal of Applied Linguistics*, 2(5), 425-438.
- Alemi, M., Meghdari, A., & Ghazisaedy, M. (2015). The impact of social robotics on L2 learners' anxiety and attitude in English vocabulary acquisition. *International Journal of Social Robotics*, 7(4), 523-535.
- Alemi, M., Sarab, M. R. A., & Lari, Z. (2012). Successful learning of academic word list via MALL: Mobile assisted language learning. *International Education Studies*, 5(6), 99-109.
- Ali, S. (2016). Towards the development of a comprehensive pedagogical framework for pronunciation training based on adapted automatic speech recognition systems. *CALL communities and culture EUROCALL*, 2(1), 7-13.
- Bernsen, N. O., Dybkjær, H., & Dybkjær, L. (1994). Wizard of oz prototyping: How and when. *CCI Working Papers in Cognitive Science and HCI*, 2(1), 1-94.
- Burgess, J., & Spencer, S. (2000). Phonology and pronunciation in integrated language teaching and teacher education. *System*, 28(2), 191-215.
- Chik, A. (2014). Digital gaming and language learning: Autonomy and community. *Language Learning and Technology*, 18(2), 85-100.
- Chiu, F. Y. (2017, July). Virtual reality for learning languages based on mobile devices. In *Information Technology Based Higher Education and Training (ITHET), 2017 16th International Conference on* (pp. 1-3). Ohrid, Macedonia. IEEE.
- Cohen, C. (2014). Probabilistic reduction and probabilistic enhancement. *Journal of Morphology*, 24(4), 291-323.
- Cook, V. (2016). *Second Language Learning and Language Teaching*. New York, NY: Routledge.

- Cruttenden, A. (2014). *Gimson's Pronunciation of English*. New York, NY: Routledge.
- Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993, September). Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques* (pp. 135-142). Anaheim, California: ACM.
- Dixon, W. W., & Foster, G. A. (2018). *A Short History of Film*. New Brunswick, NJ: Rutgers University Press.
- Eskenazi, M. (1999). Using a computer in foreign language pronunciation training: What advantages? *CALICO Journal*, 16(1), 447-469.
- Farr, F., & Murray, L. (2016, June). Digital literacies for language learning and teaching: developing a national framework. Paper presented in *HEAD 2016 Conference*. València: Universitat Politècnica de València.
- Fowler, A., Partridge, K., Chelba, C., Bi, X., Ouyang, T., & Zhai, S. (2015). Effects of language modeling and its personalization on touchscreen typing performance. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 649-658). Seoul, Korea: ACM.
- Franco, H., Bratt, H., Rossier, R., Rao Gadde, V., Shriberg, E., Abrash, V., & Precoda, K. (2010). Edu Speak: A speech recognition and pronunciation scoring toolkit for computer-aided language learning applications. *Language Testing*, 27(3), 401-418.
- Fraser, H (2000). *Coordinating Improvements in Pronunciation Teaching for Adult Learners of English as a Second Language*. Canberra, Australia: Department of Education, Training and Youth Affairs.
- Freeman, D., Katz, A., Garcia Gomez, P., & Burns, A. (2015). English for teaching: Rethinking teacher proficiency in the classroom. *ELT Journal*, 69(2), 129-139.
- Hismanoglu, M. (2009). The pronunciation of the inter-dental sounds of English: An articulation problem for Turkish learners of English and solutions. *Procedia-Social and Behavioral Sciences*, 1(1), 1697-1703.
- Kelley, E. A. (1980). *Improving School Climate: Leadership Techniques for Principals*. Reston, Virginia: National Association of Secondary School Principals.
- Kröger, B. J., Birkholz, P., Hoffmann, R., & Meng, H. (2010). Audiovisual tools for phonetic and articulatory visualization in computer-aided pronunciation training. In *Development of Multimodal Interfaces: Active Listening and Synchrony* (pp. 337-345). Springer, Berlin, Heidelberg.
- Lawrence, G., & Ahmed, F. (2020). Avatar teaching and learning: Examining language teaching and learning practices in Virtual Reality Environments. In *Recent Developments in Technology-Enhanced and Computer-Assisted Language Learning* (pp. 340-360). Hershey, PA: IGI Global.
- Li, K., Qian, X., Kang, S., Liu, P., & Meng, H. (2015). Integrating acoustic and state-transition models for free phone recognition in L2 English speech using multi-distribution deep neural networks. In *Symposium on Languages, Applications and Technologies* (pp. 119-124). Hong Kong: SLAT.
- Li, W. (2016). Computer-assisted pronunciation training: From pronunciation scoring towards spoken language learning. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2016 Asia-Pacific* (pp. 1-7). Singapore: Institute of Research, IEEE.
- Li, Y., & Somlak, T. (2017). The effects of articulatory gestures on L2 pronunciation learning: A classroom-based study. *Language Teaching Research*, 23(3), 34-56.
- Lin, T. J., & Lan, Y. J. (2015). Language learning in virtual reality environments: Past, present, and future. *Journal of Educational Technology & Society*, 18(4), 45-70.

- Mehrpour, S., Alavi Shoushtari, S., & Haghigat Nezhad Shirazi, P. (2016). Computer-assisted pronunciation training: The effect of integrating accent reduction software on Iranian EFL learners' pronunciation. *CALL-EJ*, 17(1), 97-112.
- Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, 56(3), 769-780.
- Minocha, S., Tudor, A. D., & Tilling, S. (2017, July). Affordances of mobile virtual reality and their role in learning and teaching. Presented in: *The 31st British Human Computer Interaction Conference*. University of Sunderland's St. Peter's Campus, UK.
- Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25(3), 481-520.
- Morley, J. (1994). A multidimensional curriculum design for speech-pronunciation instruction. In J. Morley (ed.), *Pronunciation Pedagogy and Theory: New Views, New Directions* (pp. 64-91). Alexandria, VA: TESOL.
- Nami, F., Marandi, S. S., & Sotoudehnama, E. (2016). CALL teacher professional growth through lesson study practice: an investigation into EFL teachers' perceptions. *Computer Assisted Language Learning*, 29(4), 658-682.
- Ong'onda, D., Anashia, N., & Muindi, M. N. (2016). Computer assisted language learning software: The effect of integrating computer assisted pronunciation training on Kenyan English phonetics class at the university level. *International Journal of Academic Research in Progressive Education and Development*, 5(2), 1-21.
- Paivio, A., & Clark, J. M. (2006). Dual coding theory and education. *Pathways to Literacy Achievement*, 1(2), 1-20.
- Rahimi, A., Golshan, N. S., & Mohebi, H. (2014). Virtual reality as a learning environment in Iranian EFL context: Personal, technical, and pedagogical. *Procedia-Social and Behavioral Sciences*, 136(4), 234-239.
- Reed, M., & Levis, J. (2019). *The Handbook of English Pronunciation*. Hoboken, NJ: John Wiley & Sons.
- Khoshsima, H., Saed, A., & Moradi, S. (2017). Computer assisted pronunciation teaching (CAPT) and pedagogy: Improving EFL learners' pronunciation using clear pronunciation 2 software. *Iranian Journal of Applied Language Studies*, 9(1), 97-126.
- Scovel, T. (1988). A critical review of the critical period research. *Annual Review of Applied Linguistics*, 20(1), 213-223.
- Shabani, S., & Alipoor, I. (2017). The relationship between cultural identity, intrinsic motivation and pronunciation knowledge of Iranian EFL learners. *International Journal of Education and Literacy Studies*, 5(2), 61-66.
- Tai, T. Y., Chen, H. H. J., & Todd, G. (2020). The impact of a virtual reality app on adolescent EFL learners' vocabulary learning. *Computer Assisted Language Learning*, 2(1), 1-26.
- Tseng, W. T., & Schmitt, N. (2008). Toward a model of motivated vocabulary learning: A structural equation modelling approach. *Journal of Language Learning*, 58(2), 357-400.
- Vahdat, S., & Behbahani, A. R. (2013). The effect of video games on Iranian EFL learners' vocabulary learning. *Reading*, 13(1), 61-71.
- Wanick-Klimczak, E., Rojczyk, A., & Porzuczek, A. (2015). *Teaching and Researching the Pronunciation of English*. New York, NY: Springer.

- Yates, L. (2017). Learning how to speak: Pronunciation, pragmatics and practicalities in the classroom and beyond. *Language Teaching*, 50(2), 227-246.
- Yoshida, K. (2003). Language education policy in Japan: The problem of espoused objectives versus practice. *The Modern Language Journal*, 87(2), 290-292.
- Yueting, S., Varasrin, P., & Prabjandee, D. (2016). An investigation of metacognitive English listening strategies used by Chinese college students at Yunnan Normal University Business School, China. *HRD Journal*, 7(1), 92-101.
- Zhang, X. Q., Zhang, W., & Chang, M. T. (2018). Synthesis of large-area of virtual reality with English Environment. *Journal of Advanced Materials*, 24(17), 2320-2325.
- Zhao, Y. (2003). Recent developments in technology and language learning: A literature review and meta-analysis. *CALICO Journal*, 45(2), 7-27.

Appendix A. Vowels of the Study

Vowels /ɪ/ and /i/	Vowels /ʊ/ and /u/
Policeman	Could
Fish	Fruit
Teacher	Cook
Fifty	Too
People	Bull
Window	You
He	Would
It's	Juice
Cheap	Good
Women	Blue
Tree	Full
Six	Soup
Green	Woman
	Beautiful
	Should

Appendix B. Participants' scores in *Speech Ace*

Experimental Group	Age	Gender	English Proficiency	Pre-test	Post-test
1	10	M	Low Intermediate	3.2	5.7
2	12	M	Low Intermediate	3.5	6.3
3	9	M	Low Intermediate	2.7	5.4
4	7	M	Low Intermediate	3.6	4.8
5	10	M	Low Intermediate	4.1	6.1
6	10	M	Low Intermediate	2.9	5.4
7	6	F	Low Intermediate	2.6	4.3
8	10	M	Low Intermediate	4.2	6.8
9	11	F	Low Intermediate	4.1	7.2
10	9	M	Low Intermediate	3.4	5.6
11	8	F	Low Intermediate	3.5	4.8
12	6	F	Low Intermediate	2.7	5.3
13	6	M	Low Intermediate	3.1	5.9
14	7	F	Low Intermediate	3.2	6.1

15	8	F	Low Intermediate	4.2	7.8
16	8	F	Low Intermediate	4.1	7.1
17	12	F	Low Intermediate	5.1	8.3

Appendix C. *Speech Ace* user environment



Figure 1. *Speech Ace* vowel categories



Figure 2. Scoring procedures of *Speech Ace* browser

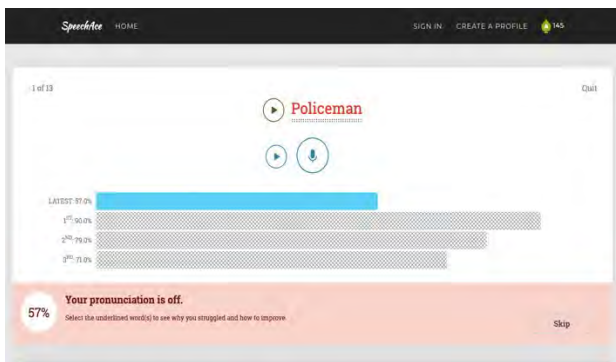


Figure 3. *Speech Ace* percentage in low mode



Figure 4. *Speech Ace* percentage in high mode

Appendix D: Virtual reality game environment and VRAPT participants

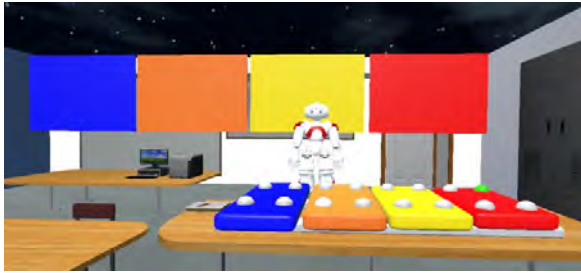


Figure 5. Virtual environment for colours

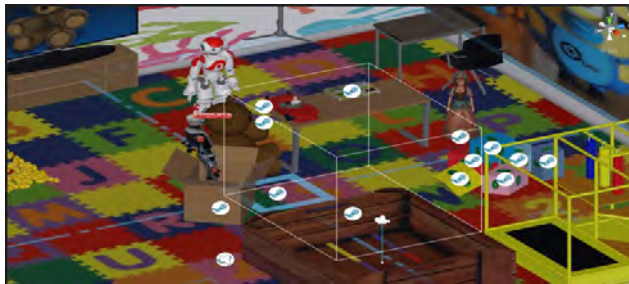


Figure 6. Virtual environment for toys

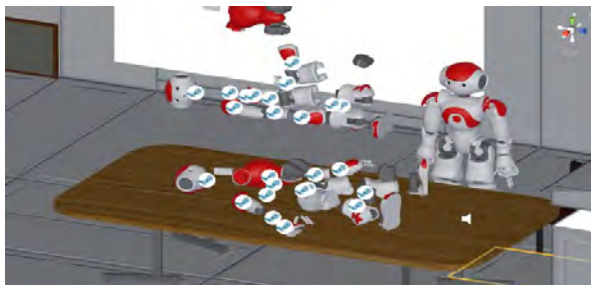


Figure 7. Virtual environment for body

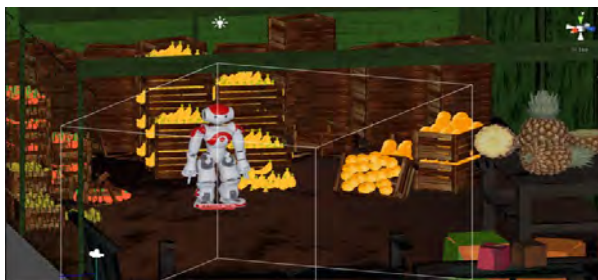


Figure 8. Virtual environment for fruits

Appendix E. Sample game scenarios

Game Scenario - Greeting

The game starts with the image of the airplane window. It's clear from the planet. With this, he finds himself out of the ground. When he looks to the left, he sees the surface of the new planet. Its surface is similar to the surface of Mars. Nima's robot comes close. Students can see the globe and the sun on the other side by turning the head. The game begins with the design of a sentence from the robot:

The robot asks: (LOOKING AROUND ANIMATION)

Hello! (BYE POINT ANIMATION). My name is Nima (POINTING TO SELF ANIMATION). What is your name (POINTING ANIMATION OR THINKING ANIMATION)

The student is expected to respond to his name. Acceptable answer for Nima

My name is If the answer was not given to the child or the answer was short (only the name implied), Robot says: (DEFEATED ANIMATION) Aha !! Say: My name is

The student should be able to repeat the sentence. If the answer is not received or incomplete, the sentence is repeated.

If the answer is correct from the child, the robot says.

Great! (HIP HOP DANCING)

The robot asks:

How are you say I'm fine. (AGREEING ANIMATION)

The student is expected to repeat the sentence. If it does not repeat or be short, the sentence is repeated again to hear an acceptable answer from the child's side.

If the answer is correct, the robot says.

I'm 10 years old. How old are you (LAUGHING ANIMATION)

If the answer was not given or was incomplete, the sentence repeated again.

(LOSER ANIMATION)

If the answer is correct, the robot says:

Excellent!!!! (AGREEING ANIMATION)

The robot asks:

I am an English Teacher! What do you do (THINKING ANIMATION)

The student goes to tell

I am a student

If the answer is correct, the robot says

Great! (PRAYING ANIMATION OR SAMBA DANCING)

If the answer is wrong, the robot says.

I am a student, Repeat after me, I am a student. (POINTING ANIMATION)

The student is expected to repeat the sentence. If the robot repeats it correctly, the robot says

Bravo! (SAMBA DANCING)

If the answer is false, the sentence is repeated to get the correct answer.

The robot says:

Nice to meet you! (PRAYING ANIMATION)

The student is expected to give her a response

Acceptable student response:

Nice to meet you too!

If the answer was wrong. The robot says

(LOSER ANIMATION)

You can say: Nice to meet you too! Repeat that! Nice to meet you too!

The student is expected to repeat the sentence. If it was true, the robot would hit him. If the mistake is repeated, the sentence is repeated again to get the correct answer.