

## *The Effectiveness of ChatGPT in Enhancing English Language Proficiency and Reducing Second Language Anxiety (L2)*

Kotaro Hayashi, Toyohashi University of Technology, Japan  
Takeshi Sato, Tokyo University of Agriculture and Technology, Japan

WorldCALL 2023 – CALL in Critical Times  
Conference Proceedings

### **Abstract**

This study investigates the effectiveness of generative AI, specifically ChatGPT3.5, in enhancing interaction skills and reducing second language (L2) anxiety in English language teaching. The study is situated in the context of the rapid advancements in AI technology and its increasing significance in educational settings, particularly in computer-assisted language learning. The aim of this study is to assess the efficacy of ChatGPT3.5 in enhancing English conversational abilities and alleviating L2 anxiety, a psychological condition frequently linked to apprehension of failure and lack of self-assurance, particularly in Asian cultures like Japan. The study involved a 4-week experimental period with 31 Japanese university students who used a user-friendly cloud-based application integrating ChatGPT3.5. The application aimed to promote natural dialogue through speech recognition and text-to-speech. The students used the AI-assisted mobile application for 10 minutes daily. Additionally, 20 Japanese university students were included in the study. They also used the AI-assisted application for 10 minutes each day for four weeks. The study assessed the effects of ChatGPT 3.5 on L2 interaction skills and anxiety levels using an English proficiency test and questionnaire. The preliminary findings demonstrated a significant reduction in L2 anxiety and an improvement in English interaction skills. These results suggest that generative AI has the potential to transform language learning and offer innovative approaches to contemporary teaching methods. This study will contribute to the current foreign language teaching environment by exploring the role of AI in language learning, and also propose a new paradigm that has the potential to bring about changes in traditional teaching models.

Keywords: Artificial Intelligence, Chat GPT, Second Language Anxiety, Interaction, Long-Term Experiments



WorldCALL Conference 2023 in Chiang Mai, Thailand

## Introduction

The integral role of interaction in second language (L2) acquisition is well established in linguistic research (McCarty et al., 2017; Suzuki & Storch, 2020). Collaborative dialogue, as conceptualized by Swain (2000), is recognized for its efficacy in promoting L2 communicative competence, particularly through facilitating vocabulary and grammar acquisition. This concept resonates with the increasingly emphasized role of computer-mediated communication (CMC) in L2 pedagogy (Nguyen, 2011; Ortega, 2009), with Swain (2000) emphasizing the utility of CMC in fostering L2 collaborative interactions. Within this framework, peer interactions are instrumental in enhancing language proficiency and knowledge acquisition through appropriation, as posited by Wertsch (1988). Empirical studies further support the benefits of CMC in L2 learning (e.g., McCarty et al., 2017).

The emergence and proliferation of CMC, coupled with the challenges posed by the COVID-19 pandemic, have brought its importance to new heights, necessitating a shift to online learning environments and digital interaction. Despite the efforts of L2 educators to facilitate peer engagement during this period, student participation in CMC has been met with reluctance (Coman et al., 2020; Dhawan, 2020). This reluctance may be related to the relationship between L2 communication readiness and the frequency of interpersonal interactions (Yashima et al., 2004), with the pandemic's limitations on relationship building likely hindering student motivation.

Furthermore, studies by Krieg et al. (2019) and brain imaging research by Jeong et al. (2016) suggest that Japanese L2 learners often experience anxiety during L2 interactions, possibly due to perceived performance gaps, leading to negative self-evaluations. In this context, we propose that AI-driven interactions, as opposed to human interactions, may provide a more conducive environment for L2 learning. This hypothesis is supported by Nomura et al. (2019), who found that individuals with social anxiety preferred interactions with robots over unfamiliar humans.

Given the critical role of cooperative dialogue in L2 acquisition (Jeong et al., 2016; Suzuki & Storch, 2020), we introduce ChatGPT as a novel tool for L2 oral interaction that aims to facilitate communication between humans and AI. While ChatGPT is commonly used for text-based dialogues and some studies examined the potential of ChatGPT for L2 writing (Barrot, 2023; Hong, 2023; Warschauer et al., 2023), its potential for improving L2 oral proficiency remains underexplored. Our study aims to fill this gap by investigating the effectiveness of open-ended L2 interactions with ChatGPT in improving L2 fluency and comprehension. Our current study hypothesizes that the absence of negative feedback during interactions with ChatGPT may increase user comfort and engagement, as negative feedback has been shown to hinder L2 performance (Tsiplakides & Keramida, 2009). In addition, the accessibility and availability of ChatGPT, which can be used with their smartphones, provides users with more opportunities to practice L2 speaking skills and seek answers to questions without the constraints of time or location.

## Proposed Conversation Training Tool

As shown in Figure 1, the research uses an online chat tool that manifests as a web-based application hosted on a HEROKU<sup>1</sup> cloud server. This tool takes the webpage form written in HTML language (Figure 2). A Python program using Flask<sup>2</sup> (a web application framework) is awaiting at the background, exchange all information (voice, response text, and speech logs) for all APIs. The functionality of this tool depends on the integration of the Web Speech API, a standard feature available in prominent web browsers such as Chrome, Microsoft Edge, and Safari, which most of the smartphone users installed. This API facilitates the recognition and processing of the user's speech input. When user push the “Speak start” button, tool start speech recognition and write down the recognized text to the “Recognition Result” field. The user can interrupt recognition by pushing the “Speak end” button.

Within the operational framework of the application, a total of 18 logs, covering both historical and current user interactions, are systematically collected and passed to ChatGPT via the application programming interface (API). Using this data, the ChatGPT algorithm generates a textual response based on the contextual information provided by these logs. This generated text is then sent to the Google Cloud Text-to-Speech service via its API.

Google Cloud Text-to-Speech, using advanced synthetic speech technology, converts the textual response from ChatGPT into an audio output. This audio output is then sent back to our online chat tool, where it is played back to the user and, generated text is rendered on the “Answer Result” field. This cyclical process of speech input, text generation, and audio feedback forms the core interaction mechanism of the tool, providing users with a dynamic conversational experience. Meanwhile, all information is logging with the “Username” the user input on the server using Papertrail which is one of the logging add-ons for HEROKU.

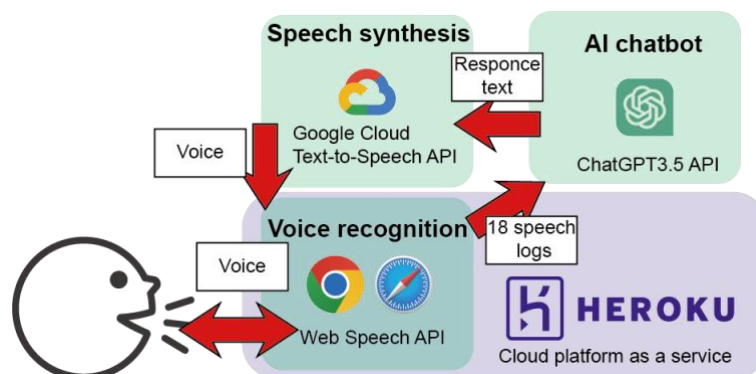


Figure 1: Conversation training tool with ChatGPT

<sup>1</sup> <https://www.heroku.com/>

<sup>2</sup> <https://flask.palletsprojects.com/en/3.0.x/>

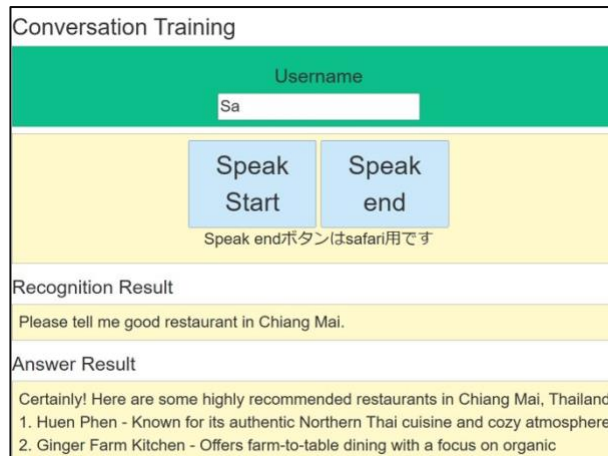


Figure 2: Conversation training tool User interface

## Research

Taking into account the discussion above, this study explores the effectiveness of the ChatGPT tool to enhance L2 speaking skills and reduce L2 anxiety during interactions by comparison with non-interactive L2 activities. Furthermore, we explore the attitude shift of the subjects after using the ChatGPT tool.

## Method

This study employed a longitudinal online experimental design. A sample of 31 undergraduate students in Japan was recruited to participate in the research. These participants were incentivized with a payment obtained through job-recruiting website (baitnet.jp), a job posting website tailored to university students in Japan. All applicants who viewed the job advertisement on this platform were considered for inclusion in the study.

For comparison purposes, participants were randomly assigned to one of two groups. The first group, consisting of 18 individuals, interacted with ChatGPT-based tools (experimental group), while the other group, consisting of 13 participants, interacted with YouTube (control group). The proficiency level of the participants was categorized as intermediate to upper intermediate, which corresponds to the B1-B2 level of the Common European Framework of Reference for Languages (CEFR).

This research was conducted exclusively in an online environment. Participants underwent pre- and post-intervention assessments using the L2 Speaking Anxiety Scale (Kumada & Okamura, 2017) to assess English speaking anxiety, as well as an online English speaking proficiency test (CASEC speaking test).

The experimental procedure consisted of the ChatGPT group using the tool for 10 minutes daily for twenty weekdays. They conducted interaction with the Chat GPT tool by asking some questions and listening to the responses. Conversely, the YouTube group was instructed to spend 10 minutes watching the YouTube programs they chose for the same weekdays as the first group. Therefore the activity for the second group did not entail any interactive activity. At the end of this intervention period, the participants of the both groups joined a semi-structured interview via Zoom to each individual, which lasted between 10 and 20 minutes. The data of their questionnaire and speaking test results were analyzed to verify

the differences between the groups. Also their interview data was also analyzed with KH conder (khcoder.net), a data-mining software.

## Findings and Discussion

A paired *t*-test revealed that the L2 speaking anxiety is reduced after treatment (ChatGPT:  $t(17) = 2.54$   $p < .05$   $d = .44$ , YouTube:  $t(12) = 2.71$   $p < .05$   $d = .46$ , Figure 3). No significant difference on the speaking test scores (ChatGPT:  $t(17) = 1.67$   $p > .05$ , YouTube:  $t(12) = 0.17$   $p > .05$ ). Two-way ANOVA for mixed design (2 groups  $\times$  pre and post) showed there is no significant difference on the interaction effect (Questionnaire:  $F(1,29) = 0.06$   $p > .05$   $\eta^2 = .00$ , Online speaking test:  $F(1,29) = 0.72$   $p > .05$   $\eta^2 = .01$ ).

The findings demonstrate that L2 interactions with ChatGPT and non-interactive activities with YouTube both yielded comparable outcomes in reducing L2 anxiety. The effect size, although rather small, is still larger than the appropriate effect size in CALL studies ( $d = 0.35$ ) according to Plonsky (2015). However, there was no significant improvement found in the L2 speaking skills between the groups. It is possible that the treatment's impact, whether it is ChatGPT or YouTube, did not differ between the groups. This may suggest that regular L2 exposure, regardless of the level of engagement, may contribute to the reduction of speaking anxiety.

Additionally, we conducted an analysis of semi-structured interviews using text-mining software. The co-occurrence network is presented in Figure 4, Table 1 displays the most frequent words used by each group. The analyses show that the ChatGPT group is characterized by active attitudes, as indicated by the use of words such as 'speak', 'use', and 'consider'. In contrast, the YouTube group is characterized by passive attitudes towards the activity, as indicated by the use of words such as 'listen' and 'watch'. These findings suggest that interactive activities through ChatGPT may promote positive learner autonomy, in contrast to the passivity observed in participants who engaged with YouTube.

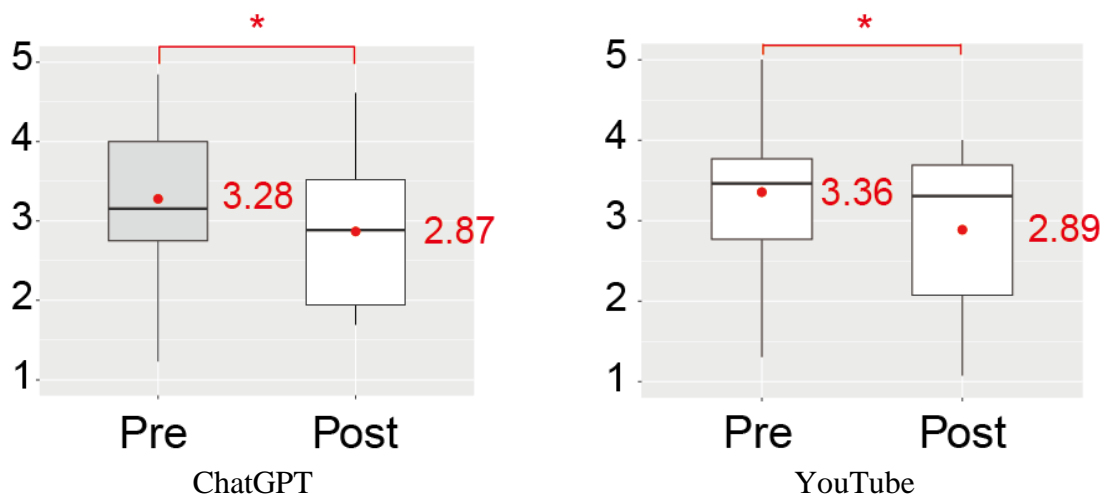


Figure 3: Results of the questionnaire surveys

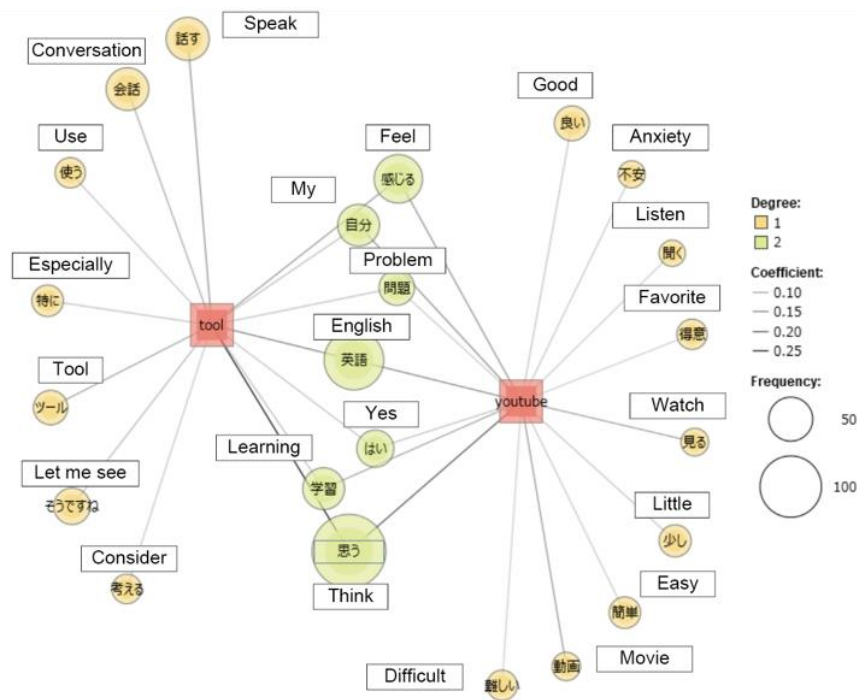


Figure 4: Co-occurrence network from interviews

Table 1: Top 10 frequent words

ChatGPT		YouTube	
think	0.282	English	0.143
speak	0.129	myself	0.134
tool	0.112	feel	0.130
conversation	0.101	video	0.124
Let me see	0.084	watch	0.115
particularly	0.072	learning	0.107
problem	0.072	yes	0.075
use	0.070	a little	0.069
consider	0.065	good	0.069
AI	0.062	anxiety	0.660

## Conclusion

This study examined the possibility of utilizing ChatGPT as an L2 interlocutor to enhance listening and speaking abilities. The hypothesis was that L2 communication with ChatGPT would decrease user apprehension and naturally encourage interaction, leading to improved L2 fluency compared to the non-interactive treatment. The data suggests that L2 speaking anxiety can be reduced through a four-week intervention, pre- and post-speaking assessments, and semi-structured interviews when L2 interacts with AI via ChatGPT. Interestingly, the control group that used YouTube showed a comparable reduction in anxiety.

Further qualitative analysis through follow-up interviews revealed attitudinal differences between the groups. The study found that participants in the YouTube group had a passive approach to L2 learning activities, while those interacting with ChatGPT had a more proactive and autonomous learning attitude.

This study proposes that L2 interaction with ChatGPT has the potential to reduce L2 anxiety, foster positive and autonomous attitudes toward L2 activities, and increase opportunities for L2 interactions. The study acknowledges its limitations, but it would provide a preliminary guide for further research on AI-based L2 learning.

One of our limitations is the potential inappropriateness of the speaking test for the participants. The high mean scores suggest a ceiling effect, and the range of the speaking test for levels A1 to B1 did not match the participants' high proficiency levels of B1 to B2. Future studies should include more challenging assessments that match the participants' actual proficiency levels. Furthermore, due to the small sample size and lack of standardization across groups, it is important to exercise caution when generalizing these results. Future studies should employ a more robust research design to address these limitations.

### **Acknowledgements**

This work was supported by JSPS KAKENHI Grant Number JP23K11284.

### **Note**

The tool we have developed can be accessed and utilized by visiting the following URL: <https://peda-chatgpt-interview.herokuapp.com/>. Additionally, the application's source code is openly available on GitHub at: <https://github.com/hayashik/chatgpt-conversation-training-tool>.

## References

- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745.
- Coman C., Țiru L.G., Meseșan-Schmitz, L., Stanciu, C., & Bularca, M.C. (2020). Online teaching and learning in higher education during the coronavirus pandemic: students' perspective. *Sustainability*, 12(24), 10367.
- Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, 49(1), 5-22.
- Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: opportunities in education and research. *Journal of Educational Technology and Innovation*, 5(1). 37-45.
- Kumada, M., & Okamura, T. (2017). Developing and validating the English speaking anxiety scale: A research for English as a subject of elementary schools in Japan. *Bulletin of Naragakuen University*, 7, 67-74.
- McCarty, S., Obari, H., & Sato, T. (2017). *Implementing mobile language learning technologies in Japan*, Singapore: Springer Education.
- Nguyen, L. V. (2011). The nature of "talk" in synchronous computer-mediated communication in a Vietnamese tertiary EFL context. *International Journal of Computer-Assisted Language Learning and Teaching*, 1(3), 14-36.
- Nomura, T., Kanda, T., Suzuki, T., & Yamada, S. (2019). Do people with social anxiety feel anxious about interacting with a robot? *AI & SOCIETY*, 35, 381-390.
- Ortega, L. (2009). Interaction and attention to form in L2 text-based computer-mediated communication. In A. Mackey & C. Polio (Eds.), *Multiple perspectives on interaction* (pp. 226-253). Routledge.
- Plonsky, L. (2015). Quantitative Considerations for Improving Replicability in CALL and Applied Linguistics. *CALICO Journal*, 32(2), 232-244.
- Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for second language writers. *Available at SSRN*.
- Wertsch, J.V. (1998). *Mind as Action*. Oxford: Oxford University Press.
- Yashima, T., Zenuk-Nishide, L., & Shimizu, K. (2004). The influence of attitudes and affect on willingness to communicate and second language communication. *Language Learning*, 54(1), 119-152.

**Contact email:** hayashik@cs.tut.ac.jp