

# Rethinking AI: bias in speech-recognition chatbots for ELT

Jaeho Jeon, Seongyong Lee, and  
Serafín M. Coronel-Molina

*Artificial intelligence (AI) technologies, particularly chatbots with speech-recognition, are gaining attention as tools for ELT. However, this frontline development in contemporary ELT seems to stand in stark contrast to the multilingual effort, another innovative trend, as chatbots' speech recognition capabilities are primarily attuned to native English speaker pronunciation. Addressing this concern, this study adopted a Global Englishes (GE) perspective and examined students' attitudes toward English after their use of a speech-recognition chatbot. Text-mining analysis of reflection papers from 38 students and thematic analysis of interviews with seven students revealed that the chatbot hindered the incorporation of diverse English use into ELT and students expressed a conception of English not as a language in flux but rather as a subject with specific content to be mastered. The findings are discussed concerning GE, AI ethics, and computer-assisted language learning, and implications for future chatbot use in ELT are provided.*

**Key words:** chatbot, Global Englishes, artificial intelligence, bias, automatic speech recognition

## Introduction

With globalization, English has become an international language of communication among speakers from diverse linguistic and cultural backgrounds, inevitably resulting in variations in English use. In response to this changing sociolinguistic landscape, scholars in the field of ELT have emphasized mutual intelligibility among users of different varieties rather than mastery of NS norms. This recognition of the legitimacy of diverse Englishes is captured in such scholarly explorations as World Englishes (WE), English as a Lingua Franca (ELF), English as an International Language (EIL), among which, Global Englishes (GE) is used as an inclusive term, challenging traditional standards of correctness

ELT Journal Volume 78/4 October 2024; <https://doi.org/10.1093/elt/ccae035>

© The Author(s) 2024. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Advance Access publication 5 September 2024

(Rose, McKinley, and Galloway 2021). The concept of GE captures how English functions as a common medium of communication in a plurilingual world and helps to address the disconnect between this practical usage and traditional instructional emphasis on native-speaker proficiency, characterized as native-speakerism in ELT curricula (Galloway and Rose 2018). In this regard, GE provides a useful framework to examining the current ELT pedagogy, as Rose, McKinley, and Galloway's (2021: 159) explained, 'it (also) encourages a critical approach, examining the impact of the dominant standard language ideology and native speakerism in learners' contexts'.

Despite the continued influence of native-speakerism in ELT (Ahn 2011), movements are underway to explore conceptual and curriculum reforms that regard language as a tool for communication, not a fixed system intolerant of deviations (e.g. Galloway and Numajiri 2020). For example, in South Korea (the setting for the current study), the National English curriculum has defined English learners as intercultural speakers who use English as an international language to negotiate meaning with speakers from different linguistic and cultural backgrounds (Fang and Baker 2018). At the classroom level, pedagogical suggestions for promoting students' GE awareness include discussions of standard language ideology (Galloway and Rose 2018), listening activities featuring different varieties of English (Galloway and Rose 2014), and engaging in conversations with speakers from diverse cultural backgrounds through videoconferencing technology (Schreiber 2021).

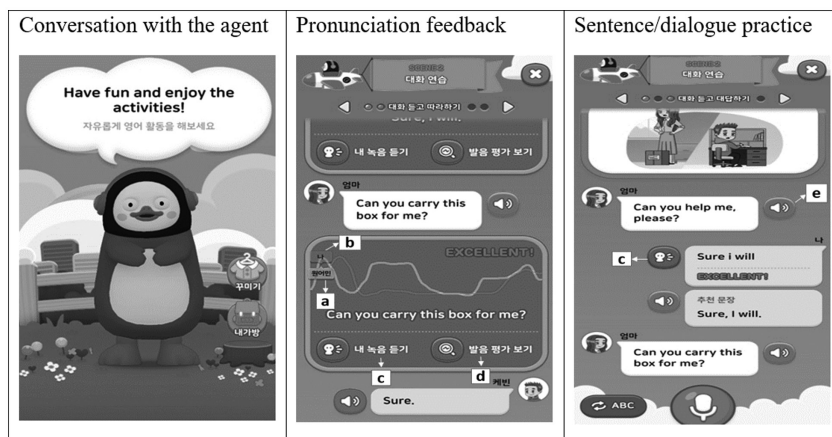
More recently, with the rapid development of artificial intelligence (AI) technologies with enhanced language capabilities, chatbots, broadly defined as virtual agents that are 'designed for verbal interaction with human users' (Lee and Jeon 2023: 443), particularly those with speech-recognition capabilities are increasingly gaining attention as tools for language learning. However, this frontline development in contemporary ELT curricula seems to stand in stark contrast to the multilingual effort, another innovative trend represented by GE. That is, despite their potential, current chatbots' speech recognition capabilities are primarily attuned to NSs' pronunciation and struggle to recognize other utterances, such as those of NNS (Lee and Jeon 2023; Martin and Wright 2023). Markl and Lai (2021) describe this discrepancy as 'predictive bias', indicating that it stems from standard NS bias in the selection of speech samples, which also raises ethical issues. As Bender *et al.* (2021: 613) argue, '... white supremacist and misogynistic, ageist, etc. views are overrepresented in the training data, not only exceeding their prevalence in the general population but also setting up models trained on these datasets to further amplify biases and harms'.

Thus, technology trained to emulate if not amplify sociolinguistic inequities reinforces the traditional privileging of certain varieties of L1 English that discredits non-native varieties spoken by L2 learners, who may eventually feel compelled to meet narrowly defined parameters of standard pronunciation. In this situation, the much-needed language practice opportunities afforded by chatbots in L2 classrooms may paradoxically undermine the promotion of the validity of diverse Englishes that is gaining traction in ELT (May 2014).

An example of a speech-recognition chatbot for L2 classrooms is PengTalk, a mobile application collaboratively developed by the Korean Ministry of Education and the Korean Education Broadcasting System (EBS) for elementary-level English instruction in Korean public schools. The pedagogical activities of the app are centred on L2 pragmatics and aligned with the Korean national standard English curriculum, enabling self-paced review or preview of the country's synchronized EFL coursework. Through the use of the application, learners can interact with a penguin character named Pengsu, who provides feedback on their English verbalizations. [Figure 1](#) displays the interface design and primary functions of the application. For example, the learner listens to an audio clip of an AI agent, illustrated either as a humanized penguin or an animated human character, and then orally responds or repeats the audio clip. After this learner–AI interaction, as shown in ‘a’ and ‘b’ in [Figure 1](#), the application uses a model of American English pronunciation as the standard to provide feedback, encouraging students to adjust their pronunciation according to the model represented by a pronunciation bar. Students can also listen to their own audio-recorded pronunciation and compare it to the native model of an audio file (see ‘c’ and ‘e’), with detailed written feedback on how to improve their pronunciation to make it sound similar to the model (see ‘d’).

In the field of computer-assisted language learning (CALL), a wealth of empirical studies and meta-analyses of research have demonstrated the benefits of chatbot technology for English language proficiency and motivation (e.g. [Bibauw et al. 2022](#)). However, given that chatbot technology predominantly relies on speech samples from standard English varieties ([Markl and Lai 2021](#); [Martin and Wright 2023](#)), most of these explorations have adhered to standard NS norms. We argue that adhering to these norms undermines GE's basic premise of the viability of diverse Englishes. Therefore, we hypothesize that the current deployment of speech-recognition chatbots may reinforce learners' attachment to these specific varieties. To examine this hypothesis, we addressed the following research question:

*How do students perceive English after their use of a speech-recognition chatbot in ELT?*



**FIGURE 1**  
App interface. Notes:  
a. native speaker;  
b. my performance;  
c. listen to my  
speech; d. detailed  
pronunciation  
feedback; e. listen  
to native speaker  
speech.

## The study

We implemented purposive sampling to recruit students aged 11 to 12 who had utilized speech-recognition chatbots for at least the one-year designated duration. A total of 38 students were recruited. These students, all users of PengTalk for a duration exceeding one year, were recruited from one public elementary school in South Korea. Upon receiving an ethical approval, we obtained formal permission to conduct the study from the school and agreement to participate from the participating learners and their parents.

We employed two methods of data collection sequentially: reflection papers and semi-structured interviews. All 38 participants completed a reflection paper responding to questions designed to elicit descriptions of their experiences with the technology in relation to their attitudes towards English (see Appendix). We analysed the papers to gain a holistic understanding of the students' experiences with the app. To obtain a deeper and more detailed understanding of these experiences, we then selected seven students for individual semi-structured interviews, based on their potential to provide rich data relevant to the research question. The interview questions focused on the elaboration of their reflection papers.

Data analysis corresponded to each method of data collection. To gain an overall understanding of the students' experiences and attitudes, we first applied sentiment analysis and word-cloud analysis to the reflection papers. We carried out sentiment analysis using VADER and SentiArt modules via Orange 3, a text-mining software. VADER analyses sentiment into four categories: positive, negative, neutral, and compound sentiment variables, while SentiArt utilizes seven categories: sentiment, anger, fear, disgust, happiness, sadness, and surprise. To obtain word-count trends that appeared in the reflection papers, we conducted word-cloud analysis using the same software.

We next conducted a thematic analysis of the interview transcripts to identify aspects of the research topic that might not have emerged in the analysis of the reflection papers. Initially, two of the authors recursively read the transcribed data to gain a general understanding. Subsequently, specific utterances regarding students' attitudes toward English and chatbot use were extracted for coding, after which codes sharing similar meanings were collated into common themes. Throughout the entire process, we took into consideration the association between students' attitudes toward English and their experiences using the chatbot. The individual analyses were compared, and disagreements were resolved through several rounds of discussions. Finally, we refined and finalized the themes to answer the research question using reconciled coding.

As shown in Table 1, our analysis of 38 reflection papers (Max = 436, Min = 65,  $M = 194.71$ ,  $SD = 93.94$ ) using VADER resulted in a higher positive sentiment score ( $M = 0.21$ ,  $SD = 0.05$ ) than the negative one ( $M = 0.08$ ,  $SD = 0.04$ ), while the compound score ( $M = 0.82$ ,  $SD = 0.45$ ), a metric of the total variance of the lexicon ratings, also indicated positivity as the value was close to 1. In sum, these results indicate the participants' positive perceptions of using the chatbot for learning English.

SentiArt analysis provided more detailed depictions of the participants' affective perceptions. Similar to the positivity identified through VADER, the sentiment score ( $M = 0.42$ ,  $SD = 0.18$ ), a compound metric of the general

## Results

### Ambivalent perceptions of English



TABLE 2  
Frequency  
counts  
of words  
mentioned  
more than 10  
times

Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency
English	196	like	119	use	82	PengTalk	80
pronuncia- tion	72	app	64	Pengsoo	64	feel	56
using	56	want	52	really	50	think	49
good	40	understand	39	speaking	39	sometimes	37
practice	35	know	35	better	35	fix	28
speak	28	makes	27	teacher	26	much	26
American	26	get	24	need	24	fun	22
well	22	feels	21	could	21	learning	21
improve	21	time	19	also	19	mistakes	19
sound	19	lot	18	confident	17	study	17
friends	17	skills	17	frustrating	17	sentences	16
feedback	16	recognize	16	even	16	say	16
keep	16	hope	16	game	16	school	16
useful	15	under- stands	15	wrong	15	words	14
right	14	helpful	14	anymore	14	scores	13
studying	13	way	13	games	12	learn	12
cute	12	however	12	make	12	cannot	11
native	11	vocabulary	11	helps	11	sure	11
experience	11	annoying	11	talking	10	saying	10
Korean	10	great	10	classroom	10	correct	10

( $n = 72$ ), 'fix' ( $n = 28$ ), 'American' ( $n = 26$ ), 'mistakes' ( $n = 19$ ), 'wrong' ( $n = 15$ ), 'right' ( $n = 14$ ), 'native' ( $n = 11$ ), 'Korean' ( $n = 10$ ), and 'correct' ( $n = 10$ ), indicating a need for a more in-depth investigation of the student perceptions behind these words.

### In-depth exploration of the possible causes for the ambivalent attitudes

Anxiety hidden  
behind confidence  
in speaking  
English

Seven students participated in individual semi-structured interviews. In line with the results for sentiment analysis (see Table 1), the interviewees all expressed their positive perceptions of using the chatbot for learning English, stating it facilitated their confidence in speaking English by enabling them to practice English in private as much as they desired. As S1 indicated, 'Now I am more confident in English than before, as I practiced a lot with the app'.

However, when commenting on the app's limitations, all the seven students expressed anxiety, unsure as to whether they could correctly pronounce English words and sentences they could not practice with the chatbot. For instance, S2 stated,

I think PengTalk is really good, and it helps me sound like an American. People will think that I am good at English. [...] However, I am also afraid that I might make some mistakes with words that I didn't practice enough in advance with the app.

In summary, although the students had generally positive perceptions of the ample practice opportunities provided by the app, which bolstered their confidence in English, using it also seemed to sharpen their consciousness of their proficiency and to encourage them to believe or strengthen their belief that they should follow a certain path to master English. This caused them to feel anxious when it was not there.



Deciding whether proper English is being used or not

Most of the students (n=6) interviewed considered receiving feedback on the correctness of their English as one of the benefits of using the chatbot. S3 stated, 'I like PengTalk because it says where I make mistakes in my pronunciation'. This kind of feedback provided by the chatbot seemed to confirm the students' belief that their Korean English accent was wrong and they should aim for American English pronunciation. Regarding this issue, S4 stated,

With the help of PengTalk, I could know whether my English is wrong or not. Before I had my pronunciation checked by the app, I did not know that I kept pronouncing "grade" wrong. It makes me find where I make mistakes and thus help me get rid of my Korean accent, which is really helpful.

On the other hand, one student expressed that, while she acknowledged it 'definitely helped me improve my English', she was confused because she received different feedback from her English teacher and the chatbot:

My English teacher understands my English well and she said that my English was okay and that foreigners could understand it, but the chatbot does not, which is honestly annoying. I am not sure whether or not it is appropriate to keep my current English.

The technology seemed to reinforce the notion that the students needed to fix their current English, which they could do by approximating the chatbot's English, making mastery of American English their ultimate goal.

Imagining target interlocutors as native speakers

When discussing their app usage experiences, all the students tended to describe their hypothetical target interlocutors as 'English speakers from America'. For example, when asked about the benefits of using the app, one student explained, 'It's like having a personal English teacher who is from America'. In addition, three students specifically mentioned that by using the app, they could ascertain whether their English could be understood by people from the United States. S5 stated, 'Pengsu understands my English, which means that my English might be understood by Americans'. Similarly, S6 mentioned, 'I like to try to talk to real Americans in English, not the penguin character, to test what I practiced with the app'.

The students' statements also revealed a tendency to divide instructional domains between their Korean English teachers and the chatbot. This point is illustrated by one student's statement that he might be able to learn vocabulary and grammar most effectively from a Korean English teacher, but 'I can improve my English-speaking skills with a native teacher, and PengTalk is similar to a native teacher'. As this statement shows, interacting with the character on the app, who spoke 'standard English' and provided feedback based on it, simulated learning English from a 'native speaker' of standard English. The interviews suggest that this limited, but consistent exposure to standard English might have led the students to believe that this was the English they would speak and listen to in their future.

**Discussion and implications**

The prevalence of NS norm-based materials, coupled with a lack of broad support for addressing it as an issue, has been consistently identified as a significant barrier to establishing a GE-based ELT curriculum. The current study reveals that chatbot technology, a frontline development

in contemporary ELT curricula, can also hinder the incorporation of diverse English use into ELT. That is, the findings suggest that this narrow presentation might have led students to conceive of English not as ‘a language in flux’ (Rose, McKinley, and Galloway 2021: 158) but rather as a subject with specific content to be mastered.

This study also shows that previous CALL research on the use of chatbot technology has typically been analysed from a standard rather than an English-as-a-global-language perspective. The students’ interviews suggest that their use of the app encouraged students to perceive English as having an ideal structure, vocabulary, and phoneme system that they must master. This distant target both seemed to bestow a sense of futility of ever being ‘good enough’ in the target language and obscure the communicative value of their current English by measuring it according to its deviation from or adherence to the idealized forms of NS English. This single-dimension judgment of their English proficiency was facilitated by the use of chatbot technology developed without considering sociolinguistic contexts and representing only ‘standard’ English (Lee and Jeon 2023). The findings of the study align with previous studies conducted within the GE framework, which identify the entrenched influence of NS norms in the ELT context (Rose, McKinley, and Galloway 2021). Furthermore, this study suggests that AI chatbot technology, as a new form of ELT materials, can potentially exacerbate this phenomenon.

Lee and Kim (2023) demonstrate that meaningful ELF interactions among Chinese students can positively influence students’ understanding of English as a global language. On a similar but different note, Ke (2010) shows that Taiwanese students who only sojourned in Anglophone countries still perceived English as a language owned by NSs, in contrast to those who engaged in NNS-NNS exchanges in NNS settings. The perceptions of the students in the current study, similar to those in Ke (2010), suggest that the limited and repetitive chatbot–student interaction, built on NS data coupled with corrective feedback, may have played a role similar to NS-NNS communication experiences in Anglophone countries, leading students to associate their English with ‘shame in not being native like’ (Fang and Baker 2018: 11) and to perceive English solely as a NS language.

This study has significant implications for the fields of GE, CALL, and AI. First, GE was adopted as a critical lens through which to examine how the use of AI technology in the CALL context influences students’ attitudes toward English. This approach represents an important preliminary step toward the integration of the premises of GE into the use of AI technology for language learning. The findings go beyond those generated from a monolithic perspective on English, providing strong support for further use of GE as a theoretical framework in future research on AI-assisted language learning.

Second, regarding existing CALL research on chatbots, which mainly explores cognitive aspects from a monolithic perspective of English, such as improvements in ‘standard’ English proficiency or in motivation to speak ‘correct’ English, this research suggests that the positive effects of using chatbots might apply mainly to cases grounded on



standard NS norms. The enhanced confidence in the correctness of their speech reported by students and their assumption that their target interlocutors are native speakers might not be compatible with the current and future reality in which English is fluidly used in plurilingual environments.

Last, this study may serve as an important reference for research on AI ethics, especially concerning policies to prevent the impact of racially, culturally, or otherwise biased programming of AI for educational uses. Building on the existing line of research, which has mostly focused on issues such as differences in the effects of varieties of English as L1 on the performance of automatic speech recognition technology (Martin and Wright 2023), this study examined conversational technology that has speech recognition functionality and addressed speakers who use English as an additional language, examining its impact on their perceptions of and attitudes toward English use.

## Conclusion

This study demonstrated how AI chatbots, which are primarily fuelled by NS data, shape and intensify students' monolingual attitude toward English in the context of CALL. This influence of chatbots may serve as a potential barrier to emergent curriculum innovation that aims to recognize and reflect diverse English use, as depicted in the GE paradigm. Despite certain limitations of the current study, which primarily result from methodology, including the reliance on self-reported data and the cross-sectional nature of the data, this research marks a significant advancement in our understanding of the intersection between GE, CALL, and AI, thereby setting the stage for future investigations within each of the fields and further interdisciplinary studies.

Given the descriptive and exploratory nature of the study, however, it should be noted that this study serves only as an initial attempt to lay the groundwork for further research with an experimental design that can rigorously confirm the causality of how the use of a similar type of application affects students' attitudes toward English. Specifically, an interesting expansion to this research will be a longitudinal examination of shifts in students' attitudes towards English that are affected by the use of a speech-recognition chatbot for learning English. In addition, it is important to conduct further research to develop chatbots that are capable of speaking different varieties of English and to examine how differently employing these chatbots affects students' perspectives towards English. Also, of significant importance is to explore methods by which teachers can effectively integrate chatbots into their classrooms in a manner that values multilingualism in ELT. Finally, we suggest that these types of AI apps include NNS models to reflect the diverse nature of English used for international communication. In order for new ELT technology to be developed and eventually introduced in a way that accurately reflects how English is used in today's globalized world—a world in which learners are likely to use English within and across local and global contexts—careful consideration of the needs and sociolinguistic reality of English language learners must be prioritized.

*Final version received July 2024*

## Funding

This work was supported by University of Nottingham Ningbo China through its open access publication funding.

## References

- Ahn, K.** 2011. 'Conceptualization of American English Native Speaker Norms: A Case Study of an English Language Classroom in South Korea.' *Asia Pacific Education Review* 12:691–702.
- Bender, E. M., T. Gebru, A. McMillan-Major, and S. Shmitchell.** 2021. 'On the dangers of stochastic parrots: Can language models be too big?' in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Canada*. ACM, pp. 610–23. <https://doi.org/10.1145/3442188.3445922>.
- Bibauw, S., T. François, W. van den Noortgate, and P. Desmet.** 2022. 'Dialogue Systems for Language Learning: A Meta-Analysis.' *Language Learning and Technology* 26(1):1–24.
- Fang, F., and W. Baker.** 2018. 'A More Inclusive Mind Towards the World': English Language Teaching and Study Abroad in China from Intercultural Citizenship and English as a Lingua Franca Perspectives.' *Language Teaching Research* 22(5):481–647.
- Galloway, N., and T. Numajiri.** 2020. 'Global Englishes Language Teaching: Bottom-up Curriculum Implementation.' *TESOL Quarterly* 54(1):118–45.
- Galloway, N., and H. Rose.** 2014. 'Using Listening Journals to Raise Awareness of Global Englishes in ELT.' *ELT Journal* 68(4):386–96.
- Galloway, N., and H. Rose.** 2018. 'Incorporating Global Englishes into the ELT Classroom.' *ELT Journal* 72(1):3–14.
- Ke, I.** 2010. 'Global English and World Culture: A Study of Taiwanese University Students' Worldviews and Conceptions of English.' *Journal of English as an International Language* 5:81–100.
- Lee, S., and J. Jeon.** 2023. 'Addressing Automatic Speech Recognition for ELT from the Global Englishes Perspective.' *ELT Journal* 77(4):435–44.
- Lee, S., and A. Kim.** 2023. 'Experiences and Practices of English as a Lingua Franca Communication in the International University: An Integrative View of Student Voices.' *Applied Linguistics Review* 14(2):271–95.
- Markl, N. and C. Lai.** 2021. 'Context-Sensitive Evaluation of Automatic Speech Recognition: Considering User Experience & Language Variation.' In *Proceedings of the First Workshop on Bridging Human-Computer Interaction and Natural Language Processing*, edited by S.L. Blodgett, Association for Computational Linguistics.

<https://aclanthology.org/2021.hcinlp-1.6> (Accessed 15 August 2022).

- Martin, J. L., and K. E. Wright.** 2023. 'Bias in Automatic Speech Recognition: The Case of African American language.' *Applied Linguistics* 44(4):613–30.
- May, S.** 2014. 'Introducing the "multilingual turn".' In *The Multilingual Turn: Implications for SLA, TESOL and Bilingual Education*, edited by S. May, 1–6. New York: Routledge.
- Rose, H., J. McKinley, and N. Galloway.** 2021. 'Global Englishes and Language Teaching: A Review of Pedagogical Research.' *Language Teaching* 54(2):157–89.
- Schreiber, B. R.** 2021. "'More Like You': Disrupting Native Speakerism Through a Multimodal Online Intercultural Exchange.' *TESOL Quarterly* 53(4):1115–138.

## The authors

**Jaeho Jeon** is a PhD candidate in the Department of Curriculum and Instruction in the School of Education at Indiana University Bloomington, US. His research interests include AI in language education, computer-assisted language learning, and dynamic assessment. His articles have been published in *Computer Assisted Language Learning*, *Computers & Education*, *TESOL Quarterly*, *Education and Information Technologies*, among others.  
**Email:** [jaehojeon21@gmail.com](mailto:jaehojeon21@gmail.com)

**Seongyong Lee** (corresponding author) is an Assistant Professor in School of Education and English at University of Nottingham Ningbo China. His research focuses on computer-assisted language learning and global Englishes. His articles have appeared in *Computer Assisted Language Learning*, *Computers & Education*, *System*, *Interactive Learning Environments*, among others.  
**Email:** [seongyonglee77@gmail.com](mailto:seongyonglee77@gmail.com)

**Serafin M. Coronel-Molina** is a Full Professor in the Department of Curriculum and Instruction in the School of Education at Indiana University Bloomington, US. His books, special issue volumes and articles on sociolinguistics, translanguaging literacies and language revitalization have been published by top international journals and publishing houses.  
**Email:** [scoronel@iu.edu](mailto:scoronel@iu.edu)

## Appendix

### Interview questions

What are some of the advantages you have found in using PengTalk for learning English? Do you like to continue using it for future English studies? Why?

What challenges have you faced while using PengTalk? Why do they interfere your English learning?

What aspects do you think could be enhanced for better learning experience in the future?

When PengTalk successfully recognizes or does not recognize your English, how do you feel?

How would you describe your overall experience with PengTalk?