Pak. j. life soc. Sci. (2024), 22(2): 7367-7382 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2024-22.2.00556

RESEARCH ARTICLE

Usability Study of GenAI for English Learning in VR

Bee Sian Tan[1]*, Shen Hoi Ong[2], Khin Huat Tan, Andrew[3], Thein Lai, Wong[4], P Sharimila Bai AP Pandurenga Rao[5]

[1,2,3,4,5]Tunku Abdul Rahman University of Management and Technology, Kuala Lumpur, Malaysia

ARTICLE INFO

Received: Jul 24, 2024
Accepted: Sep 15, 2024

Keywords

Virtual Reality

Generative Artificial Intelligence

Educational Technology

English Language

Education

***Corresponding Authors:**
tanbs@tarc.edu.my

ABSTRACT

Traditional English learning environments in university often hindered by outdated focus on reading and writing, limited textbook content, insufficient speaking practice opportunities, and pre-programmed artificial intelligence (AI) in English speaking practice. This study explores the potential of leveraging VR technologies and generative AI (GenAI) to overcome these barriers. *EasyEnglish* is a real-time conversation game with GenAI non-player character (NPC)s. This game utilizes voice input recognition, large language model (LLM) for language assessment and text-to-speech (TTS) for NPC lip-sync animation. The content validity was assessed by 3 experts to evaluate the conversation quality. A usability test was conducted using a purposive sampling method with 7 undergraduate non-native English speakers who have less than 1 year experience in using VR technology. This study employed System Usability Scale (SUS) and Content Validity Index (CVI) metrics for assessment. The CVI result showed satisfactory agreement in conversation quality but highlighted areas for improvement in learning objectives. The SUS result revealed satisfaction in consistency of user interface (UI) and learnability of *EasyEnglish*, while also highlighting the need for improvement in UI, setup, and visual cues. The significance of the study lies in the GenAI's ability to provide diverse response, avoid repetitive dialogue and speak using gestures to undergraduate students. GenAI effectively identify and assess irrelevant words in conversation, provide immediate grammar and vocabulary correction and suggestion of conversation improvement accurately. Future research should focus on improving accessibility, include more multimodal interactions, and mapping learning objectives with the soft skills emphasized by the World Economic Forum (WEF).

INTRODUCTION

Proficient English communication skills are vital for global connectivity and career advancement (World Economic Forum [WEF], 2019). However, non-native undergraduate speakers often lack confidence in spoken English (Hossain, 2024; Romero et al., 2024). This lack of proficiency in creates barriers for students, educators, and administrative staff to participate in the academic process (Romero et al., 2024). The World Economic Forum (WEF) highlights the need for improved language proficiency for workplace success (WEF, 2019). Employers also priotize listening and speaking skills for workplace communication (CambridgeEnglish, 2024). Traditional English language learning methods, which primarily emphasize reading and writing skills, are outdated (Hossain, 2024). In addition, undergraduate English language learners have limited learning content on textbook (Hossain, 2024; Madani, 2021). Approaches like this unfairly burden learners, especially those with weaker proficiency in grammar and vocabulary (Chan, 2024). Consequently, undergraduates have

limited opportunities to practice and improve their spoken English abilities, leading to a lack of confident in engaging in casual conversation in the workplace and daily life (Chan, 2024).

Virtual reality (VR) technology was applied in learning environment to ensure the learners to study the English as a Second Language Learning (ESL) in the conservation in future (Hossain, 2024; Jayes et al., 2022). VR technology continued to gain momentum in 2014 because Mark Zuckerberg, the CEO of META, which was formerly called Facebook, spent $2 billion buying the Oculus VR company and subsequently introduced the Metaverse (Kraus et al., 2022; Saleekongchai et al., 2024). This action prompted many companies to explore VR possibilities, further enhancing VR technology. Hence, VR is described as a new dimension of participation, bringing learners to synthetic environments. It is an invaluable tool with potential benefits for language learning, offering ample opportunities to improve skills at anywhere and anytime (Huang et al., 2023). Former study found VR has the potential to develop language learners' cognitive abilities in terms of improve memory retention, reduce cognitive load and critical thinking. However, the learning process often hindered by technical issues, slowing down the speed of completing task and low accuracy in terms of feedback. Most of the past study employed low immersion VR, which is by watching 360-degree video to learn language. The research also highlights the need of content customization and interaction needed for immersive learning experience. In addition to VR, generative artificial intelligence (GenAI) is known as one of the top ten emerging technologies. which can generate the content based on user input, offer personalized learning experiences (Jurgens, 2023; Nasir et al., 2024). However, integration of GenAI with VR for language learning requires further exploitation to address the existing challenges.

Therefore, the vision of the study is to assess the effectiveness of VR and Generative AI (GenAI) technologies in enhancing the English language learning outcomes. The vision is broken down into two research objectives to be achieved in this study as following

1) To assess the content validity of VR GenAI application by gathering expert feedback on the content.
2) To evaluate the usability of the integrated VR and GenAI to understand how effectively it enhances language learning.

LITERATURE REVIEW

Virtual Reality (VR)

VR technology is rapidly advancing in fields like education, travel, and interviews (Philippe et al., 2020). It offers dynamic virtual environments using head-mounted display (HMD) and controller, which employ sensors to mirror avatar actions with visual, auditory, and haptic feedback, thereby enhancing spatial memory and learning outcomes (Chen & Chen, 2022). VR has been shown to simulate real-life examples which can help them to overcome issues of low self-confidence in the process of learning language (Huang et al., 2023). Moreover, it motivates learners to practice English speaking and listening in various VR scenarios and with non-player characters (NPCs) without the need to travel abroad (Huang et al., 2023). However, challenges remain, particularly regarding the setup time of VR devices and students' familiarity with the technology (Holly et al., 2021). This study aims to address this gap by designing user friendly VR features for interaction when learning language (Holly et al., 2021). Furthermore, existing virtual reality assistant language learning (VRALL) possess limitation in terms of pre-programmed NPC, which limit the variety of content to practice English speaking (Chen et al., 2022). By focusing this novel aspect, the study will contribute to the existing body of knowledge on improving language learning through personalized learning experiences VR.

Generative Artificial Intelligence (GenAI)

GenAI creates new content based on user prompts in natural language, leveraging advanced natural language processing (NLP) and large language models (LLMs) (Lv, 2023, UNESCO, 2023). Despite its benefits, concerns about GenAI's drawbacks include job displacement, particularly in tasks like conversation, writing software code, and promotion (Dwivedi et al., 2023). Ethical debates also arise over its dual-use potential, as it can be employed for both beneficial and harmful purposes without discerning intent, which could lead to political and social issues (Coeckelbergh & Gunkel, 2023). Additionally, there's anxiety over creativity and copyright, as GenAI might generate content from existing works, potentially infringing on the rights of original authors and artists (Farina & Lavazza, 2023).

ChatGPT, a notable GenAI tool, was first launched by OpenAI in 2018, simulating human-like text conversations through deep learning (Ollivier et al., 2023). The technology evolved with the release of GPT-3 in 2020, and later, the enhanced GPT-3.5, designed specifically for conversational AI (Dwivedi et al., 2023). ChatGPT employs Whisper API for speech-to-text (STT) functions, using automatic speech recognition (ASR) algorithms to convert voice into text (Alharbi et al., 2021). To ensure security, HTTPS web requests encrypt audio data during transmission, maintaining privacy and integrity (Huang et al., 2022). The latest version of ChatGPT is noted for its improved accuracy in executing prompts.

GenAI may introduce biased if the data trained consists of biased (Huang et al., 2022). Researchers and educators are advised to filter sensitive information using keywords to mitigate this risk (Huang et al., 2022). Second ethical consideration is who is responsible for the content generation and the potential to be used in destructive ways (WEF, 2023). For instance, the user may spread misinformation and deepfake videos (WEF,2023). Therefore, a proper copyright that should be addressed if the content is generated by GenAI (WEF2023).

English Language Education

Language learning requires structured guidance and disciplines, with motivation often fading due to tedious learning materials and process (Schorr et al., 2024). English is the most common foreign language learned by non-native speakers, focusing on vocabulary mastery (Schorr et al., 2024). Previous study showed various educational technologies applied in language education (Huang et al., 2023). The common challenges applying technology in language learning includes differing language levels to achieve the similar learning outcomes (Huang et al., 2023). Most current language learning applications were predefined and lacking reliable speech production (Huang et al., 2023). Chatbots are common applications in HE for guidance (Romero et al., 2023). However, the previous chatbot content is predefined and display in a text form, limiting personalized experience of learners (Romero et al., 2023). his study explores the usability of ChatGPT's speech-to-text (STT) features for conversational learning. By focusing on this innovative aspect, the research aims to provide insights for designing user experiences and interfaces (UX/UI) that facilitate language learning through speech.

GenAI in English Language Education

AI has revolutionized language education through automated grading, tutoring services, speech and pronunciation training, and personalized learning experiences. (Huang et al., 2023). GenAI adjusts content difficulty based on learners' input, has shown to improved learning outcomes compare to traditional learning method (Huang et al., 2023; Yu & Guo, 2023). However, uncertainties remain regarding the accuracy and reliability of GenAI responses, particularly when compared to human responses (Rosario & Noever, 2023). Educators have shown reluctance to adopt GenAI due to a lack of technology experience and doubts about its reliability (Kohnke et al., 2024). This is shown where

feedback generated by AI tools may be biased, not logical and not accurate (Huang et al., 2023; UNESCO, 2023). To address these issues, prompt need to be clear and specific (Huang & Chen, 2023). Additionally, the performance of automatic speech recognition (ASR) and natural language processing (NLP) should be tested before implementation to account for variability in user input (Huang et al., 2023). Despite its potential, GenAI currently falls short in delivering personalized feedback to learners. (Romero et al., 2024). This study aims to mitigate these issues by investigating the right prompt instructions for designing conversation contexts and automated error checking, aiming to increase educators' confidence in using VR for language teaching.

METHODOLOGY

The Analysis, Design, Development, Implementation and Evaluation (ADDIE) model is applied in this study to design the educational content to align with the virtual activities (Yu et al., 2021). The steps involve analyse and plan, design, test, build, review, and launch. The analyse and plan stage involves the discussion of learning content with English Language experts. This is to determine the VR English learning objective based on the level of English proficiency among undergraduate students. The user and scenario within VR environment is determined as well. In the design stage, the VR application called *EasyEnglish* is created using the Unity game engine. The ChatGPT prompt is designed by setting up background story for the NPC. The storyboard of *EasyEnglish* is sketched to visualize the entire game flow. Upon completion of implementation, the content of *EasyEnglish* is then validated by content experts to ensure accuracy and effectiveness of *EasyEnglish*. A usability test is conducted using system usability scale (SUS) test to evaluate the usability of integration between VR and ChatGPT. This research procedure was approved by the university ethics community. Participants were fully informed about the purpose of the study and they have right to withdraw anytime. Informed consent was obtained before the tests were conducted. Their identities were anonymized to ensure confidentiality and privacy. They were assigned to use wear *MetaQuest* headset and controller, explored the VR environment between 10 to 15 minutes to avoid nausea and discomfort (Meta, 2024).

EasyEnglish VR Game Features

The *EasyEnglish* VR game offers learners an exciting opportunity to select their own characters and role-play in a dynamic game scenario. For instance, in this challenging scenario, an NPC wants to invite friends to a birthday party, and it is the player's mission to communicate effectively with various guest NPCs at the party using the English language. Each NPC has their own unique personality, conversation style, and interests, so learners must always remain sharp and attentive. The game presents numerous challenging missions, such as giving gifts and fulfilling requirements, which require learners to demonstrate exceptional communication skills. The game features a wide array of interactive items, including gifts, food, and drinks, which require learners to master specific sets of vocabulary to interact effectively with NPCs. With this game, learners can significantly enhance their English language skills while enjoying an immersive and thrilling gameplay experience.

Technical Implementation

Architecture Design of EasyEnglish

The *EasyEnglish* system, illustrated in Figure 1, integrates several NLP tasks to enhance language learning. Learners interact with ChatGPT NPCs, virtual conversational partners equipped to assess grammar and vocabulary usage. Through the Whisper 4.0 API, spoken input is seamlessly converted to text, facilitating interaction with NPCs and the VR environment. Text responses from NPCs undergo TTS conversion, enabling spoken feedback synchronized with lip movements via Oculus Lip-sync technology. Additionally, ChatGPT evaluates learners' grammar and vocabulary usage, providing personalized feedback through the Assessment Menu.

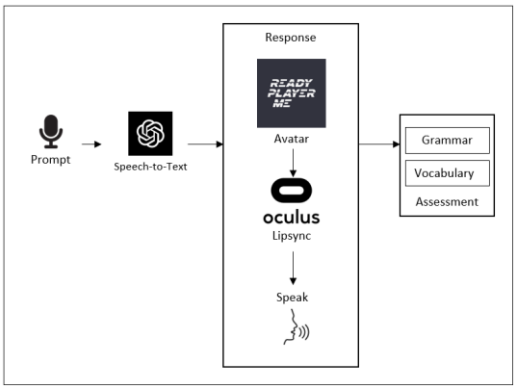

***Figure 1*. The software architecture of VR English learning application, *EasyEnglish*.**

Design and Implementation of User Prompt Features

Designing user prompts begins with a scenario background, NPC characteristics, and conversation instructions (Figure 2). To avoid lengthy responses, NPC replies have a word count limit. The prompt also includes grammar error detection and suggestion features, distinguishing between grammar checks and reply instructions. To maintain immersion, ChatGPT is programmed not to identify itself during conversations. ChatGPT then trims incoming messages and generates responses. In the game, ChatGPT's replies are shown in the dialog scene. If a grammar error is detected, a warning panel pops up for immediate learner review. The main menu features a conversation history section, allowing learners to review past interactions with NPCs.

Speech-to-Text (STT) Process

Integrating Whisper API into this VR English conversation game has transformed speech-to-text transcription, providing an immersive experience for learners. Implemented using the whisper-1 model in Unity, Whisper processes speech with real-time accuracy, recognizing multilingual inputs and translating them into English efficiently. It captures low-volume speech precisely, allowing natural interactions with NPCs in the virtual environment. The script scans available microphones, presenting them in a dropdown menu for learners to choose from. Recording duration is set to 5 seconds, with a progress bar indicating remaining time. Once recording is complete, Whisper API transcribes the voice to text, which is then sent to ChatGPT to generate a response during the conversation.

Text-to-Speech (TTS) Process

To enhance immersion, OpenAI's TTS model has been integrated to convert text into natural-sounding speech. This AI model captures nuances, inflections, and emotions, giving NPCs lifelike qualities. The TTS model enables diverse, engaging personas with distinct voices. Two models are used: TTS-1 for speed and TTS-1-HD for quality. In this project, the TTS-1 model implements the speech endpoint in the Audio API, offering six built-in voices for different NPCs. When ChatGPT generates a reply, the text is sent to the TTS Manager for sound synthesis. The TTS Manager uses the selected model and voice to convert text to audio, which is then played through the AudioSource API in Unity.

Instruction: I want you to chat with an English learner later. Assume that you are in the birthday party scenario and act as Billy, the host of this birthday party to celebrate your birthday. You also role play as the learner's best friend since you met in primary school. Since you are a quiet and socially awkward person, you do not invite many people to your party. In this invitation, you mainly want to meet the learner because you have not chat face to face for a long time since graduating from the university. I will signal you when to start the conversation with the learner by typing "start" I will signal you to end the conversation with the learner by typing "end". Do not create more than 15 words as possible for one sentence in the conversation. If you detect any grammar error from the learner's sentence, please mention the correct sentence and provide explanation of the grammar error by using the following format.

Conversation format:

<Grammar Error Detection!!> The correct sentence is <Correct full sentence>. This is because <Explanation>

<Your reply sentence>

Figure 2. **The conversation prompt design for NPC using ChatGPT.**

NPCs Integration with Dynamic Facial Expression

The NPC is created from the combination of the avatar from *Ready Player Me*. Before loading the avatar from Ready Player Me, it needs to create its own avatar configuration to determine the morph target of the avatar such as Oculus Visemes, mouth smile, eye blink left and eye blink right. This avatar configuration will be placed in the *Ready Player Me* avatar load menu and the avatar designed in the *Ready Player Me* website can be imported now. The avatar with the avatar configuration will provide a render avatar which will determine the blend shapes for the setting of the morph target. After putting the render avatar into the skinned mesh render section of *Oculus Lip Sync*, the face animation will work normally and increase the realism of NPCs.

NPCs Facial Realism Enhancement

To enhance realism of NPC, the facial animation was implemented to enable the mouth to move like a human's. The Oculus Lip Sync is integrated into this project which can act as an add-on plugin to sync avatar lip movements of an NPC to speech sounds from pre-recorded audio or live microphone input in real-time. This Lip Sync analyses the audio input stream from an audio file and predicts a set of values called visemes such as SIL, PP, FF, TH, DD, and KK which represent with its own phonemes to convert TTS. This can help to generate the gestures or expressions of the lips and face according to the speech sound. Finally, these visemes can animate avatars' mouths so that they look like they are speaking, especially when combined with some postures defined.

Language Assessment and Feedback

In the process of language learning, learners often encounter challenges with grammar and vocabulary usage, leading to the needs of error correction for better understanding and improvement. Unlike the conventional methods where learners check grammar manually, *EasyEnglish* automate the process by leveraging ChatGPT's capabilities within the VR platform.

Grammar Correction Assistance in *EasyEnglish*

During the conservation, the learners may use wrong grammar in their sentences. Therefore, the correction needs to be provided so that the learners can understand the mistake that they made, and how to correct it. Normally, if browsing the ChatGPT website, Learners can check the grammar of sentences by utilizing the prompt "check grammar". In *EasyEnglish,* the grammar checking function is implemented in Unity software. Therefore, there is a need for importing OpenAI API packages in Unity. After the implementing, the standard HTTP libraries in C# is used to make the REST API calls. Therefore, ChatGPT is applied to do the conversation and it can help to check the grammar errors that may have occurred in the learners' conversation as shown in Figure 3. For example, if the learners say "I happy about it." with the NPC, ChatGPT will detect the error and show the correct version below the learners' sentence. An explanation of the corrected version will be provided after the learners have completed their sentence. In this case, ChatGPT will suggest changing the words to "I am also happy about it" so that the learners can understand grammatical error in the sentence used.

***Figure 3*. Grammar Error Detection when Chatting with NPC.**

Vocabulary Learning Through Interactive Object

In the virtual environment, multiple objects are labelled, so that learners can improve on their vocabulary. Whenever players interact with any of these objects by picking them up, the pronunciation of the object will be played in audio. (Figure 4). Through this learning method, learners can expand their vocabulary of nouns which are relevant to the virtual scenario, and hence improve their conversational ability.

Content Validation Process

The content validation was conducted to ensure the game's elements, mechanics and features align with the intended goals and the experience in the whole game process. This form can also be used to gather feedback from the experts to ensure the game content pertains to the consistency and coherence of the original design. The content validation was held in a physical meeting by presenting *EasyEnglish* prototype directly. After presenting the application, the experts filled up their rating and feedback on the content validity form which contains six content aspects based on their expert judgment. A 4-point Likert scale ranging from "strongly disagree" to "strongly agree" was used to rate this content.

During this content validation process, 3 different identity experts were selected to take part in the rating and comments. The first expert is a senior lecturer and game technology expert. He has a rich experience in VR and AR related knowledge. The second expert is an experienced lecturer whose expertise in Linguistics study. Her rich experience in English Studies plays an important part in validating the learning content. Finally, the last expert is a research leader and assistant professor. His expertise in the field of research can contribute to the overall quality of the research project. Table 1 displays the questions for the content validation.

Figure 4. **Audio sound of vocabulary is played once player picked up the items.**

Table 1Content *Validation Form for EasyEnglish*

	Content Aspects	Ratings			
		1	2	3	4
1	Grammar Accuracy: The content demonstrates proper grammar usage and structure.				
2	Vocabulary Accuracy: The choice of vocabulary is appropriate and relevant.				
3	Coherence of Conversation: The conversation flows logically and coherently.				
4	Relevance to Learning objectives: The content aligns with the intended learning outcomes.				
5	Appropriateness for Target Audience: The content is suitable and engaging for the specified audience.				
6	Overall Quality and Effectiveness: Overall, the meets high standards of quality and effectiveness for language learning.				

Usability Testing

The usability of the *EasyEnglish* is divided into two primary stages. These stages include conduct survey using questionnaire and interview procedure. *EasyEnglish* VR game's usability was assessed using the System Usability Scale (SUS) questionnaire developed by Brooke (1996) (Vlachogianni & Tselios, 2022). This collects valuable participant insights directly, aiding developers in enhancing game efficiency and user satisfaction (Vlachogianni & Tselios, 2022). Upon the completion of usability testing, the participants were interviewed to gather open-ended responses, concluding the evaluation process.

According to usability study guideline from Nielson, 5 sample sizes are sufficient to identify mistakes found in a complete systems as larger sample sizes do not reveal more problem but time consuming (Nielson, 2000). In this study, 7 undergraduate students aged 18-25 who has less than 1 year experience in using VR devices were randomly selected. They are required to wear MetaQuest2 VR headsets for testing. After completed the testing, participants complete a Google Form with a ten-item SUS questionnaire, indicating usability aspects on a 5-point Likert scale (Vlachogianni & Tselios, 2022). These questions address satisfaction, helpfulness, and participant ideas, serving as a high-validity measure. All participants are undergraduates from various disciplines, only two have prior

VR experience. Lack of VR familiarity complicates testing but most participants successfully complete it. The collected data was calculated with the formula as shown in Figure 5 where 5 points will be deducted for questions with odd numbers and 25 points deducted for questions with even numbers.

X = total points for all odd-numbered questions -5

Y = 25 – total points for all even-numbered question

SUS Score = (X=Y) *2.5

Total SUS Score for the project = Total SUS Score / participant involved.

Figure 5. Formula for Calculation of SUS score

RESULTS

After the testing process, the result for the Content Validation Form and the System Usability Scale (SUS) Form had been collected from different experts and participants' feedback. The result from the data collected can be used to determine this project's satisfaction and perfect scale.

Content Aspects and Ratings

Agreements are calculated by adding the relevant ratings for each item provided by all experts. Ratings of 3 or above are considered as 1 (Yusoff, 2019). For example, experts in agreement for item 1 (1+1+1) = 3 (Yusoff, 2019). The item-level content validity index (I-CVI) is then calculated by dividing the number of agreed items by the number of experts (Yusoff, 2019). For instance, the I-CVI of item 1 is 1 because 3 divided by 3 experts equals 1, while item 4's I-CVI is 0.6667 as 2 divided by 3 experts yields 0.6667. Next, the Scale-Level Content Validity Index (S-CVI/Average) is determined by averaging the I-CVI scores across all items using (sum of I-CVI)/ (number of items). Here, S-CVI/Average is 0.8889 [(1+1+1+0.6667+0.6667+1)/6]. The Scale-Level Content Validity Index (S-CVI/Universal Agreement) is then calculated as (total agreement)/(number of items). With a total agreement of 4, divided by 6 items, the final score is 0.6667.

Although the ideally acceptable S-CVI/Universal Agreement value is 1 based on the number of experts, calculations show that I-CVI, S-CVI/Ave, and S-CVI/UA have met satisfactory levels. However, the questionnaire's scale has an unsatisfactory level of content validity for items 4 and 5, indicating disagreement on relevance to learning objectives and appropriateness for the target audience. Hence, the I-CVI falls short of a full mark of 1 for each item, requiring improvement and further practice for setting learning objectives and revise the learning content for undergraduate students.

SUS RESULTS

Overall SUS Score

The System Usability Score (SUS) result is calculated based on the formula outlined in Figure 6, utilizing data obtained from the evaluation process. This VR application project received a score of 50.71, derived from the summation of all SUS Scores calculated. This score suggests that the application passed the usability test, with an average usability rating. A standard good usability score is typically 68, placing this application in the "OKAY" rating category (Shari et al., 2021). This indicates room for improvement within the VR application. Overall, the score of 50.71 suggests that learners found the application moderately usable, but with notable areas requiring enhancement. Consequently, some features of the application may cause confusion or dissatisfaction among learners, necessitating adjustments or enhancements to improve the overall user experience.

Total SUS Score for this project = (42.5 + 27.5 + 57.5 + 77.5 + 47.5 + 52.5 + 50)

= 355 / 7

= 50.71

Figure 6. **Calculation of total SUS score for Easy English application.**

Individual Item Scores

Figure 7 displays the individual score for each question. Question 1, 3, 5, 7 has the highest score of 3. Meanwhile, question 9 "I felt confident using this system" has the lowest score of 2. The results show feelings of neutrality among participants when it came to using this VR frequently. Question 2 has the highest score of 4, indicating that the participants agreed that the VR application is complex to be used. The result suggests that they may be some features of the systems are challenging for participants to access and navigate. This complexity could potentially contribute to the lower confidence level as shown in Question 9. Meanwhile Question 6, 8 and 10 has the lowest score of 2, indicating that participants disagree that there is inconsistency in the VR application and disagree that they need to learn a lot things before using this system. This result suggest that they felt that the learning process for using this application was relatively manageable.

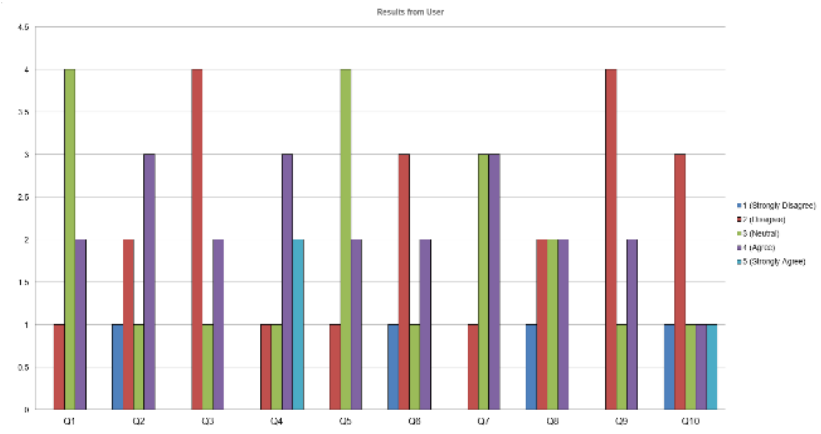

Figure 7. **Individual score for each item**

Qualitative Insights

Upon the completion of SUS questionnaire, open-ended comments are received during the discussion to gain deeper insights into user experiences and perceptions. The participants' feedback is gathered with the following themes

a) UI Interaction Enhancement

The feedback regarding frustration with direct UI interaction aligns with the usability result indicating that participants found the VR application to be complex as shown in question 2. Participants noted frustration with the direct UI interaction, where they found it challenging in pressing the button precisely using this interaction method. This also contributes to the confidence of using the system (Question 9). To address this issue, VR ray interactor is suggested to improve user experience by facilitating easier button pressing compared to VR direct input. Furthermore, user

feedback that continuous input causes motion sickness after exploring the application for a certain time. Therefore, suggest to use teleportation for navigation within the scene.

b) Event Feedback and Completion

Majority of the participants found the game interesting but highlighted issues with the setup due to lack of experiences in VR, contributes to the low score for question 9. In addition, they felt that there is a lack of visual cues upon the game completion, indicating potential issues with the overall experiences (Question 1, 3, 5, 7). This absence of visual cues gave them a feeling that all missions had not been completed even though they had already been done. In this case, some improvement needs to be added such as importing the particle effect, victory sound and warning text to give a sense of progression.

DISCUSSIONS

Achievement

Immersive Conversation Experience

This application provides an immersive experience through VR technology to enhance user engagement and enjoyment in learning grammar and vocabulary. The design of GenAI NPC in *EasyEnglish* offers a diverse range of responses to player's prompt, avoids repetitive dialogue, and provides a sense of realism to the conversation. As highlighted by experts, the conversation maintains coherence, seamlessly aligning with the game scenario, earning an average rating of 3. For instance, in cases where learners did not speak fluent English or provided irrelevant input, the NPC would apologize and inform them that it did not comprehend the meaning of the words mentioned. (Figure 8). Additionally, the NPC could assign task to the player and express gratitude with a "thank you" gesture upon completion of the task as shown in Figure 9.

Real-time feedback

This application provides real-time feedback on conversation progress, grammar, and vocabulary errors, allowing learners to promptly identify and correct mistakes, which is crucial for language learning. As depicted in Figure 9, reaching 100% on the progress label signifies that learners have completed 4 to 5 conversations with the NPC. The smooth operation of these NLP tasks ensures uninterrupted interaction. Overall, the content adheres to high standards of quality and effectiveness in English language learning, as indicated by the average rating of 3 provided by the experts.

Figure 8. **NPC was able to identify irrelevant words in the conversation.**

Figure 9. **NPC showed gesture when saying thank you to the learners for completing missions.**

Limitation

Although the Whisper 4.0 API is powerful, it has a limitation noted by experts: it automatically corrects grammar errors in the STT function in this conversation game. While the API accurately transcribes speech, its autonomous grammar correction may be limited by the complexity of language structure and contextual nuances. This presents an intriguing challenge as it auto-corrects grammar errors from learner input before sending it to ChatGPT. For instance, if learners' pronunciation is not perfect, the API may produce text that does not accurately reflect the learners' intended meaning. This confusion arises because the API selects the nearest text representation, which may not align with the learners' original intent. Consequently, this can lead to inaccuracies in the corrected or original meaning of the output, as the auto-correction process may prioritize literal grammar rules over contextual accuracy, resulting in semantically incorrect transcriptions.

FUTURE STUDY

Improvement on User Interface (UI) Accessibility

Actual learners who have completed SUS commented and provided feedback in terms of UI accessibility as mentioned in section 4.2.3. To further improve usability, it is advisable to consider users' preferences for input methods. Incorporating raycast input for menu selection, rather than direct input, aligns with users' desire for smoother interactions. Similarly, integrating teleportation for navigation, instead of continuous input, enhances user comfort and reduces motion sickness concerns. For an enhanced user experience, the design of the STT menu must prioritize clarity and simplicity, catering to the needs of learners. The panel's scale should be optimized to prevent obstruction of reply prompts and NPCs, ensuring uninterrupted engagement. Additionally, minimizing the number of visual buttons reduces redundancy and streamlines operations. By relocating these functions to VR controller buttons, efficiency is improved for recording, cancelling, and sending actions. Similarly, the accessibility of the main menu in the party scene is crucial. Presently, learners are required to traverse to a specific location, potentially hindering immersion. To address this shortcoming, the main menu should be effortlessly accessible from anywhere via a simple click of a VR controller button, improving any sense of inconvenience and enhancing overall user freedom.

Multimodal Interaction

Furthermore, there remains untapped potential in leveraging the advanced capabilities of VR within this project. Take for instance, the intriguing possibility of integrating eye tracking technology. This innovative feature allows for the precise monitoring of learners' gaze within the virtual environment, enhancing immersion by providing subtle cues about where learners are focusing. Through seamless integration with gaze tracking, this functionality can enable specialized events such as activating buttons or initiating conversations with NPCs based on eye contact. This not only facilitates more natural and intuitive interactions but also offers an exciting avenue for researching body language

cues in virtual environments, thereby fostering confidence in speaking by allowing learners to practice maintaining eye contact and observing visual cues during conversations.

Additionally, hand tracking emerges as a promising avenue for enriching user engagement. Presently, VR hand presentation relies on predefined models and animations, limiting gesture capabilities to basic actions like grabbing and poking. However, the integration of hand tracking technology opens up new possibilities for real-time gesture recognition, allowing learners' hands and fingers to be accurately represented within the virtual environment. This advancement eliminates the need for physical controllers, further enhancing the realism of interactions and providing a fertile ground for studying the nuances of non-verbal communication in virtual spaces. By enabling learners to learn and practice body language through hand gestures, this technology facilitates a deeper understanding of interpersonal communication and enhances the naturalness of NPC conversation. While undoubtedly a challenge, ongoing advancements in hand tracking technology hold the promise of delivering even greater accuracy and realism, thereby enriching the overall user experience in profound ways.

Enhancement on Conversation Flow

Speech Recording Process

The STT function has a limitation in the chat recording duration, set at 5 seconds. Learners may feel a lack of control over their input as they rush to speak within this timeframe. Some learners must wait for the countdown to finish even if they have no speech to record. Improvements are needed for more flexibility. For instance, allowing learners to start and stop recording at their discretion would enhance usability. Alternatively, recording could begin simply by approaching the NPC, eliminating the need for controller buttons, and creating a more natural interaction experience. Through the minimize usage of button and controller, the NPC may be able to handle turn-taking and interruptions naturally during the conversation.

Vocabulary Usage and Conversation Topic Integration

Concerns were raised about the learning of vocabulary words in the game's scenarios, noted by experts. Some learners feel disconnected and hindered in the learning process due to the simple usage of vocabulary terms. Future studies should introduce vocabulary relevant in developing self-efficacy and management skills acquired by World Economic Forum (WEF) such as self-efficacy, management, and engagement skills (World Economic Forum [WEF], 2024). These skills are important as it focuses on working with people (WEF, 2024). Furthermore, VR prepare a play-based, hands-on and safe environment for learning and practicing the conversation (WEF, 2024). Next, conversations should allow NPC to initiate responses autonomously, creating a more natural and realistic experience. For instance, if learners stand near the NPC without initiating a prompt, ChatGPT should initiate the conversation. Furthermore, concluding conversation topics selected by learners from the main menu in the starting scene can enhance educational value and ensure relevance to real-life scenarios, fostering an immersive learning experience.

Learning Content Personalization

Future study may include features to adjust the conversations difficulty according to individual learner’s English proficiency levels. This could be done designing prompt instruction which can generate personalized conversation. This could reduce the teaching load of educators in monitoring the students’ performance during the class.

CONCLUSION

In conclusion, this research project has successfully demonstrated the potential of immersive VR technology coupled with GenAI to enhance language learning experiences. Through the utilization of VR technology, users were provided with an engaging and realistic environment conducive to grammar and vocabulary acquisition. The GenAI NPC creates dynamic conversation, offering non-repetitive responses according to the user's prompts and maintaining coherence within the game scenario. Real-time feedback on conversation progress, grammar, and vocabulary errors proved invaluable for learners, effectively correct mistakes for the users. While the VR showcased significant achievements, several limitations and areas for improvement were identified. The integration of Whisper 4.0 API is limited by its auto-correction grammar features, which impacting the accuracy of assessment. Improvements in user interface accessibility, such as simplifying menu navigation and reducing controller functions, were recommended to enhance user experience. Additionally, the incorporation of multimodal interactions, including eye tracking and hand tracking technologies has potentially used for training attention and body language to foster confidence in conducting a conversation. Future studies should focus on enhancing conversation flow by addressing limitations in speech recording processes in VR environment and vocabulary usage distribution. Furthermore, the implementation of personalized conversation prompts based on individual learner proficiency levels could significantly enhance the efficacy of the application. Overall, this research represents a significant step towards leveraging emerging technologies to revolutionize language learning methodologies, paving the way for more immersive and effective educational experiences.

REFERENCES

Alharbi, S., et al. (2021). Automatic Speech Recognition: Systematic Literature Review. *IEEE Access,* 9, 131858-131876. https://ieeexplore.ieee.org/document/9536732/authors#authors

Brooke, J. (1996). SUS: A quick and dirty usability scale. In P.W. Jordan, B. Thomas, B. A. Weerdmeester & I. L. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 189-194). London: Taylor & Francis.

Chan, C. S. C. (2024). University graduates' transition into the workplace: how they learn to use English for work and cope with language-related challenges. *System*. https://doi.org/10.1016/j.system.2021.102530

Chen, C. C. & Chen, L. Y. (2022). Exploring effect of spatial ability and learning achievement on learning effect in VR assisted learning environment. *Educational Technology & Society,* 25(3), 74-90. https://www.jstor.org/stable/48673726

Cambridge English. (2024). Impact of Linguaskill in Malaysia. Cambridge Assessment English. Retrieved from https://www.cambridgeenglish.org/why-choose-us/impact-monitoring-and-evaluation/impact-evaluation-study-examples/impact-of-linguaskill-in-malaysia/

Chen, Y.L., Hsu, C.C., Lin, C.Y. & Hsu, H.H. (2022) Robot-Assisted Language Learning: Integrating Artificial Intelligence and Virtual Reality into English Tour Guide Practice. *Educ. Sci*. 12, 437.https://doi.org/10.3390/educsci12070437

Coeckelbergh, M., & Gunkel, D. J. (2023). ChatGPT: Deconstructing the debate and moving it forward. *AI & Society*. https://link.springer.com/article/10.1007/s00146-023-01710-4

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., & Ahuja, M. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Farina, M., & Lavazza, A. (2023). ChatGPT in society: Emerging issues. *Frontiers in Artificial Intelligence.* https://doi.org/10.3389/frai.2023.1130913

Holly, M., Pirker, J., Resch, S., Brettschuh, S., & Gütl, C. (2021). Designing VR experiences – expectations for teaching and learning in VR. *Educational Technology & Society*, 24(2), 107-119. https://www.jstor.org/stable/27004935

Hossain, K. J. (2024). Reviewing the role of culture in English language learning: challenges and opportunities for educators. https://doi.org/10.1016/j.ssaho.2023.100781

Hua, C. C. & Wang, J. (2023). Virtual reality assisted language learning: a follow up review (2018-2022). *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2023.1153642

Huang, Q.-X., Chiu, M.-Y., Chen, Y.-F., & Sun, H.-M. (2022). Attacking Websites: Detecting and Preventing HTTP Request Smuggling Attacks. *Security and Communication Networks*, 2022, Article ID 3121177. https://doi.org/10.1155/2022/3121177

Huang, X. Y., Zou, D., Cheng, G., Chen., X. L. & Xie., H. R. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*. 26(1), 112-131. https://doi.org/10.30191/ETS.202301_26(1).0009

Hwang, G. J. & Chen, N. S. (2023). Exploring the potential of generative artificial intelligence in education: application, challenges, and future research directions. *Educational Technology & Society,* 26(2). https://doi.org/10.30191/ETS.202304_26(2).0014

Jayes, J. D., Noraini Said, Wardatul Akmam Din, & Megawati Soekarno. (2022). Virtual reality in Malaysian English as a second language learning: A systematic review and implications for practice and research. *International Journal of Education Psychology and Counseling*, 7(48), 263-277. http://www.ijepc.com/PDF/IJEPC-2022-48-12-19.pdf

Kasnec, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kim, S., Shim, J., & Shim, J. (2023). A Study on the Utilization of OpenAI ChatGPT as a Second Language Learning Tool. *J Multimed Inf Syst*, 10(1), 79-88. https://doi.org/10.33851/JMIS.2023.10.1.79

Kohnke, L., Zou, D. & Moorhouse, B. L. (2024). Technostress and English language teaching in the age of generative AI. *Educational Technology & Society*, 27(2), 306-320. https://doi.org/10.30191/ETS.202404_27

Kraus, S., Kanbach, D. K., Krysta, P. M., Steinhoff, M. M., & Tomini, N. (2022). Facebook and the creation of the metaverse: Radical business model innovation or incremental transformation? *International Journal of Entrepreneurial Behavior & Research.* https://www.emerald.com/insight/content/doi/10.1108/IJEBR-12-2021-0984/full/html

Lund, B. D., Wang, T., & Mannuru, N. R. (2023). ChatGPT and a new academic reality: Artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*.

Lv, Z. H. (2023). Generative Artificial Intelligence in the Metaverse Era. *Cognitive Robotics*, 3, 208-217. https://doi.org/10.1016/j.cogr.2023.06.001

Madani, F. M. (2021). Student Perception of Traditional English Teaching Methods (CLT approach) and Comparison to Modern Methods (Using Technology), *International Journal of Education and Information Technologies* 15, 35-43. http://dx.doi.org/10.46300/9109.2021.15.5

Meta. (2024). Meta Quest 2 Health & Safety Warnings. https://www.meta.com/quest/safety-center/quest-2/

Nielson, J. (2000). *Why you only need to test with 5 users.* Nielson Norman Group. https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

Ollivier, M., Pareek, A., Dahmen, J., Kayaalp, M. E., Winkler, P. W., Hirschmann, M. T., & Karlsson, J. (2023). A deeper dive into ChatGPT: History, use and future perspectives for orthopaedic research. Knee Surgery, *Sports Traumatology, Arthroscopy,* 31, 1190–1192. https://link.springer.com/article/10.1007/s00167-023-07372-5

Philippe, S., Souchet, A. D., Lameras, P., Petridis, P., Caporal, J., Coldeboeuf, G., ... & Duzan, H. (2020). Multimodal teaching, learning and training in virtual reality: A review and case study. *Virtual Reality & Intelligent Hardware,* 2(5), 421-442. https://doi.org/10.1016/j.vrih.2020.07.008

Rokade, K. M., & Saroj, S. C. (2022). Virtual reality: History, application, and future. *International Journal For Research in Applied Science and Engineering Technology*, 10(4). https://www.ijraset.com/best-journal/virtual-reality-history-application-and-future

Romero, M., Reyes, J. & Kostakos, P. (2024). Generative artificial intelligence in higher education. *Palgrave Studies in Creativity and Culture*. https://doi.org/10.1007/978-3-031-55272-4_10

Rosario, G., & Noever, D. A. (2023). Grading Conversational Responses Of Chatbots. *ArXiV.*

Schorr, I., Plecher, D. A., Eichhorn, C. & Klinker, G. (2024). Foreign language learning using augmented reality environments: a systematic review. Frontiers in Virtual Reality. 5:1288824. https://doi.org/10.3389/frvir.2024.1288824

Shari, A. A., Ibrahim, S., Sofi, I. M., Noordin, M. R. M., Shari, A. S., & Fadzil, M. F. B. M. (2021). The Usability of Mobile Application for Interior Design via Augmented Reality. In 2021 *6th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)* (pp. 1-5). https://doi.org/10.1109/ICRAIE52900.2021.9703984

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*. https://slejournal.springeropen.com/articles/10.1186/s40561-023-00237-x

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2023). Guidance for generative AI in education and research. *UNESCO*

Vlachogianni, P., & Tselios, N. (2021). Investigating the impact of personality traits on perceived usability evaluation of e-learning platforms. *Interactive Technology and Smart Education*, 19(2), 202–221. https://doi.org/10.1108/itse-02-2021-0024

World Economic Forum. (2019). Towards a Reskilling Revolution: A Future of Jobs for All. Retrieved from https://www3.weforum.org/docs/WEF_Towards_a_Reskilling_Revolution.pdf

WEF. (2023). Artificial intelligence. Strategic intelligence briefing. World Economic Forum.

Jurgens, J. (2023). Top 10 emerging technologies of 2023 flagship report.

Vlachogianni, P & Tselios, N. (2022). Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *Journal of Research on Technology in Education,* 54 (3), 392-409. https://doi.org/10.1080/15391523.2020.1867938

World Economic Forum (2024). Ten 21st century skills every student needs. https://www.weforum.org/agenda/2016/03/21st-century-skills-future-jobs-students/

Yu, H., & Guo, Y. (2023). Generative artificial intelligence empowers educational reform: Current status, issues, and prospects. Front. Educ., 8. https://doi.org/10.3389/feduc.2023.1183162

Yu, S. H., Hsueh, Y. L., Sun J. C.Y, Liu, H. Z. (2021). Developing an intelligent virtual reality interactive system based on the coffee model for learning pour-over coffee brewing. Computers and Education: Artificial Intelligence, 2. https://doi.org/10.1016/j.caeai.2021.100030

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Educational Resource*, 11(2). https://eduimed.usm.my/EIMJ20191102/EIMJ20191102_06.pdf

Saleekongchai, S., Bengthong, S., Boonphak, K., Kiddee, K., & Pimdee, P. (2024). Development Assessment of a Thai University's Demonstration School Student Behavior Monitoring System. *Pakistan Journal of Life and Social Science*, *22*(2).

Nasir, A. M., & Mustapa, I. R. (2024). Building The Nigerian Corporate Governance Index (NCGI) and Intellectual Capital Disclosure Practices. *Pakistan Journal of Life and Social Sciences*, *22*(1).