



Castledown

 OPEN ACCESS

Technology in Language Teaching & Learning

ISSN 2652-1687

<https://www.castledown.com/journals/tltl/>

Technology in Language Teaching & Learning, 6(3), 1716 (2024)

<https://doi.org/10.29140/tltl.v6n3.1716>

Pecha, A Language Practice Peer: Guiding Language Learning Interactions Through Large Language Models



RYAN LEGE^a 

EUAN BONNER^b 

TAKAKO AIKAWA^c 

^a*Kanda University of International Studies, Japan*
lege-r@kanda.kuis.ac.jp

^b*Kanda University of International Studies, Japan*
bonner-e@kanda.kuis.ac.jp

^c*Massachusetts Institute of Technology, United States*
taikawa@mit.edu

Abstract

The interaction hypothesis of second language acquisition (Long, 1981) states that negotiated interaction is necessary for language development. In many language learning contexts, educators and stakeholders seek to provide opportunities for learners to engage in meaningful real-life interactions that help them build linguistic, semantic, and rhetorical competence. However, the opportunities provided for interaction can vary in their degree of effectiveness and may only sometimes lead to increased language ability. If these interactions are scaffolded correctly, they can be *tuned* to maximize their benefits (Loewen & Sato, 2018). Unfortunately, this is not always practical from a temporal and economic perspective. Artificial intelligence (AI) could be the solution for providing learners with individualized, comprehensive assistance during their learning interactions. Accordingly, the authors developed a bespoke application that employs advanced natural language processing and large language model AI technologies to support learner interactions. The application was developed to support students in two different contexts: Kanda University of International Studies in Japan, where students study English, and Massachusetts Institute of Technology in the United States, where learners study Japanese. The rationale for creating the application and selecting its major features is discussed. This is followed by a discussion of how the application functions and how it will be used.

Copyright: © 2024 Ryan Lege, Euan Bonner and Takako Aikawa. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within this paper.

The authors will then discuss their plans for implementation into both informal and formal learning contexts at the two universities. They conclude by discussing potential limitations and plans for improving the application.

Keywords: CALL, artificial intelligence, AI, large language models, LLM, language learning, interaction

Introduction

Language learning contexts and rationales vary considerably worldwide. Language skill can be necessary for career advancement, daily survival, or can be pursued out of intellectual curiosity (Gallagher-Brett, 2004). In some contexts, language learning is a mandatory part of the educational system. Sometimes, when learning languages, access to speakers of the target language may be limited, so learning to communicate can present unique challenges. Though language exists to transmit information between parties, its study can often be divorced from this purpose if there are few opportunities to apply a target language in authentic, meaningful contexts.

The significance of real language practice in second language acquisition (SLA) is a cornerstone of Long's (1981) interaction hypothesis, one of the most influential theories in SLA literature (Tran, 2009). Long (1981) observed that the conditions for language acquisition are best fostered through modification and negotiation in authentic conversations with target language speakers. Namaziandost and Nasri (2019) note that while learners can possess both spoken fluency and accuracy, they may still have difficulty participating in interactions, highlighting the important role of interaction as a mediator of communication. Ellis (1991) further expanded on this, positing that the effectiveness of input is enhanced through the negotiation of communication issues, which tunes the input to the appropriate level for language learners. This concept aligns with Krashen's (1982) theory of comprehensible input, though rather than ensuring an initially suitable input state, Long focuses on the interactive, adaptable nature of the input (Tran, 2009). While the interaction hypothesis is widely accepted as a leading explanation of SLA, and the literature broadly supports its claims, scholars (Tran, 2009; Ebrahimi, 2015) concur that it only partially elucidates the intricate language acquisition process. Nonetheless, it is widely acknowledged that interaction and negotiation of input are vital components of SLA.

In contexts with limited access to target language speakers, the challenge of providing opportunities for authentic language practice is being addressed in many ways. Advances in computer-mediated communication (CMC) allow people to engage with others remotely, providing more opportunities to practice than were ever available before (Pouromid, 2019). Collaborative international online learning (COIL) has become a favored approach to connecting language learners with speakers of their target languages. Effective COIL, however, must navigate around some unique challenges. These challenges include scheduling, high-speed internet access, the cost of arranging exchanges, and the constraints of CMC (Kučerová, 2023; Minei et al., 2021). Even when there is access to synchronous video conferencing, the field of view limits nonverbal cues (Koester, 2022), and audio channels tend to overlap, making some of the natural constituents of conversation, like backchanneling and interrupting difficult (Boland et al., 2022). Combined, these factors mean that while COIL is effective in many contexts, it cannot be universally applied in all contexts that need interactive language practice opportunities.

Recently, the hegemony of generative AI has pushed it to the forefront of academic discourse in all disciplines, including second language acquisition. This creates another potential avenue for addressing the need for interaction and negotiation in a target language. Educators and researchers have begun to

explore how Large Language Models (LLMs), AI models trained on enormous amounts of language texts that can generate human-like text by predicting the next word in a sequence, may be applied as practice partners for language learning (Belda-Medina & Calvo-Ferrer, 2022). However, there is still a lack of consensus on whether these models can provide authentic communicative or interactive experiences, with researchers noting that there are still many shortcomings (Zhai, 2023). Wood and Moss (2024) argue that human interaction is “irreplaceable” and crucial for “enhancing students’ critical thinking abilities, fostering creativity and providing essential context and perspective” (p. 164). Resnick (2024) proposes that AI tools should be used to support “creative learning experiences [that] involve people learning with and from one another” (p. 8). In accordance with Resnick’s proposal, and given the current state of LLMs, instead of pursuing AI as a replacement for human interaction, it may be more beneficial to consider the potential of AI to *augment and support human interaction*.

Language educators in many contexts still rely on creating face-to-face opportunities for interaction, despite the current focus on the capabilities of LLMs. This is the case for the institutions where the authors of this paper teach foreign languages. The authors currently teach in very different contexts and are responsible for teaching different languages, Japanese and English. The first context is teaching Japanese at a private research institution in the United States, Massachusetts Institute of Technology (MIT). The second context is teaching English to students at a private university, Kanda University of International Studies (KUIS) in Japan. Each context deals with the issue of providing opportunities for real language practice in different ways.

Japanese Practice at MIT

Students studying Japanese at MIT can participate in the *Japanese Lunch Table*, where students can practice Japanese with native speakers while having their lunch. The sessions are not limited to MIT students; they are also available to exchange students, visiting scholars and their spouses. The Japanese Lunch Table is very informal and welcomes students of all proficiency levels, from beginners to advanced speakers. However, the lack of structure may result in participants having limited real conversations in Japanese.

In a third-year Japanese class at MIT Japanese speakers are invited to join the class via *Zoom* twice a semester, allowing students to practice their Japanese in *Zoom*’s breakout rooms. During these sessions, students are encouraged to ask questions, with each breakout room assigned one native Japanese speaker and two or three students. Each session lasts 45 minutes. The students enjoy this experience as it provides a rare opportunity to engage in in-depth conversations with Japanese speakers, beyond just greetings and self-introductions. However, the sessions occur only twice a semester, which is not frequent enough, and providing real-time feedback during the sessions is challenging despite visits to all the breakout rooms.

English Practice at KUIS

At KUIS, all students study English in addition to other languages, global topics, and liberal arts subjects. To provide opportunities to practice the target language, the university has prepared facilities and services for students’ benefit. They can meet with lecturers for formalized practice sessions or join informal group discussions in the English lounge, an all-English area available to students during the university’s operating hours. In addition, students can choose language practice partners from the body of exchange students at the university to provide mutual practice opportunities for both parties. The wide variety of formal and informal opportunities for students to use their language of study is one of the strong points of the university, which has also been expanded to offer online, synchronous video sessions. However, the availability of self-access language practice facilities and services is not

without its issues; sometimes, students may find it challenging to fit practice sessions into their busy schedules and may come out of sessions unsure of how they benefited (see Mynard et al., 2020). This results in some students not using the services offered or not using them as much as they need for practice.

Issues with Language Practice Sessions

In the current language practice sessions offered in both contexts, learners have valuable opportunities to utilize their language skills. However, there are significant areas for improvement. When learners engage in real interaction and negotiation of meaning, it can be challenging to document and retain new phrases, insights, and other information in real time. Even if students engage in a post-interaction reflective task, they may struggle to recall important conversation details. This can lead to a situation where the experiences are valuable yet have little long-term effect on language development. The effectiveness of informal language practice sessions can vary greatly depending on the session structure and individual learner variables. As Loewen and Sato (2018) stress, mere interaction is not enough; educators should focus on maximizing the impact of interaction for the benefit of language learners. AI, specifically LLMs, can provide solutions to these problems.

AI as a Solution

What Are LLMs?

First proposed by Google researchers (Vaswani et al., 2017), deep learning transformer-based LLMs are AI systems trained on vast quantities of human-generated content and can produce human-like text and other media based on their training data. As the technology has improved, they have become expert-level text generators in many fields including language acquisition (O'Grady & Lee, 2023). Since the advent of these LLMs, most notably the LLM powering OpenAI's *ChatGPT*, language teachers, researchers, and learners have begun exploring and applying them in both formalized and informal instruction.

Many of the current LLMs available for use, such as *OpenAI's GPT* range, *Google's Gemini*, and *Anthropic's Claude*, are also available as application programming interfaces (APIs), allowing for application developers to integrate them into their own applications. Applications that have begun to incorporate LLMs to implement such intelligent systems and personalization include *Duolingo*, *Squirrel AI*, *Mathway*, *Grammarly*, *Coursera*, *VIPKid*, and *Khan Academy*.

How Are LLMs Being Applied in Language Education?

While the field of computer-assisted language learning (CALL) has been using natural language processing (NLP) models for years (Antoniadis et al., 2013), widespread access to LLMs that are simple and easy to use has made NLP more accessible than ever before. Gao et al. (2023) demonstrated through empirical means that LLMs can exhibit high spoken language intelligence and impressive accuracy when prompted with specific domain knowledge and best prompting techniques. They note that LLMs "hold considerable promise for improving conversational spoken language learning" (p. 15). LLMs have even been shown to be an effective method for generating written feedback and can be a positive and motivating tool to improve student writing (Meyer et al., 2024).

Additionally, LLMs are increasingly being used as tutors to help support learners and give them practice opportunities. Kwon (2023) designed a custom tutor, *GPTutor*, to generate personalized feedback to learners for study at the appropriate language level for students. Liu et al. (2024) investigated using

LLMs to enhance an intelligent tutoring system that used pictures as prompts for conversation with an LLM. They found that the LLM could offer various types of cognitively scaffolded feedback based on constructivist learning theories to “align with personalized learning needs” (p. 6). The increasing use of LLM-powered chatbots in educational settings like this, both on and off-campus, is transforming student support and practice opportunities. These chatbots, with their advanced capabilities and accessibility, are fostering engaging and interactive learning environments that encourage independent learning and boost student engagement (Kooli, 2023; Wang & Xue, 2024).

Moving beyond tutoring systems, Yu and Guo (2023) discussed additional avenues for artificial intelligence systems powered by LLMs to enhance educational outcomes. They suggest that intelligent systems could be used for marking homework and additional speech interaction with the content. They also discuss using LLMs to personalize assessment, learning content, teaching routes, and learning experiences.

How Are the Authors Applying LLMs in Their Contexts?

The authors at KUIS have been utilizing this access to LLM APIs since 2021 in several projects aimed at investigating its usage in language acquisition. In 2021, a study was conducted on the potential for LLMs to be utilized by teachers and students in the classroom (Bonner, Lege, & Frazier, 2023). In 2022, the authors created and tested a student-customizable AI conversation partner, where students were able to create characters with individual appearances, personalities, and life histories for classroom conversation practice (Loretzen & Bonner, 2023). Current projects include a classroom material-trained AI teacher assistant web application for students to query during class to aid in their understanding and AI-generated live-action cartoons that students can adjust the content and language level of in real time to suit their learning goals.

The author at MIT has been adapting the use of LLMs since 2023 for her third-year Japanese course. She has explored whether AI could assist by providing instant feedback on student writing, thus freeing up time for instructors to focus more on designing in-class learning activities. The initial prompt provided to the LLM used for this assignment is simple: “modify my Japanese.” Students are then required to perform a writing assignment that consists of three elements: (i) their original sentence, (ii) their modified version by AI, and (iii) their reflection and/or analysis on the differences between (i) and (ii). Overall, the feedback from students regarding the use of AI for this assignment has been positive. They particularly appreciate being able to consult with AI anytime and anywhere and find that they can learn vocabulary and grammar beyond the scope of the textbook. However, some students have expressed doubts about the efficacy of AI in improving their writing skills.

In 2023, the authors at KUIS and MIT agreed to combine their LLM research experience by collaborating on a project to investigate the role that LLMs could play as third-party participants in language conversation practice, customized to the student’s needs and able to provide post conversation feedback based on their specific needs and performance. Across society, much of the focus is on fears that AI will replace human teachers and tutors in the current language teaching and learning landscape. While the ability for AI to complete tasks with near human efficiency is possible, the authors felt that instead of using AI as replacements, it would be more meaningful to consider AI’s potential to augment the human-to-human component of language learning. Rather than replacing the practice partners or tutors with AI versions, it would be better to focus on the ability of AI to provide scaffolding support and feedback. Therefore, the authors are developing an application, *Pecha* (named after the Japanese phrase *pecha kucha*, which is an onomatopoeic expression that means chattering), to allow learners to maximize their access to the affordances of face-to-face interaction. To do this, the

authors have reconceptualized the role of AI not as a replacement conversation partner or teacher, but as a peer that supports learners before interactions, participates during them, and helps the learners reflect on their experience afterwards.

Application Description

In both the KUIS and MIT contexts, language learners often practice what they have learned by engaging in language practice sessions with another speaker of the target language. For this project, the authors are developing Pecha to support learners in these interactions. A learner sits down with a practice partner or instructor and engages in a 15-minute conversation on a topic of their choice in the target language. The application is displayed on a monitor and sits to the side of the table between the two speakers. The application harnesses a combination of LLM generative AI, speech-to-text, and text-to-speech technologies, initially supporting both Japanese and English modes, with a modular design that allows for the inclusion of additional languages. This technical integration is facilitated through a modular framework utilizing cloud server requests to several different online services for all core functionality. The backbone design of the application is platform agnostic, allowing for the testing and use of various interchangeable AI services depending on the circumstances. The flexible nature of the application allows it to be adapted for different contexts where there may be varying pedagogical needs, such as scaffolding for lower-level learners in their first language.

For AI, the application can switch between accessing LLM platforms such as those offered by OpenAI, Anthropic, Google, and even open-source fine-tuned language-specific LLMs, ensuring adaptability to the evolving landscape of LLM AI technologies. For example, in cases where a significant amount of text needs to be sent for analysis and summarization, LLMs with APIs supporting extremely long context windows such as Gemini (which currently supports up to approximately 2 million short words, or tokens, of content) can be accessed and used.

Speech-to-text is handled by individual speaker diarization via speech-to-text cloud services. Diarization enables each speaker to be transcribed separately, enabling both overall conversation and individual speaker analysis by isolating specific speakers during conversations. Services such as *Amazon Web Services*, *Assembly.ai*, and *Deepgram* offer varying levels of accuracy depending on the audio environment and language. This speaker role separation subsequently allows for the isolation of both the learner's and interlocutor's output which can then be forwarded individually to the LLM for processing while still allowing the separate processing of the conversation as a whole. As the ongoing diarized transcript is received, it is paired with specialized commands that instruct the AI at regular intervals to monitor and respond to issues with vocabulary misapplication, sentence structure, and conversation coherence. Additionally, it catalogs these issues along with an overall conversation summary for personalized post-session feedback.

Privacy and data protection are paramount in the design of this system. Voice recordings are immediately discarded after they are converted to text. When submitting content for analysis, the application utilizes LLM platforms with strict policies against using user data for training or other purposes. Additionally, standard secure online databases handle and store all incoming and outgoing information. Students are also fully informed about data handling practices at the beginning of their sessions and are advised to avoid discussing personal or identifying information during their interactions. Finally, to ensure students only access their own content, the system generates a five-digit alphanumeric passcode for them to use at the beginning of each session.

Text-to-speech is integrated to provide the option for auditory as well as text-based responses by the AI and utilizes cloud services that provide realistic life-like voice synthesis such as *Resemble*.

ai, Google Cloud Voices, and OpenAI text-to-speech. The application also features an intuitive touch interface paired with an animated avatar for enhanced interaction in language acquisition scenarios. Based on existing artwork used in a previous AI chatbot project (Lorentzen & Bonner, 2023), the user can customize the AI's appearance and voice (Figure 1). UI and UX appearance and flow are based on current best practices. The combination of realistic voice synthesis, customizable avatars, and user-friendly design elements works to create a more immersive and less intimidating learning experience, potentially boosting learner confidence and participation.

Needs Assessment

For Pecha to meet learner needs, it needs to know more about their background and learning goals. On first use learners create an account, and then an intuitive onboarding system guides them through a series of simple questions to elicit key information about their language learning needs and goals. For example, if the learner is taking a language aptitude test, they can specify which areas of the test are the most problematic and give their desired outcome/score. They are also prompted to add any language structures or vocabulary they want to use during their session. If they wish to add any additional materials or texts, they can also do that here (Figure 1).

After learners have used the application once, the pre-session process is simplified for subsequent uses, with learners having the option to revise or supplement this information. If the learner has completed a prior session, Pecha welcomes returning learners with a succinct overview of their previous session, bridging past and present learning while promoting progress through strategic repetition.

Interaction Phase

During the conversation practice session, the AI can be set to play the role of an active participant in addition to listening and taking notes for post-session feedback. When Pecha is in active mode, it listens

Figure 1. Pre-interaction needs assessment prototype menu.

and requests permission to intervene via a non-intrusive visual/audio cue to offer real-time assistance (Figure 2). The AI can participate in the discussion, inject comments, ask probing questions, and subtly correct incorrect vocabulary by weaving the right words into its responses. Users can prompt the AI for suggestions or examples and even set it to challenge their opinions, enhancing critical thinking through conversational role-play. As Pecha is not placed directly between the participants, they can choose to ignore any of these cues.

On a technical level, this is achieved by having the application record audio in fifteen-second portions and send them sequentially to a speech-to-text transcription service, which diarizes and returns the transcription to the application. The application then sends that transcription regularly to an LLM for summarization and language learning analysis.

The analysis itself consists of monitoring the transcription to identify the topic of conversation, the opinions of the speakers, and potential areas for the application to chime in with its own opinions and suggestions. Additionally, on the language learning assistance side, the application also monitors the type of vocabulary and sentence structures being used so that when it chimes in, it can attempt to utilize them in a contextually appropriate manner. This content is also provided to the learner at the end of the conversation for self-study.

Pecha determines when to chime in based on the contents of the conversation, what the students set as their learning goals, and the amount of time that has passed since it last contributed. This is determined by comparing the student's stated goals with the content of the ongoing conversation and identifying patterns of vocabulary misapplication, conversation topic expansion, areas for potential sentence structure improvement, and other goal-related issues. When the application determines that it is suitable for it to chime in, the screen provides a visual cue, and a light audio tone is played. The learner can then decide to ignore the request or tap the screen to allow Pecha to begin speaking using

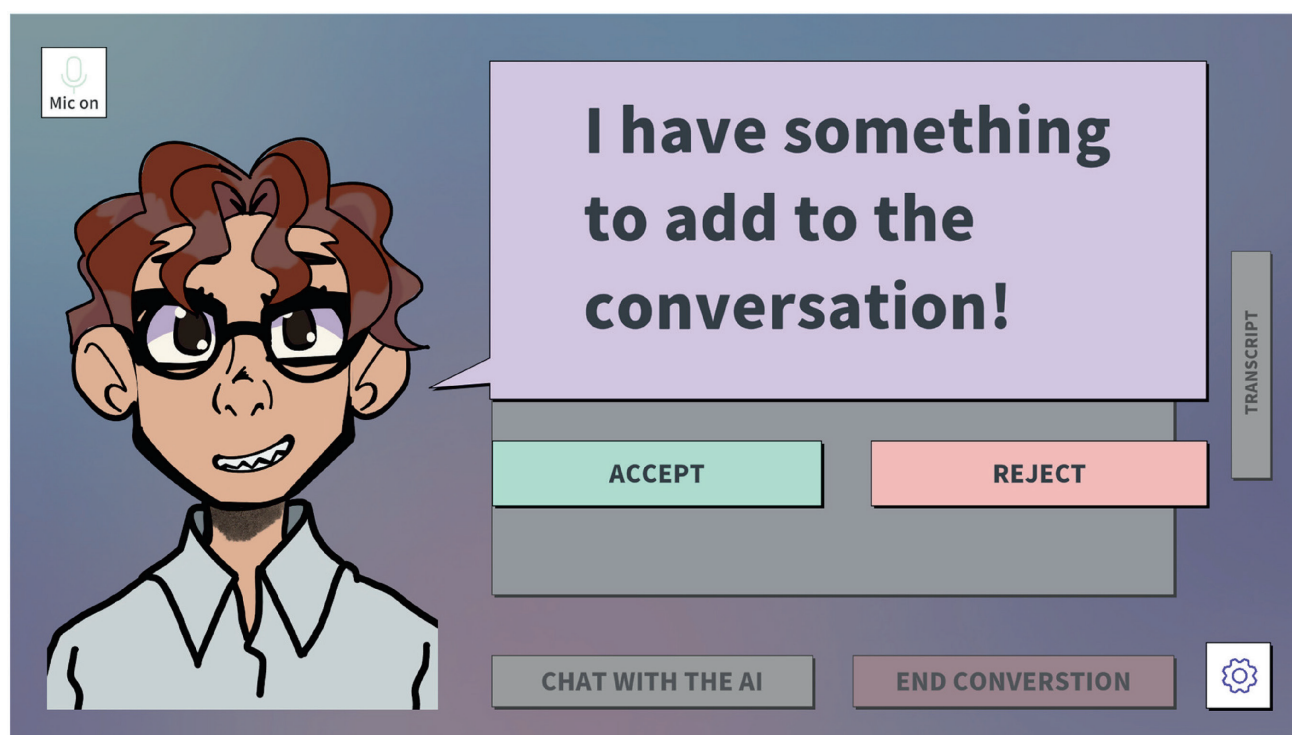


Figure 2. During interaction prototype screen.

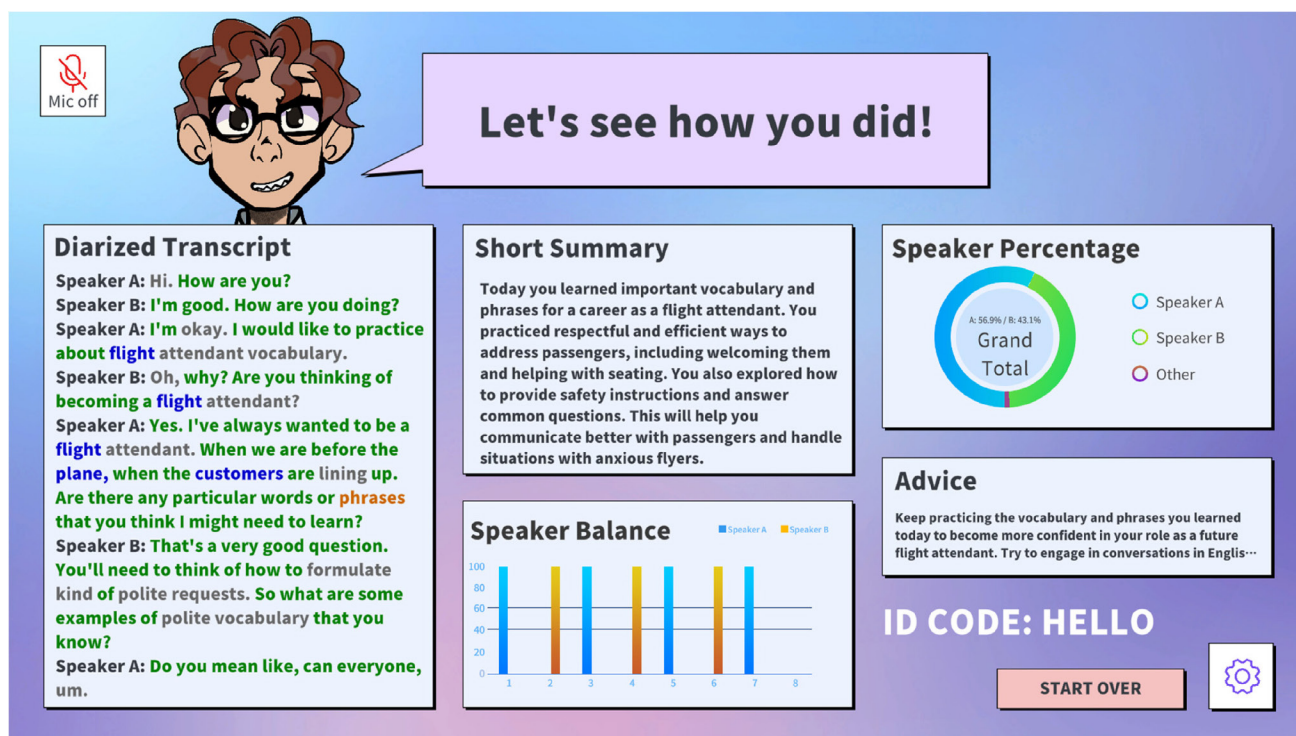


Figure 3. Post-interaction dashboard prototype screen.

text-to-speech. Significant attention has been paid to making sure that the chime-in cues do not overly distract the participants from their conversation and students are also free to press a button to initiate a direct conversation with the AI to get specific feedback whenever desired. These features strike a balance between learner autonomy and guided instruction, allowing users to choose when they receive feedback while ensuring all insights are firmly grounded in the context of their interaction.

Feedback and Reflective Analysis

Post-interaction, Pecha presents the learner with a diarized interaction transcript. The application also analyzes the transcript for the complexity of lexical items and applies formatting to highlight vocabulary used based on word lists and the target vocabulary identified by the student. The transcript is segmented by topic, with summaries of the interaction provided in a panel of the user interface. This allows the learners to recall and reflect on their performance during the interaction. Finally, Pecha offers suggestions about vocabulary and language to improve future interactions (Figure 3).

Additional information and features can also be easily provided, such as suggested learning plans. This data is also fed back into the pre-interaction system to supplement the original data provided in the needs analysis step for subsequent interactions.

Limitations

Diarization Accuracy

One of the most important elements of the system is its ability to discern individual speakers through diarization. The ability to differentiate the learner from their interlocutor in the transcription and

provide advice relevant only to the learner is paramount. The technology has advanced significantly and continues to do so, but it is not infallible. While rare, instances where both speakers are at times incorrectly merged into a single line of dialogue attributed to only one speaker do occur. Such inaccuracies can impede the application's ability to provide precise, personalized feedback even though it does not affect the AI's understanding of the overall conversation. Fortunately, the diarization system is overall quite capable of differentiating speakers by voice, and the merger of two speakers into a single line of dialogue has so far not resulted in the rest of the transcription subsequently mixing up roles as to who is Speaker A or Speaker B. However, one of the more common problems being faced is that the diarization system always assigns the first speaker heard to the role of Speaker A, necessitating an opening message being displayed on the screen that the learner always commence the conversation for the application to provide assistance to the correct speaker.

Transcription Accuracy

The effectiveness of the application is also reliant on the clarity of the learner's pronunciation and correct language usage. Speech-to-text technologies have long struggled with accuracy due to their training data, which results in less accurate transcripts when the user has an accent that differs from the training data (Koencke et al., 2020). Hirai and Kovalyova (2024) examined accuracy rates for Japanese learners and found that the speech-to-text models they investigated struggled with incorrect syllabification and phonemic misapplication. This problem is particularly problematic for Japanese speakers who may misapply the /r/ and /l/ sounds or syllabify words with consonant clusters. However, their study was conducted in 2020 using tools developed before today's more powerful models. Modern speech-to-text APIs also benefit from the ability to create custom lexicons to aid the models' accuracy. The researchers can also see the confidence levels for each word transcribed by the software. That said, in cases where students have pronounced accents or significantly misapply complex grammar structures or vocabulary, the accuracy and usefulness of the system will be affected as it will struggle to transcribe dialogue lines accurately, leading to less effective feedback. This can be further complicated by extended periods of silence or minimal verbal interaction from shy or less proficient learners. Therefore, this application will be predominantly helpful for language learners at intermediate or advanced levels who wish to work on higher-level skills rather than acquire basic ones.

Audio Environment Sensitivity

Transcription performance is also highly dependent on the audio environment. While the system can handle a certain level of indistinct background noise, nearby conversations or loud ambient sounds can disrupt the transcription and diarization processes. For instance, a conversation taking place in a busy language practice space on campus might result in the AI capturing snippets of adjacent tables, potentially confusing these with the learner's dialogue and thus distorting the feedback. To address this, Pecha will support multiple microphone inputs in the future to allow for individual lapel microphones.

Reliability of LLM Models for Extended Conversations

The application's dependency on LLMs introduces another layer of complexity when working with extended conversations. As dialogues grow in length and the amount of transcribed text increases, the AI is more likely to provide irrelevant or inaccurate feedback. This risk is exacerbated if the conversation summary becomes overly long, which can lead to feedback errors. To mitigate this challenge, the

application attempts to truncate summaries after they reach a significant length by re-submitting them for further summarization at regular intervals. However, some conversation details can be lost in this process, so the project has limited conversation sessions to fifteen minutes.

AI Interjection Frequency

The design of the AI's interjection system to engage and participate in the conversation also presents challenges. The AI is programmed to interject when it feels appropriate within calculated interval windows, such as a maximum of once every forty-five seconds or more if it already recently participated. Due to the nature of LLMs and their tendency to fall into self-determined patterns, this can lead to a situation where the AI may incorrectly detect a pattern in its past participation behavior and start to participate only at self-determined set intervals, regardless of the content of the actual conversation. This can cause it to miss opportunities for more naturally timed contributions that could better enhance learning. For example, if a learner is determined to misunderstand something, the AI might wait too long to interject rather than provide more immediate help. To address this, the application is designed to recall only the time since the most recent interjection and forget all previous interjection timings.

Future Plans and Conclusion

Implementation Plans

The authors plan to trial Pecha within the Japanese and English language practice areas of MIT and KUIS. The application will allow learners to maximize the benefits of those interactions at both universities, where robust systems already support students in practicing their target language in real contextual situations with peers and teachers. It will focus on students' goals and learning needs and provide comprehensive post-interaction support, including advice, self-study tasks, transcripts, and conversation summaries. This reflective and iterative analysis offers instructive feedback to students, allowing them to see their progress over time in terms of proficiency. This holistic feedback helps students reflect deeply on their experiences. Given students' busy schedules, Pecha increases the effectiveness of their limited interaction time. It can function seamlessly in both face-to-face and online settings, thus expanding accessibility, but will probably be most advantageous during face-to-face interactions, where the added affordances of a shared environment and fully visible non-verbal cues enrich the interaction. Pecha will be a supportive system that builds learners' self-confidence, like an encouraging peer that assists learners on their journey towards improved communicative competence.

The focus of this ongoing project is on development of the prototype and testing its performance with Japanese and English learners across various learning contexts at both MIT and KUIS. Once sufficient data has been gathered and issues addressed, there are several potential next steps for development and expansion, including expanding support for additional languages and investigating what other forms of feedback the system can provide students both during and after their conversation sessions. Additionally, for language educators, the application could compile session summaries and offer customizable performance feedback that is also made available directly to teachers. This feedback could be collated along with other students to provide teachers insight into the performance of whole classes, providing insights into their progress and areas needing attention, allowing for targeted support and enhanced teaching efficacy.

There are few fields with such an obvious natural connection to LLMs, as the field of language learning. Though the purposes for language learning may vary, the concept of language as the medium for

transmitting and negotiating information remains key. Just how LLMs fit into fulfilling this ultimate purpose must be explored and evaluated. Following the conclusion of the prototyping phase of this project, the authors plan to conduct research evaluating the following questions:

- To what degree can an LLM fulfill the role of communicative facilitator within the informal learning space at MIT and the formal learning space at KUIS?
- Will the support provided by the LLM assist students in improving their communicative competence and acquisition of language structures?

Multiple roles of LLMs in language learning and teaching need to be explored in the field of language education so that their roles can be defined based on empirical evidence. For this to occur, implementations of LLMs must be created and applied in real situations with actual language learners. The authors hope that Pecha can provide a way to explore how LLMs can be used to augment and improve current paradigms of language practice. As more research is conducted into LLM use in language learning, it is hoped that research will unveil where this technology can really be best put to use in aiding learners and educators the most.

References

- Antoniadis, G., Granger, S., Kraif, O., Ponton, C., & Zampa, V. (2013). NLP and CALL: integration is working. *arXiv preprint*, <http://arxiv.org/abs/1302.4814>
- Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using Chatbots as AI conversational partners in language learning. *Applied Sciences*, 12(17), 8427. <https://doi.org/10.3390/app12178427>
- Boland, J. E., Fonseca, P., Mermelstein, I., & Williamson, M. (2022). Zoom disrupts the rhythm of conversation. *Journal of Experimental Psychology: General*, 151(6), 1272–1282. <https://doi.org/10.1037/xge0001150>
- Bonner, E., Lege, R., & Frazier, E. (2023). Large language model-based artificial intelligence in the language classroom: practical ideas for teaching. *Teaching English with Technology*, 23(1). <https://dx.doi.org/10.56297/BKAM1691/WIEO1749>
- Ebrahimi, S. (2015). Interaction hypothesis: An insufficient explanation for second language acquisition. *International Journal of Basic Sciences & Applied Research*, 4(6), 350–353. <http://isicenter.org/fulltext/paper-22052015123907.pdf>
- Ellis, R. (1991). The interaction hypothesis: A critical evaluation. *Regional Language Centre Seminar*, Singapore.
- Gallagher-Brett, A. (2004). *Seven hundred reasons for studying languages*. University of Southampton. https://eprints.soton.ac.uk/362669/1/700_reasons.pdf
- Gao, Y., Nuchged, B., Li, Y., & Peng, L. (2023). An investigation of applying large language models to spoken language learning. *Applied Sciences*, 14(1), 224–239. <https://doi.org/10.3390/app14010224>
- Hirai, A., & Kovalyova, A. (2024). Speech-to-text applications' accuracy in English language learners' speech transcription. *Language Learning & Technology*, 28(1), 1–21. <https://hdl.handle.net/10125/73555>
- Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., ... & Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the national academy of sciences*, 117(14), 7684–7689. <https://doi.org/10.1073/pnas.1915768117>
- Koester, A. (2022). Why face-to-face communication matters: A comparison of face-to-face and computer-mediated communication. In *COVID-19, Communication and Culture* (pp. 115–134). Routledge.

- Kooli, C. (2023). Chatbots in education and research: a critical examination of ethical implications and solutions. *Sustainability*, 15(7), 5614. <https://doi.org/10.3390/su15075614>
- Krashen, S. (1982). *Principles and practices in second language acquisition*. Pergamon.
- Kučerová, K. (2023). Benefits and challenges of conducting a Collaborative Online International Learning Class (COIL). *International Journal on Studies in Education (IJonSE)*, 5(2), 192–212. <http://dx.doi.org/10.46328/ijonse.110>
- Kwon, T. (2023). *Interfaces for personalized language learning with generative language models* [Masters Thesis, Columbia University]. <http://doi.org/10.7916/26qy-3n10>
- Liu, Z., Yin, S. X., Lee, C., & Chen, N. F. (2024). Scaffolding language learning via multi-modal tutoring systems with pedagogical instructions. *arXiv preprints*. <http://arxiv.org/abs/2404.03429>
- Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching*, 51(3), 285–329. <https://doi.org/10.1017/S0261444818000125>
- Long, M. H. (1981). Input, interaction, and second-language acquisition. *Annals of the New York Academy of Sciences*, 379(1), 259–278. <https://doi.org/10.1111/j.1749-6632.1981.tb42014.x>
- Lorentzen, A., & Bonner, E. (2023). Customizable ChatGPT AI chatbots for conversation practice. *The FLTMAG*. <https://www.doi.org/10.69732/JLOQ2431>
- Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, 6, 100199. <https://doi.org/10.1016/j.caeai.2023.100199>
- Minei, E., Razuvaeva, T., & Dyshko, D. (2021). Modern day digital pen pals: A semester-long Collaborative Online International Learning (COIL) project. *Communication Teacher*, 35(4), 336–344. <https://doi.org/10.1080/17404622.2021.1887906>
- Mynard, J., Burke, M., Hooper, D., Kushida, B., Lyon, P., Sampson, R., & Taw, P. (2020). Dynamics of a social language learning community: Beliefs, membership and identity. *Multilingual Matters*. <https://doi.org/10.21832/9781788928915>
- Namaziandost, E., & Nasri, M. (2019). A meticulous look at Long's (1981) Interaction Hypothesis: Does it have any effect on speaking skill? *Journal of Applied Linguistics and Language Research*, 6(2), 218–230. <https://www.jallr.com/index.php/JALLR/article/view/1023/pdf1023>
- O'Grady, W., & Lee, M. (2023). Natural syntax, artificial intelligence and language acquisition. *Information*, 14(7), 418. <https://doi.org/10.3390/info14070418>
- Pouromid, S. (2019). Towards multimodal interactions in the multilingual EFL classroom: Lessons from a COIL experience. *Indonesian Journal of Applied Linguistics*, 8(3). <https://doi.org/10.17509/ijal.v8i3.15262>
- Resnick, M. (2024). Generative AI and creative learning: Concerns, opportunities, and choices. *An MIT Exploration of Generative AI*. <https://doi.org/10.21428/e4baedd9.cf3e35e5>
- Tran, T.H. (2009). The interaction hypothesis: A literature review. *Education Resources Information Center*. <https://files.eric.ed.gov/fulltext/ED507194.pdf>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. <https://research.google/pubs/attention-is-all-you-need/>
- Wang, Y., & Xue, L. (2024). Using AI-driven chatbots to foster Chinese EFL students' academic engagement: An intervention study. *Computers in Human Behavior*, 108353–108353. <https://doi.org/10.1016/j.chb.2024.108353>

- Wood, D., & Moss, S. H. (2024). Evaluating the impact of students' generative AI use in educational contexts. *Journal of Research in Innovative Teaching & Learning*, 17(2), 152–167. <https://doi.org/10.1108/JRIT-06-2024-0151>
- Yu, H., & Guo, Y. (2023). Generative artificial intelligence empowers educational reform: Current status, issues, and prospects. *Generative Artificial Intelligence Empowers Educational Reform: Current Status, Issues, and Prospects*, 8. <https://doi.org/10.3389/feduc.2023.1183162>
- Zhai, C. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence*, 4, 100134. <https://doi.org/10.1016/j.caeai.2023.100134>