Trends

JALTCALL Trends
ISSN 2982-3749, 2(1), 103496 (2026)
https://doi.org/10.29140/jct.v2n1.103496

Castledown

Developing an AI-powered VR chatbot for immersive medical English training

Takeru Kanahama
Tokyo Denki University, Japan
gandmf15@gmail.com

Jeanette Dennisson
St. Marianna University School of Medicine, Japan
dennisson@marianna-u.ac.jp

James York
Meiji University, Japan
jamesyorkjp@gmail.com

Gary Ross
Kanazawa University, Japan
gary.ross@icloud.com

Hiroshi Nakayama
Tokyo Denki University, Japan
nhiroshi001@gmail.com

Japanese medical students have limited opportunities for medical English communication training. To overcome these limitations, we have been developing AI-driven medical English communication tools that utilize speech-to-text technology. Here, we developed a tool that leverages virtual reality (VR) to practice medical interviews with an AI patient. We tested this tool on pre-clinical Japanese medical students in comparison with a browser-based AI-driven communication tool. Based on the post-training student survey, the VR-AI tool was preferred over all other alternative training methods, including those involving human simulated patients and the browser-based chatbot tool. Further development and testing are needed to address technical parameters and reliability, allowing for improved future medical English communication training tools.

Keywords: virtual reality, artificial intelligence, chatbot, speech synthesis, English for specific purposes

JALTCALL Trends
Vol. 2 No. 1
2025 JALTCALL Conference Papers Special Issue

Background

English for Specific Purposes (ESP) programs are increasingly essential to meet the rising demand for professionals with effective English communication skills. In response to the growing number of foreign patients in Japanese hospitals (Ministry of Health, Labour and Welfare, 2024), medical schools need to better prepare students for real-world clinical interactions. However, English for Medical Purposes (EMP) programs often lack instructional time and resources to provide effective practical medical communication training. Table 1 compares commonly used instructional methods and their limitations.

Table 1. Comparison of medical interview training methods in EMP curricula in Japan

Method	Outcome	Limitations
Scripted dialogues (listening/ reading/shadowing/creating) (Scripted)	Targeted vocabulary & grammar building	No speaking; limited feedback; reading/ memorization focused
Student-student role play (Peer)	Basic speaking & listening practice	High anxiety; limited feedback; reading/ memorization focused
Simulated patient-student role play (SP)	Authentic speaking & listening practice; Quality, real-time feedback	High anxiety; limited practice time; limited budget

To address these limitations and provide more targeted, practical speaking opportunities, we have been developing medical communication training tools that integrate speech recognition and chatbot technologies – collectively referred to as Medical Language Interactive Learning Agents (MedLiLAs). Our initial browser-based version of MedLiLA (hereinafter, browser-MedLiLA) was designed to simulate a medical interview with an AI patient (Dennisson & Ross, 2023). This study investigates the feasibility of extending MedLiLA to a virtual reality (VR) environment (hereinafter, VR-MedLiLA) to provide immersive, realistic interview training, and compares it to browser-MedLiLA.

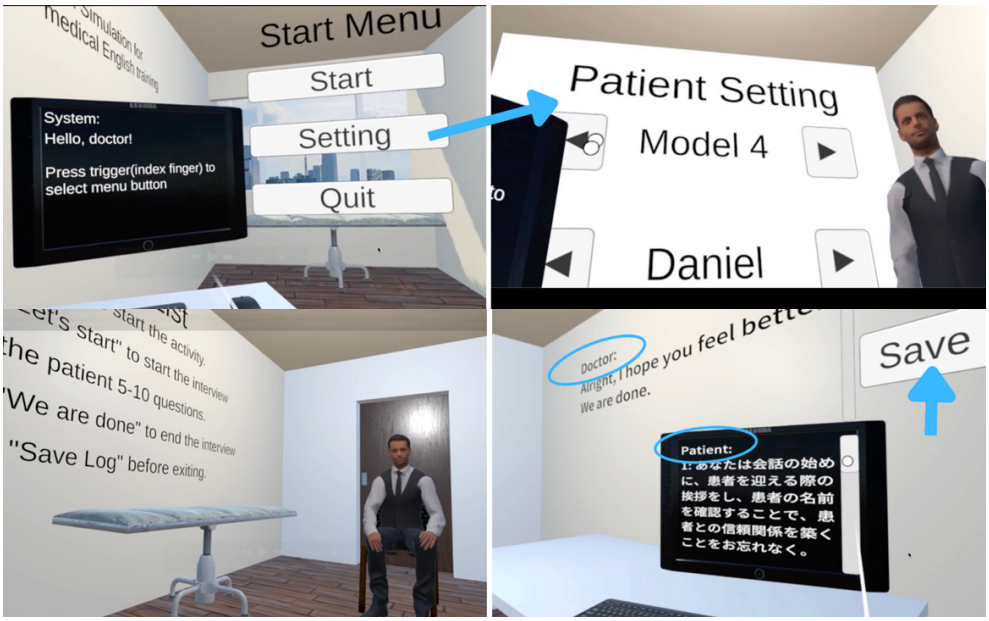
A systematic review of VR learning tools in medical education identified significant improvements in knowledge acquisition, practical skills, and learner satisfaction (Lampropoulos et al., 2025). Moreover, VR learning environments have shown the capacity to reduce anxiety and increase learners' willingness to communicate (York et al., 2021). Compared to conventional role-play, VR allows for repeated, adaptive practice across different scenarios, helping to improve fluency, lexical precision, and communicative confidence. One VR application in an EMP Collaborative Online International Learning context has already been reported (Yamashita & Vehmaskoski, 2022). However, to date, no study has applied VR to the EMP setting specifically for medical interview training. Therefore, this study provides insight into the potential of VR communicative training in medical English education.

Methods

We developed a VR environment for the MedLiLA system that simulates a medical interview examination – a high-stakes examination in all medical programs (i.e., Objective Structured Clinical Examination or OSCE). A small cohort of students were recruited to test the feasibility of VR-MedLiLA for OSCE training. They compared their experience of VR-MedLiLA with browser-MedLiLA and provided feedback via a post-training survey.

System description

The VR environment of VR-MedLiLA was developed in Unity, with Whisper used for speech recognition, to replicate a realistic examination room. The software program was deployed on a Meta Quest 2 head-mounted display (HMD) (see Figure 2). In the program, the user plays the role of the doctor and the AI chatbot (ChatGPT API) plays the role of patient. The examination room includes a desk with a PC, chair with a 3D avatar patient, window, door, and examination bed (see Figure 1).



Note. *A) Start Menu; B) Settings; C) Checklist; and D) Text Output and Save functionality.

Figure 1. Overview of the VR-MedLiLa environment

At program startup, users access “Start Menu” which includes “Start”, “Setting”, and “Quit” (see Figure 1A). Under Setting, users select between four 3D avatar models and four patient scenarios (see Figure 1B).

The user’s output is displayed on the right wall and the AI patient’s output is displayed on the PC (see Figure 1D). As a guide, a checklist of tasks is provided on the left wall (see Figure 1C). Each scenario is programmed to provide feedback in Japanese at the end of the interview by saying: “We are done” (see example in Figure 1D); this instructional language was maintained between MedLiLA modalities.

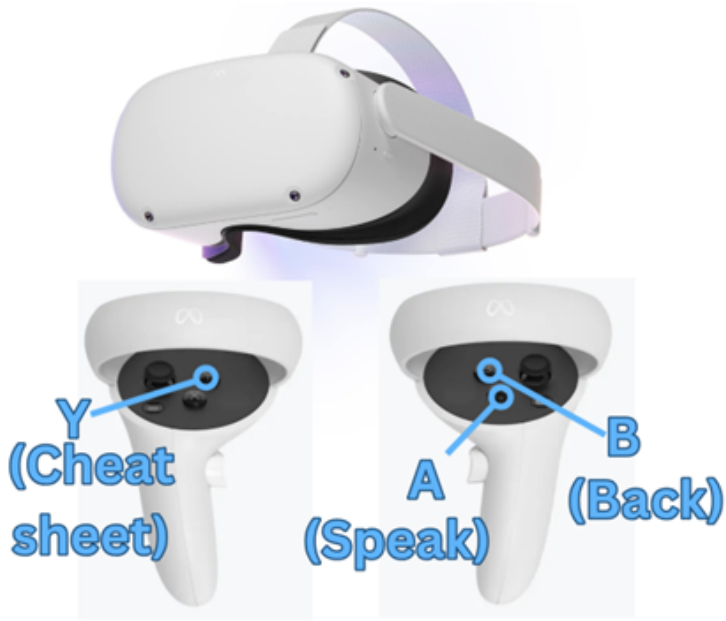


Figure 2. Meta Quest 2 HMD and controllers

Controllers of the HMD allow users to speak (holding down A with thumb), go back to the main menu (clicking B), and view a "Cheat Sheet" of sample interview questions (clicking Y) (see Figure 2). In addition, either trigger finger button allows navigation within the program. At the task start, a tutorial acclimatizes users to the controller functions; basic instructions of how to wear the device and hold the controllers were demonstrated by the researchers.

Patient scenarios and post-training feedback prompting structures designed for browser MedLiLA were adapted and designed to accommodate VR-MedLiLA. These scenarios were then incorporated within the VR program. Within VR-MedLiLA, user's spoken text is recognized by ChatGPT's API through speech-to-text conversion via Whisper, and the API responds to user text (e.g., "Why did you come to the hospital today?") using text-to-speech conversion. Users then hear a response as spoken text (e.g., "I have a headache").

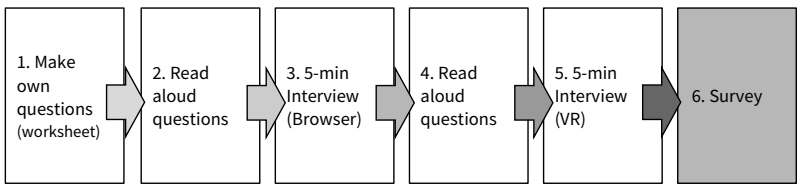


Figure 3. Study design to compare browser- vs VR-MedLiLA platforms

This comparative study recruited 24 pre-clinical (fourth-year) Japanese medical students (with English proficiency averaging around CEFR B1-B2) to test the initial version of VR-MedLiLA. All students had already experienced the set of browser-MedLiLA activities in a medical English course immediately prior to the study.

Therefore, they were already familiar with the browser-MedLiLA system before recruitment. The study design of the comparison involved interaction with reading and creating interview questions, followed by training with the browser-MedLiLA and then VR-MedLiLA, as illustrated in Figure 3. After completion of the study, a post-training survey was administered. For the preference comparison, participants ranked the five modalities, and responses were analyzed using a simple weighted ranking approach (1st = 5 points, 2nd = 4 points, etc.). Additional Likert-scale items assessed perceived task difficulty, usefulness for English practice, alignment with course objectives, and system usability, along with a small number of open-ended response items.

Results

Students reported almost no prior use of VR tools for daily activities, studying, career, or fun; only one student reported using it for studying purposes before this study. Thus, all students needed basic instructions on how to use the Meta Quest 2 HMD and controllers.

The post-training survey measured perceived effectiveness of the VR tool over other teaching methods. Students were asked to rank the five methods from 1 to 5. In the rating system, we weighted each rank by giving 5 points for first place down to 1 point for fifth place, thereby the higher the total score, the more perceived effectiveness by students. Based on this rating system, role-play training with the VR patient was ranked the highest among the five methods, slightly more effective than role-play training with a human simulated patient and browser AI patient; listening to dialogues was considered the least effective (see Table 2). As the survey was not formally validated and the ranking method was exploratory, findings should be interpreted as indicative rather than conclusive.

Table 2. Perceived effectiveness ranking of teaching methods of medical interview training

	1st	2nd	3rd	4th	5th	Rating score
Scripted	1	2	4	10	7	52
Peer	2	2	9	7	4	63
SP	9	6	3	6	0	90
Browser	4	7	11	1	1	84
VR	11	10	2	1	0	103

JALTCALL Trends
Vol. 2 No. 1
2025 JALTCALL Conference Papers Special Issue

Table 3. Perceived reliability of VR-MedLiLA by students

	Never	Sometimes	Always
Good opportunity to practice English	1	4	19
Met course objectives*	1	5	18
Easy to complete	1	17	6
Challenging to complete	6	16	2
AI responded appropriately	1	4	19
AI spoke too fast	11	11	2
AI spoke difficult words	12	9	3
Understood the output of AI	2	9	13
Unable to complete sentences	8	12	4
AI gave feedback	3	6	15
Feedback was helpful	3	8	13

Note. *Course objectives refers to those for the compulsory EMP course.

Table 3 details both reliability of perceived effectiveness and technical parameters of VR-MedLiLA. Overall, almost all students found the task to be a "good opportunity to practice English" and believed it "met course objectives" all the time. These responses suggest generally positive perceptions of VR-MedLiLA's usefulness and alignment with course objectives.

The AI patient generally "responded appropriately" (96%). Responses to the paired items "easy to complete" and "challenging to complete" showed a consistent pattern: most students found the task easy to complete at least sometimes (96%), while only two students reported that it was always challenging.

Regarding AI patient output, about half of the students found no problems with speaking speed (46%) and vocabulary difficulty (50%); two students found the output to always be too fast (8%) and three found the vocabulary too difficult (12.5%). Two thirds of students (67%) reported that they were unable to complete their sentences during the interaction; this appears to be related to the system design, in which the AI patient was programmed to respond after a short period of detected silence, potentially interrupting user speech rather than reflecting limitations in student language ability. Finally, feedback functionality worked for most students (87.5%) and was perceived as helpful (87.5%).

Discussion

Overall, VR-MedLiLA was positively received as a medical interview training tool in English and was perceived as the best training method overall by fourth-year medical students in a Japanese university. The students were able to use VR-MedLiLA despite having little or no prior experience with VR, suggesting that the tool was accessible to first-time users while still requiring active engagement with the interview task.

Immediate informal feedback from students following VR-MedLiLA testing included spontaneous comments such as "sugoi" [wow] and "this one [VR device] is better." Some students also recognized a useful training effect of the VR format, noting that it limited their ability to rely on prepared notes and required them to formulate questions from memory. This added cognitive demand may encourage

greater recall and spontaneous language production, making the interaction feel closer to a real medical interview. In addition, follow-up discussions with fourth- and fifth-year medical students suggested potential interest in adapting VR-MedLiLA for pre-clinical clerkship training in Japanese. Taken together with the perceived effectiveness ranking results, students indicated a positive initial reaction to VR-MedLiLA as a tool for medical interview training.

The preference for VR-MedLiLA was observed despite some system problems and usability issues, including AI speaking speed, linguistic difficulty and length of the AI patient's output, interruption of user speech before sentence completion, and variability in feedback quality. The VR tool may be improved by adjusting the current prompt structure. For example, the patient scenarios could be scaffolded by language level using English as a Foreign Language (EFL) frequency word lists (e.g., NGSL, K2000) or language complexity parameters (e.g., sentence length, CEFR level). While VR-MedLiLA already supports listening comprehension by allowing users to read their own and AI's output live, different speech synthesis mechanisms could also be explored to adjust the speed at which the AI speaks to accommodate lower level EFL students. Finally, while there is a "save" function that creates a file of the interaction in a researcher-accessible folder upon completion, making these resulting interactions accessible to the teacher and student would be useful in allowing both to evaluate student performance and to conduct post-training reflection and teacher-led feedback sessions.

Conclusion

VR-MedLiLA shows promise as an immersive authentic tool of medical interview training for EMP programs in Japanese universities. VR-MedLiLA addresses some of the limitations of existing EMP training methods by providing repeated domain-specific speaking practice with immediate AI-generated feedback in a low-anxiety, low-stakes, low-invasive environment. Based on the results of this pilot study, we will address technical issues and improve the environment to provide a more effective and reliable training tool, adaptable to any EMP program and pre-clinical clerkship training courses in Japanese.

References

Dennisson J. & Ross G., (2023). Medical English communication training using ChatGPT and speech recognition. *Journal of Medical English Education, 22*(2/3), 87–90.

Lampropoulos, G., del Bosque, A., Fernández-Arias, P., & Vergara, D. (2025). Virtual reality in medical education, healthcare education, and nursing education: An overview. *Multimodal Technologies and Interaction, 9*(7), 75. https://doi.org/10.3390/mti9070075

Ministry of Health, Labour and Welfare. (2024, March). *2023 Survey on the acceptance of foreign patients at medical institutions.* https://www.mhlw.go.jp/content/10800000/001292094.pdf

Yamashita, Y., & Vehmaskoski, K. (2022). Practice report: On the development of medical English ESP learning model via VR space in collaboration with Finland. *AsiaCALL Online Journal, 13*(2), 161–171. https://asiacall.info/proceedings/index.php/articles/article/view/161

York J., Shibata K., Tokutake H., & Nakayama H. (2021). Effect of SCMC on foreign language anxiety and learning experience: A comparison of voice, video, and VR-based oral interaction. *ReCALL, 33*(1), 49–70. https://doi.org/10.1017/S0958344020000154

JALTCALL Trends
Vol. 2 No. 1
2025 JALTCALL Conference Papers
Special Issue

Author biodata

Takeru Kanahama is a masters student at Tokyo Denki University exploring research on the use of VR in language learning contexts.

Jeanette Dennisson is a university professor developing AI-driven communication tools using speech recognition and chatbots for medical English education.

James York is an associate professor at Meiji University researching the intersection of games and education.

Gary Ross is a professor at Kanazawa University researching speech recognition and AI in language education.

Hiroshi Nakayama is a professor at Tokyo Denki University and supervises students in the Information Systems and Design Department.