JOURNAL OF E-LEARNING AND KNOWLEDGE SOCIETY
Vol. 19, No. 4 (2023), pp. 17-25

A methodology to design immersive Virtual Reality experiences for foreign language pronunciation training

Ilaria Compagnoni[1]

[a]*Ca' Foscari University of Venice, Department of Linguistics and Comparative Cultural Studies – Venice (Italy)*

(submitted: 20/10/2022; accepted: 28/12/2023; published: 29/12/2023)

Abstract

Applications of immersive virtual reality (iVR) for language learning are on the rise due to enhanced interaction possibilities with virtual agents and objects. However, iVR environments for language learning are not yet supporting speech recognition, hence limiting the potential to use iVR for fostering users' pronunciation training. This is due to a lack of methodological guidelines helping education professionals to create iVR environments effectively supporting speech recognition and pronunciation feedback. Moreover, no matter how close to reality iVR settings are designed to be, users may miss the correspondence between virtuality and reality due to design issues and lack of realistic input, with detrimental effects on enabling users to transfer iVR-acquired language skills to real-life interactions. Attempting to address these gaps, this study examines the methodology adopted by a group of researchers at the AP University of Applied Sciences and Arts in Antwerp to design an iVR experience for learning the pronunciation of Flemish words. Following a participatory design approach, an iVR prototype was planned and designed with the game engine Unity and targeted to prospective international students at the University of Antwerp. Findings from user trials revealed that virtual agents' timed feedback, environmental real-likeness and users' involvement in completing linguistic challenges were considered to be essential tenets for fostering iVR-based pronunciation training. Additionally, results suggested that further developments are needed to develop a technology for speech recognition in iVR environments that foster students' pronunciation skills and cultural exposure through pre-reality language training.

KEYWORDS: Immersive Virtual Reality, Language Learning, Pronunciation Training, Spatial Affordances, Participatory Design.

DOI
https://doi.org/10.20368/1971-8829/1135769

CITE AS
Compagnoni, I. (2023). A methodology to design immersive Virtual Reality experiences for foreign language pronunciation training. *Journal of e-Learning and Knowledge Society*, *19*(4), 17-25.
https://doi.org/10.20368/1971-8829/1135769

1. Introduction

Recent technological developments in language education have fostered advancements in the design of immersive and participatory learning experiences allowing students to interact with one another and their virtual surroundings. One type of technology which has received a lot of attention from scholars and instructional designers is immersive Virtual Reality (iVR), a three-dimensional digital space where users interact in a naturalistic fashion and which becomes their only perceivable reality (Eichenberg, 2012). Consequently, a rising number of iVR applications for language learning have appeared, targeting any schooling level and implying steadfast changes in teachers' roles as instructional designers who use technology to create experientially-rich and multisensorial language learning experiences (World Economic Forum, 2020). However, this process has also evidenced the inefficiency of in-person only instruction for boosting digital literacy, which is vital for the development of the necessary skills to professionally and socially interact in an increasingly digital world (European Commission, 2022). However, while the use of iVR in language education is encouraged, many teachers lack the necessary training to use this type of technology for designing language learning experiences.

[1] corresponding author - email: ilaria.compagnoni@unive.it

Moreover, while iVR language learning applications generally focus on hands-on experiences with virtual objects and avatar interactions, there is little evidence of existing iVR applications with embedded speech recognition features enabling live feedback on users' oral input. Recent developments in the field of Artificial Intelligence (AI) have opened new horizon for researching virtual speech recognition. In fact, VR companies have used AI-powered agents to provide real-time linguistic feedback to avatar-embodying users. These are pre-trained agents ready to respond to a limited set of questions sourcing information from AI platforms (WondaVR, 2023; Immerse 2023). The growing interest of companies to invite users to enhance VR experiences by programming virtual agents to recognize real-time linguistic feedback using AI powered tools highlight a major limitation of these technologies in terms of personalization of linguistic feedback. In fact, these agents are likely to provide a limited set of linguistic answers devoid of empathetic and realistic feedback which are likely to negatively affect the linguistic performance of application users. Moreover, AI speech may present limitations related to sound quality and type of conversational speech originating from written prompts (Mitsui et al., 2021). Therefore, there is a gap in the literature that needs to be addressed; as stated by Kumar Jang Bahadur (2022), providing effective solutions for iVR-based speech recognition may contribute to improve iVR design. However, guidelines are missing on the practices and training required to create pedagogically-effective language learning activities with iVR (Cook et al, 2019; Holly et al., 2021). This forces most teachers to rely on free web-based resources that lack immersive features and customizability. Moreover, from the perspective of designing effective pronunciation training experiences, language learning in iVVR must support motivation and engagement within culturally significant environments (Scovel, 1994), hence the need of incorporating these parameters into virtual platforms. This paper presents a methodology for designing iVR-based language learning experiences supporting pronunciation training in iVR and provides indications on the virtual environmental affordances needed to foster linguistic output. The two research questions that this paper addressed are:

- Which design practices can be applied by teachers to create language learning experiences for pronunciation training with iVR?
- Which environmental affordances can best support pronunciation training in iVR?

In an attempt to answer these queries, this paper illustrates the approach adopted by a group of researchers at the AP University of Applied Sciences and Arts in Antwerp (Belgium) to design a prototype of an iVR application for learning to pronounce basic Flemish words. The prototype was devised for prospective international students of the University of Antwerp. By following a participatory design methodology (Spinuzzi, 2005), researchers reviewed current gaps in the literature of iVR-based language learning design and pedagogy, and identified a profile of potential users. Subsequently, they planned the iVR prototype through story-making, spatial design, audio editing and user flow analysis. To design the iVR prototype, the researchers used the game engine Unity and subsequently tested the prototype on 8 participants who provided their feedback on usability and pronunciation training efficacy. Overall, this paper provides suggestions to structure language learning environments to support pronunciation training and highlights potential future developments in creating immersive language learning activities for speech recognition and pronunciation training. It is hoped that this research may also encourage language teachers to foster their instructional design skills with iVR to create interactive and socially immersive language learning activities.

2. Methods and Results

2.1 Prototype development

The study was planned as a pilot for the creation of iVR experiences incorporating real-time speech recognition using authentic people's speech. To plan and create the iVR experience, 4 researchers attending the summer school in "Storytelling with Virtual Reality" at AP University of Applied Sciences and Arts in Antwerp followed the 3 phases of participatory design, which include exploration, discovery and prototyping in Table 1. The group was given 2 weeks to design the iVR prototype, comprising pre-design data collection, prototype planning and post-design validation. Researchers utilized the facilities of the Immersive Lab

PARTICIPATORY DESIGN

Exploration	Discovery	Prototyping
Investigation on user types and needs Context identification Strategy formulation	Ideate a storyline Envision and develop solutions to task challenges Choose and storyboard ideas	Prototype building for usability focus Validation through prototype trial to real users Identification of potential adjustments and future developments of the prototype

Table 1 - Participatory design utilized by the researchers for iVR planning and design.

of AP University of Applied Sciences and Arts to test the iVR experience with Meta Quest 2 headsets and computers equipped with Intel i9 10940x core processors and Nvidia RTX 3070 graphic processing units.

The first part of the study consisted in identifying potential prototype users as prospective international students at AP University of Applied Sciences and Arts. Course interests as well as the academic and social involvement of these users were identified as paramount factors influencing their selections. Therefore, the group decided to set the iVR experience in a realistic environment. Inspiration was taken from the recommendations given by a group of 20 university students on local bars, which were subsequently visited for ambience recording through photos and videos. The environment of a bar was chosen to add familiarity to the experiences as it constituted a familiar setting for students to meet future friends and classmates. A survey on the identification of existing gaps in iVR language learning applications led to acknowledge a lack of speech recognition and live pronunciation feedback from a virtual agent in iVR environments. Further analysis also demonstrated a scant presence of phone-integrated iVR applications for the geo-localization of local bars and cultural landmarks of Antwerp. This process helped researchers to devise a phone-based application integrating pronunciation training with virtual representations of local bars in which users could collect virtual tokens redeemable as discounts through potential visits to the real bars portrayed in iVR. In terms of linguistic gains, users would benefit from a pronunciation training in Flemish provided by virtual agents whilst completing a series of challenges prompting them to utter words related to local food and landmarks. Once the above-mentioned exploration phase was completed, the group engaged in adopting a suitable methodology for designing the iVR prototype as a pronunciation-training experience.

2.2 The proposed method

In the discovery phase, researchers developed a methodology for designing effective learning experiences supporting iVR-based pronunciation training and speech recognition. This led the group to consider 4 planning aspects:

- Storytelling
- Immersion
- User flow
- Experience creation

Knowledge of storytelling principles was pivotal to create compelling iVR language learning experiences supporting pronunciation training since key aspects of user engagement, story immersion and willingness of continuous participation in iVR depend from it (Calvert & Abadia, 2020). For this reason, the researchers decided to opt for a linear narrative structure where events unfolded in a specific order following the interactional affordances of the iVR space. Once the storyline was completed, the experiences were planned considering the points of view of users entering the virtual space with VR headsets affording 6 Degrees of Freedom (6DOF) and enabling depth motion mapping (forward-backward, lateral-vertical).

To enhance empathy and emotional engagement, a second tenet of this method was identified as immersion in its strategic and narrative typologies. While the former happens in task-centered activities where individuals focus on goal achievement, the latter is the result of users' interest in the narrative dimension of immersive stories (Moeller, 2017; Visch et al., 2010). Therefore, both typologies are reliant on task design and user flow. Defined as the enjoyment resulting from balancing individual skills with challenges in task performance, user flow is contingent to the attainment of virtual goals with language and strategic skills (Brockmyer et al., 2009). For this reason, the researchers decided to have the iVR experience preceded by user training on headset use and content debriefing. The process followed for iVR design is outlined in Figure 1. The virtual experience was preceded by an explanation and tutorial phase where users were trained on the technology before emerging into real-life through an experiential reflection phase. Emersion was the last element to be considered within this methodology was emersion. Described as a "reawakening of human consciousness through recreational activities" (Chang 2021, p.26), emersion happens as users remove their headsets and return to real-life. To facilitate users' psychological adjustment to the virtual experiences and reflective practices on their virtual experience, the researchers decided to add post-iVR self-assessment phases on pronunciation skills and perceptual feedback. It was foreseen that these practices could help users to reconcile reality with virtuality and effectively reap the rewards of iVR exposure.

With language output being successfully triggered by social interactions, one of the main challenges of creating this iVR experience was ensuring that users' sense of immersion was maintained when interacting with virtual agents. In fact, due to time constraints of prototype design, the experience was not set as multiplayer. Therefore, it was decided to incorporate in the iVR experience humanoid characters reacting to user output to virtual agents' pronunciation prompts. In this way, interactional dynamics were maintained as real-like, enhancing immersion and user flow. Pedagogical considerations emerged when researchers integrated attainable and contextually-relevant tasks in the iVR experience ensuring the pronunciation accuracy of virtual agents interacting with real users. Following the guidelines of Saberi, Bernardet and DiPaola (2014) on fundamental parameters to facilitate users' feedback, virtual agents were designed with realistic attributes in order to be experientially well-received and avoid potential drawbacks related to the uncanny valley effect mentioned by Seyama and Nagayama (2017).

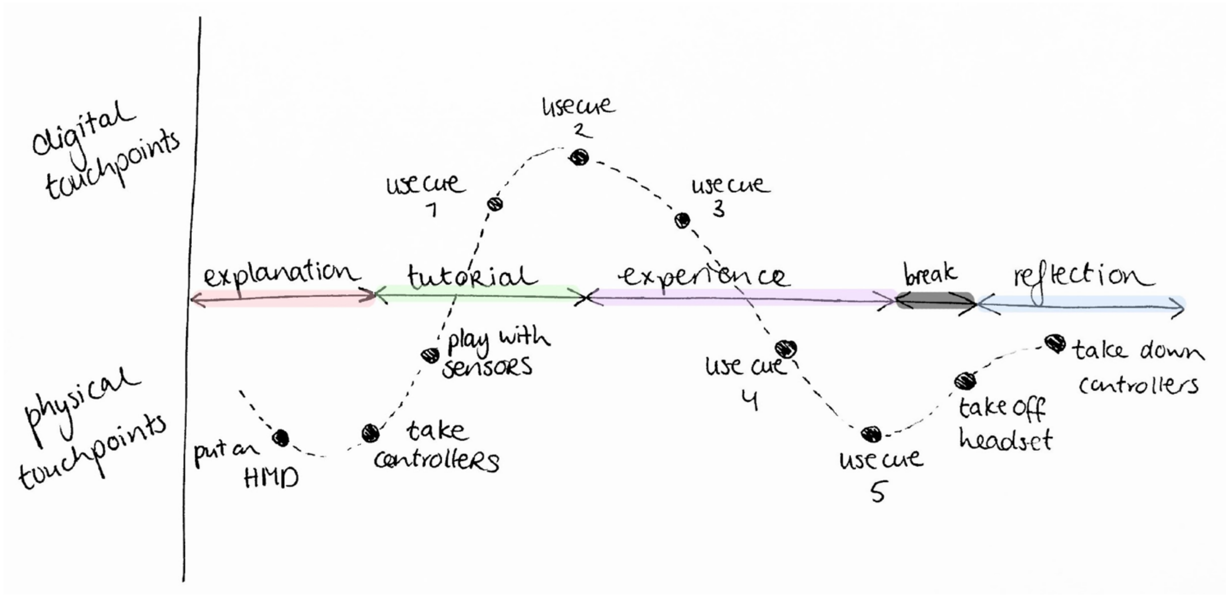

Figure 1 - User flow scheme followed by the research group in the proposed methodology.

Once the above-mentioned conditions were set, researchers proceeded to create the iVR prototype using the game engine Unity and virtual design applications.

2.3 The creation of iVR prototype

To create the iVR experience, researchers followed the participatory design outlined in Table 1. The prototype had to be activated on mobile phones, so that users could search for locations to visit on an app-integrated map. Upon selecting a location, users would be prompted to wear their Meta Quest 2 headsets before they were taken into the iVR experience.

Initially, the event sequence was planned using storytelling strategies such as reading out-loud the story script mimicking the potential movements required by users entering the virtual space. To encourage strategic immersion, the research group had to understand action feasibility in the iVR space and revise story flow, object location, interaction enactment and complexity as well as linguistic output. Guidelines for testing the above-mentioned parameters were taken from experimental research on screenplay writing in iVR (Luthy, 2017) and are outlined in Figure 2. In fact, in the planning phase, researchers utilized an A3 map mockup to reproduce the diagrams outlined in Figure 2. By placing user point of view at the axis intersection upon entering the iVR space, researchers could map user actions by locating objects, remove potential obstacles to virtual movements and perform gamified language challenges, which enabled the systematic planning of the virtual space and its affordances. Moreover, in order to ensure the native-likeness of word production, 2 native Flemish speakers (1 male and 1 female) were individually recorded whilst reciting a script in a quiet setting with a Zoom H2next recorder. An additional speaker was recorded while welcoming in English potential users in the virtual space. To increase resemblance to real-life ambience, sounds from the online database Freesound were incorporated in the experience together with background chatting and music. Audio files were then assembled with the software Audacity and inserted in the iVR space as monodirectional sounds coming from the virtual agents.

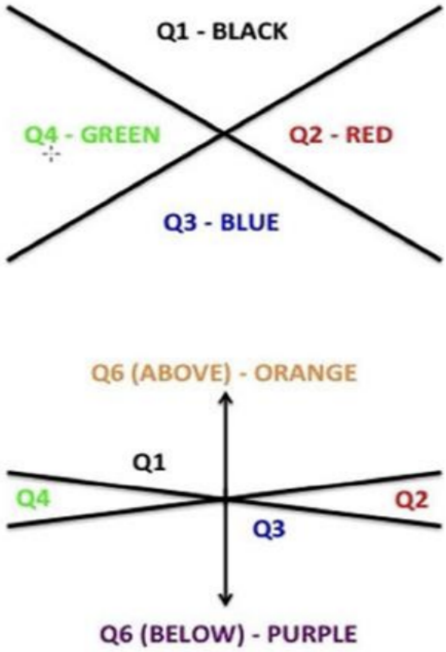

Figure 2 - Environmental mapping of iVR environments. The user point of view is considered to depart from the convergence of all lines. The first picture indicates forward-backward movements, the second lateral-vertical ones. Colors are used to identify spaces and their related affordances.

To implement user flow and enable its efficacy from pedagogical perspectives, the iVR experience was designed with the game engine Unity. Motion capture was recorded with the software Rokoko as one group member performed the actions of virtual agents wearing a sensor-equipped suit, which were transferred on Mixamo for character automation and matched with the bodies of virtual characters. Data was subsequently transferred to the prototype on Unity as virtual representations of students and bar personnel. User movement was enabled through hand controllers instead of teleportation to enhance real-like actions. User point of view was placed slightly higher than agents' height to enable a full view of spatial affordances. Using features of timeline adjustment, audio files were linked to cubes delimiting the range of motion of virtual agents and sequentially timed to play sounds at specific moments in the story flow. After each utterance, a time lapse of 5 seconds was left for word repetition. It was decided that virtual agents should speak in English apart from the targeted Flemish words so that an international audience could understand instructions. Haptic feedback was provided through vibrations in hand controllers each time users gained a token. Environmental navigation was fostered by interactions with virtual agents which directed users to interactional areas. The setting was created by modifying a bar environment downloaded from Unity Asset Store according to the features of a real bar in Antwerp. Daytime sky for lighting and a rooftop were subsequently downloaded from the store. Collectable tokens were designed with the software Cinema 4D while texture was added through Unity. Researchers also raised vertices allowance so that users' ability to explore and move around the virtual space could run smoothly. Figure 3 provides an outline of the prototype creation on Unity. Each step of the design process was tested on an Meta Quest 2 device set in developer mode to enable quick troubleshooting.

2.4 The final prototype

The researchers named the iVR experience "The FlemishTerp Bar" and made it last for 5 minutes. Upon entering the virtual bar, Character 1 would welcome users and gesture towards the counter to meet Character 2. The first language task would consist of Character 2 inviting the user to pronounce the equivalent of "beer" in Flemish. After each utterance, a time lapse of 5 seconds would be left for word repetition. Then, a third character would be introduced in the form of a lady coming from behind the user. The character would be the first token repository; by pointing at the top of her head, a token would appear in the shape of a hand-shaped biscuit representing a local delicacy called Antwerpse handjes (Antwerp hands). The token appearance would be followed by a "ding" sound. The character would start gesturing towards a table, on top of which users would see three dishes containing French fries, mussels and croquettes. Upon pronouncing the words and collecting a second token, the user would go to a gaming area consisting of a dart board and four darts. Upon being prompted to throw the darts at the board by Character 3, the user would see appearing 2D pictures of four Antwerp landmarks after each hit. In the last part of the experience, the user would return to the entrance to meet Character 1 summing up the number of tokens gained in the experience and encouraging users to visit the bar in Antwerp to redeem their discounts.

A full version of the characters' script was provided:

Character 1: "Welcome to the FlemishTerp bar! To get your discount, you will need to perform language tasks. If you are successful, you will collect tokens in the shape of Antwerp hands. Follow the instructions that you will be given starting from the bar!"

Character 2: "Come over here! - In Flemish, we call this 'bier'. If you repeat it after me, you will get your first Antwerp hand which will be your first discount. Ok, let's start: 'bier'... 'bier'."

Character 3: "Sorry, I did not hear you well, let's pronounce it together 'bier... 'bier'. Cool, you'll manage anyway. Here is your first token. Come with me and let's have something to eat!"

Character 3: "Now let's move on to our dining area. We have unique local dishes here in Antwerp, try to repeat them after me:

- *"Frieten" ... "Frieten"*
- *"Mosselen" ... "Mosselen"*
- *"Kroketten" ... "Kroketten"*

Great job! here is your second hand token!"

Character 3: "Ok, let's move to the next area and play darts! Grab your dart and try to hit one of the Antwerp

Figure 3 - Layout of "The FlemishTerp Bar" on the game engine Unity.

locations on the board. Once you hit a location successfully, a picture of it will pop up with its name in Flemish. Then, you will get your hand token! You can play the game more than once."

- *Brabo fontein*
- *Onze-Lieve-Vrouwkathedraal*
- *Havengebouw*
- *Treinstation*

Well done! here is your third hand token!"
Character 1: "Congratulations! You have now collected 3 Antwerp hands. You can claim them once you visit us at the FlemishTerp! You can also collect additional discounts by spreading the word about us! See you next time!"

2.5 Post-design survey

After the product design was completed, usability testing was conducted on 8 participants who tried the iVR experience. Participants performed their experiential evaluation on a voluntary basis and belonged to a population of undergraduate and graduate students attending summer courses at the AP University of Applied Sciences and Arts. This small number of participants resulted from the experimental nature of the study and the limited time availability in the recruitment process. They were all experienced iVR users but had received none or very little exposure to Flemish. They used a Meta Quest 2 headset connected to a computer monitor via Airlink, so that researchers could see participants' activities on the iVR platform. Once the experience ended, participants provided written comments on their experience by answering the question: "Which feedback would you give to the creators of this experience in terms of its design, effectiveness of pronunciation training and future developments?". Answers were provided in Table 2 and listed according to feedback type.

2.6 Design challenges

Researchers initially debated on whether opting for immersing users in a 360° environment or in iVR. In fact, the former would have enabled users to increase cultural exposure while limiting environmental interactivity. Moreover, capturing 360° videos and photos of real environments would have required permissions to reproduce locations and customers in the virtual space. Therefore, researchers opted for iVR due to its customizable interactive affordances. Another challenge encountered by the researchers consisted in creating an iVR experience based on linguistic aspects and bars of the city of Antwerp as none of the researchers was Flemish or had been in the city long enough to know its pubs. Therefore, researchers sought the advice of local students on cultural aspects Antwerp and its pubs. In this way, researchers gathered information on local dishes, popular pubs and student societies. Not only this helped with iVR spatial planning but also contributed to the realism of the virtual environment, such as coat of arms of student societies and university logos hanging on bar walls. Researchers were challenged to develop a shared imagery of the virtual environment where the activities would be set. They also struggled to design pedagogically effective activities for pronunciation training. In fact, only one of the researchers had previous experience of language learning in iVR. This required researchers to constantly negotiate and clarify planning operations. The group was also faced with challenges related to the types of actions allowed in the virtual space, which was overcome by refining motion capture and simplify agents' script.

Audio design constituted an additional challenge in the design process as the researchers struggled to balance background chatter with timed language prompts from virtual agents. Furthermore, agents' speech was edited as a single string of sounds incorporating pauses to allow for user production, which implied accurate audio editing though the software Audacity. The resulting audio files were then inserted into Unity as positional cubes activated upon user proximity as users moved within the iVR space. To enable audio production, virtual agents had to be set as audio sources, which consequentially restricted aerial availability to hear pronunciation prompts. Through C-sharp scripting, four additional audio sources were placed in the environment to amplify virtual agents' sounds. However, delays were reported between virtual agents' sounds and body movements as there was a mismatch between their gestures and speech content. This required further adjustments in virtual agents' movements and sound in order not to affect users' sense of immersion.

3. Discussion

The methodology outlined in this study has highlighted design practices for creating pronunciation training experiences with iVR. In fact, by adopting a participatory design methodology, researchers crafted an iVR experience by accurately designing environmental affordances and planning pedagogically-informed language learning activities. In particular, the methodology outlined that environmental credibility, sound quality and task challenge successfully determined the pedagogic effectiveness of the iVR experience for pronunciation training purposes. These findings were consistent with the results of a user trial, which highlighted that pre-reality exposure to socially-relevant spaces can help international students prepare to live in foreign social contexts. In fact, iVR design and its interactional affordances were praised by the participants who tested the prototype functionalities and appreciated the real-life accuracy of its environmental features, stating that they contributed to enhance sense of presence (Table 1; b, c, d). Positive feedback was also given to awarding pronunciation attempts with redeemable bar discounts (Table 2, a, d). Satisfactory comments were also provided on story flow as it facilitated spatial movement and prompted users to interact with virtual agents and the digital environment (Table 2, c, d). On the flip side, some participants were less positive about the richness of environmental details which distracted them from achieving activity goals. Motion sickness and viewpoint alteration were also cited as limiting experiential factors (Table 2; f). Others regretted not noticing when earning hand tokens and said they would have preferred to experience enhanced visual effects connected to their gains (Table 2, g). However, since positive comments were given to the experience outweigh the negatives, confirming the effectiveness of a participatory design methodology for creating an iVR environment supporting pronunciation training. In fact, participants highlighted that the culturally-rich immersive atmosphere of the experience fostered their motivation, engagement, curiosity and interest in the Flemish language, with benefits for their pronunciation skills (Table 2, i, l, m).

<table>
<tr><th>Design</th><th>Language learning experience</th></tr>
<tr><td rowspan="3">a. “From my experience, the storytelling, the 3D model environment settings are very elegant in this game, which really gave me the feeling of walking into a local pub. The idea of creating some game rounds to get discount was also brilliant, which made me feel immersed in this process and excited to see what would happen next.”
b. “I have no remarks about the design of the VR environment as it was pretty well depicted and cozy enough for its purpose. I have to highlight the sound work, which was magnificent regarding 3D forms and binaural features. It immensely helped the orientation in the VR space.”
c. “During the experience the first thing that caught my attention was a nicely made environment with well-devised interactions and avatars. I didn’t have any issues with understanding the experience and the order of the actions to perform. The idea itself I really liked, mostly the fact that the place was inspired by a real bar that users could visit during their stay in Antwerp.”
d. “Notion of space was interesting, with perceivable scene references to the Dutch culture. The atmosphere facilitated my sense of presence and spatial movements. Sound cues were seamlessly distributed to enable the exploration of the environment. Directional cues were easy to follow and I was engaged in trying to pronounce words. Since I was testing a low-fidelity prototype my expectations were not high, but the story flow worked to maintain me engaged until the end of the experience. I think this engagement and sense of presence associated with the actual contact with words in another language really support the language learning process.”
e. “I did admire the whole experience, and I was honestly astonished at what a team of experts from different backgrounds who have never met before can achieve in only a couple of days. And I am sincerely curious about what further development of this VR experience could be.”
f. “The only issue I had was feeling motion sickness during movement in the environment that disturbed the experience itself. My avatar was also placed in a lower position than it should have, which made me feel that ⅓ of my body was below the floor. It caused me to miss the intended purpose.”
g. “At some point I felt disconnected from the first activity because I could not notice what I earned in it. I was also expecting some visual effects expressing my success in the activity.”
h. “I regret that the challenge of playing darts in VR was a bit clumsy, because none of my darts actually reached the target.”</td><td>i. “After experiencing the Flemish bar in VR, I felt it was a very productive and interactive tool to learn a language. I was able to learn basic words that were taught to me there. It also came with the benefit of collecting rewards that can be utilised in a real bar setting. As a prototype, the experience was well made.”
l. “I had a nice and interesting experience visiting the Flemish bar, I learned how to pronounce one of the Flemish signature dishes and later on I played a game of darts. An immersive experience!”
m. “With “The Flemishterp Bar”, I experienced a unique, engaging as well as impactful way to learn Flemish. This immersive experience demonstrated that using virtual reality for educational purposes is efficient, and very promising!”</td></tr>
<tr><th>Suggestions for prototype development</th></tr>
<tr><td>n. “I can imagine with speech recognition technology, this iVR experience can do wonders.”
o. “Maybe the iVR prototype could be a good marketing experience for other places? If so, I believe that this VR experience could be used as a marketing tool for places wanting to be more recognizable for tourists and even local people, and not only as a way to learn a new language!”
p. “I think the graphical part of the language could also be explored further. Since I was saying a new word it would help me if I could see the written word and maybe the syllabic structure of it.”</td></tr>
</table>

Table 2 - Participants’ responses after iVR prototype testing.

Looking at potential future developments of this iVR experience, participants suggested increasing the number of visitable locations in order to promote local businesses and cultural amenities (Table 2; n, o). This entails that iVR experiences of this kind could enhance the quality of international students' academic and social lives, serving both as a learning and a promotional tool for local bars and cultural amenities. Furthermore, students hinted at further developments in speech recognition technology so that users could receive immediate feedback on their performance and improve language parsing (Table 2; n, p). This confirms the necessity to expand current research on pronunciation training in iVR by fostering real-time pronunciation feedback from agents and environmental affordances.

4. Limitations

Significant limitations of this study consisted in a lack of available technologies needed to incorporate iVR-based speech recognition and pronunciation training. Although this was partly compensated by timing virtual agents' responses to potential real-user feedback and environmental interactions, it highlighted that future investigations in iVR-based pronunciation training should consider a wide range of speech-recognition technologies for integration assessment. It is believed that joint efforts between instructional designers and technologists may help to develop interactionally effective language applications for speech recognition. Moreover, out of the 4 researchers who designed this experience, only one was an expert in language pedagogy, while just another had enough knowledge of Unity to design iVR components. A bigger research team with at least two highly skilled members per discipline might have widened the range of available iVR environments, added higher interactional and linguistic complexity to the experience once planified. The number of evaluators would have need to be expanded. As part of the evaluation process, these people would be requested to express their judgement in terms of acceptance, sense of presence, usability and comfort, using reference systems such as the Technology Acceptance Model of Davis (1989), Brooke (1995) and of Witmer et al. (1998). Lastly, time constraints significantly limited the team's possibilities to refine interactional and environmental details, which could have been avoided with a wider time availability to complete the experience. Further application development needs to be conducted within premises granting the availability of cutting-edge iVR tools and computers enabling effective design and testing.

5. Conclusions

This study has attempted to outline the design practices needed by language teachers to create an iVR prototype enhancing the pronunciation skills of students interested in pursuing their studies in Antwerp. To meet this goal, the iVR experience was designed to immerse users into social and cultural aspects of the city of Antwerp while learning to pronounce basic Flemish words. By using Unity, 3D designs and audio files, iVR realism was enhanced through cultural aspects of local bars in Antwerp, a story line depicting possible social scenarios and authentic examples of Flemish vocabulary related to food, drinks and cultural landmarks. Although the technology to record and provide real-time feedback on user pronunciation was not available at the time of this study, the level of interactivity provided by the iVR experience successfully encouraged users to repeat words said by virtual agents with consequential praises for environmental immersion. In fact, results from a user testing demonstrated that virtual environmental affordances, interactional realism and situated practice successfully fostered pronunciation training in a foreign language through enhanced immersion and action stimulation. However, the study also highlighted that further investigations are needed to develop suitable technologies for real-time speech recognition in iVR required to improve students' pronunciation skills. Considerations of this kind would need to include virtual agents' appraisal of users' non-native accents and further pedagogic planning of language activities, as well as the deployment of AI. As this study was specifically centered on Flemish, voice recordings of native Antwerp people would increase the real-likeness and pedagogical impact of spoken feedback in the iVR environment.

To conclude, the creation of this language learning experience has evidenced how iVR can foster students' interest in learning the local language via iVR involvement. Due to this efficacy, it is believed that designing iVR language learning experiences should be integrated in teachers' training background. In this way, language teachers could be part of a community of instructional designers with diversified skills deployed to create iVR-based language learning experiences. This may enable teachers to customise iVR activities for curriculum development and encourage partnerships between universities and VR-equipped laboratories, which are necessary practices in light of the progressive incorporation of iVR technology in language education.

Acknowledgements

Special thanks are given to the AP University of Applied Sciences and Arts in Antwerp for organizing the summer school in "Storytelling with Virtual Reality: An Immersive Encounter" and for providing researchers with the facilities to create the iVR language learning experience. Gratitude is expressed to the researchers of the Immersive Lab and to fellow group members who worked on this project.

References

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., & Pidruzny, J. N. (2009), The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. Journal of Experimental Social Psychology, 45(4), 624-634.

Brooke, J. (1995). SUS: A quick and dirty usability scale. In: P.W. Jordan, B. Thomas, I.L. McClelland, B. Weerdmeester (ed.), *Usability Evaluation in Industry*. London: CRC Press.

Calvert, J., & Abadia, R. (2020). Impact of immersing university and high school students in educational linear narratives using virtual reality technology. Computers & Education, 159, 104005.

Chang, H. (2021). Emersive VR: An Expanded Immersive VR Practice (Master thesis, University of California Santa Cruz, Santa Cruz, CA). URL: https://escholarship.org/uc/item/0k21m2sf (accessed on 9th September 2022).

Cook, M., Lischer-Katz, Z., Hall, N., Hardesty, J., Johnson, J., McDonald, R., & Carlisle, T. (2019). Challenges and Strategies for Educational Virtual Reality, Information Technology and Libraries, 38(4), 25–48.

Eichenberg, C. (2012). Virtual reality in psychological, medical and pedagogical applications, InTech, New York, NY.

European Commission, (2022). Digital education initiatives, URL: https://guides.library.uq.edu.au/referencing/apa6/webpage (accessed on 7th September 2022)

Holly, M., Pirker, J., Resch, S., Brettschuh, S., & Guetl, C. (2021). Designing VR Experiences -Expectations for Teaching and Learning in VR, Educational Technology & Society, 24, 107–119.

Immerse (2023). AI Conversation Practice. URL: https://www.immerse.com/ (accessed on 23rd December 2023).

Kumar Jang Bahadur, S. (2022), Solving Automatic Speech Recognition Deployment Challenges, URL: https://developer.nvidia.com/blog/solving-automatic-speech-recognition-deployment-challenges/?ncid=so-link-148320-vt46&=&linkId=100000147748589#cid=dl20_so-link_en-us (accessed on 7th September 2022)

Luthy, C. (2017), 5 ways virtual reality is flipping the script on screenwriting, URL: https://www.uncsa.edu/news/20170307-virtual-reality-screenwriting.aspx (accessed on 8th September 2022)

Mitsui, K., Hono, Y., & Sawada, K. (2023). Towards human-like spoken dialogue generation between AI agents from written dialogue, *arXiv*. https://doi.org/10.48550/arXiv.2310.01088

Moeller, D. (2017), Different types of immersion and how they work, URL: https://ispr.info/2012/06/28/different-types-of-immersion-and-how-they-work/ (accessed on 7th September 2022)

Saberi, M., Bernardet, U., & DiPaola, S. (2014), An Architecture for Personality-based, Nonverbal Behavior in Affective Virtual Humanoid Character. Procedia Computer Science, 41, 204–211.

Scovel, T. (1994), The role of culture in second language pedagogy, System, 22 (2), 205-219.

Seyama, J., & Nagayama, R. S. (2007), The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. Presence, 16(4), 337–351.

Spinuzzi, C. (2005), The Methodology of Participatory Design. Technical Communication, 52(2), 163–174.

Visch, V. T., Tan, E. S. & Molenaar, D. (2010), The emotional and cognitive effect of immersion in lm viewing, Cognition and Emotion 24(8), 1439-1445.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225–240.

WondaVR (2023), Harvard Students Build Their Own AI Game to Learn a New Language, URL: https://www.wondavr.com/case-studies/ai-mystery (accessed on 23th December 2023).

World Economic Forum (2020), Technology is changing the face of education. Here’s how, (accessed on 9h September 2022)