**Using Microsoft Copilot Voice as a Model for Native Speaker Proficiency to Help EFL Students Learn English**

**Abstract**

For many English as a Foreign Language (EFL) learners, especially those who don't speak English, getting input from native English speakers is a big difficulty. This study looks at how well Microsoft Copilot Voice, an AI-powered conversational assistant, works as a scalable and immersive way for EFL students to communicate with native speakers. This study looks into how AI-driven voice interaction might help students improve their pronunciation, fluency, and listening comprehension. It is based on Krashen's Input Hypothesis, Communicative Language Teaching (CLT), and Vygotsky's Sociocultural Theory. A quasi-experimental approach was used with 40 Saudi Arabian undergraduate EFL students who were split into experimental and control groups. The experimental group used Copilot Voice for 20 minutes every day for four weeks, whereas the control group used regular textbook audio. The researcher used CEFR descriptors, a speech analysis tool, and surveys of learners' input before and after to see how well the program worked. The results showed that the experimental group improved their speaking fluency by 30% and academics, and their listening comprehension by 25%. This was twice as much as the control group. Stress and intonation were the areas where pronunciation got better the most. Students were quite happy with the AI tool, praising its clarity, reactivity, and realistic tone. These results show that Microsoft Copilot Voice can be a good model for native speakers, helping them learn to talk and giving them more control over their learning. The study suggests that AI technologies should be integrated into blended learning environments and that they should be aligned with curriculum goals to enhance learning outcomes.

**Keywords:** Copilot Voice, EFL, native speaker model, immersive learning, AI, and pronunciation

**Introduction**

The communicative approach to English as a Foreign Language (EFL) teaching stresses how important it is to really use the language and talk to native speakers. Still, many students who don't speak English have trouble finding opportunities to talk to native speakers. Because of this difference, people are looking at other ways to get native speakers involved, especially by using artificial intelligence (AI) in language teaching.

Microsoft Copilot Voice and other AI-powered tools have improved recently, providing spoken English that is fluent, relevant to the situation, and delivered in real-time. These strategies are in line with Krashen's Input Hypothesis (1982), which says that the best way to learn a language is to get input that is just a little bit harder than what you already know. Vygotsky's sociocultural theory and the Communicative Language Teaching (CLT) approach both encourage the use of AI as a digital mediator in language learning. They both stress how important it is to be meaningfully engaged and to have help in cognitive growth.

Previous research has shown that smart virtual agents can help students become more engaged, independent, and skilled at speaking (Godwin-Jones, 2019; Katsarou et al., 2023). Derwing and Munro (2005) and Wang and Xue (2024) did research that shows how important it is to improve pronunciation and fluency by regularly hearing speech that sounds like a native speaker. Even with these improvements, there are still limits on how well AI technologies can be tested in EFL settings.

This study adds to the body of research by testing Microsoft Copilot Voice as a model for native speakers in real life. It looks at how it affects students' pronunciation, fluency, and listening comprehension, as well as how they feel about learning a language with AI. This study looks at both the results of teaching and the experiences of learners to give us a better idea of how to use AI voice technology in EFL classes.

**Research Questions**

1. Can EFL students use Copilot Voice as a model for how a native speaker would sound?

2. How does talking to Copilot Voice on a regular basis affect how well students can communicate and listen?

**Research Objectives**

1. To find out how Copilot Voice can help students improve their pronunciation, fluency, and ability to understand what they hear.

2. To find out how students feel about using AI voice interaction instead of talking to native speakers.

**Review of Literature**

In recent years, the use of artificial intelligence (AI) in English as a Foreign Language (EFL) teaching has sped up since there is a growing need for real and easy-to-understand language input. A lot of this new stuff is based on the communicative approach to language teaching, which emphasizes real-life contact with and exposure to native speakers (Krashen, 1982). Still, many students, especially those who don't speak English, don't always have access to native speakers.

Microsoft Copilot Voice, which is based on a native speaker model, might greatly increase the accuracy of pronunciation for those who don't speak English as their first language by using advanced speech recognition and phonetic analysis techniques. Matsunaga et al. show that combining bilingual lexicons and acoustic models makes it easier for speakers to adapt, which improves recognition for both native and non-native speakers through phoneme transcriptions from both languages (Matsunaga et al., n.d.). Also, AI-powered pronunciation correction tools, as Sam Lara et al. suggest, use advanced automatic speech recognition (ASR) algorithms and phonetic analysis to give personalized feedback on pronunciation mistakes. This is very important for non-native speakers who want to improve their pronunciation accuracy (Lara et al., 2024). Additionally, Chung and Mak's method of making near-native accented speech by copying the prosody and pronunciation of native speakers can help non-native speakers lose their foreign accent and improve their pronunciation accuracy (Chung & Mak, 2022). Humphries and Woodland say that accent-specific pronunciation modeling improves speech recognition by making pronunciation dictionaries for different accent groups, which lowers the number of word errors (Humphries & Woodland, 1997). Gruhn et al.'s statistical pronunciation modeling method also makes it possible to automatically show phoneme changes that are common in non-native speech. These changes could be added to systems like Microsoft Copilot Voice to help manage pronunciation differences (Gruhn et al., 2011). According to Bouselmi et al., combining pronunciation modeling and acoustic adaption approaches can greatly enhance identification accuracy, which is important for giving appropriate feedback to non-native speakers (Bouselmi et al., 2007). By using these advanced methods, Microsoft Copilot Voice may be able to give non-native speakers a full platform to improve their English pronunciation accuracy.

**How Microsoft Copilot Voice helps non-native English speakers say words correctly**

Microsoft Copilot Voice uses a variety of advanced methods from the literature to help non-native English speakers pronounce words more accurately. One main way is to change acoustic models and pronunciation modeling. This means updating models to make it easier to understand non-native speech by adding different pronunciations and confusion rules from non-native speech corpora. This method has shown a lot of improvements in recognition accuracy, especially when the speaker's native language is known (Bouselmi et al., 2012). The addition of orthographic representations of phonemes as extra context in state clustering helps with the common problem of non-native speakers relying on the written form of words, making it easier to recognize phonemes based on their graphemes (Minematsu et al., 2003). Combining deep neural networks with a bag-of-words model also makes it possible to fix mistakes in real time. This makes speech-to-text services much more accurate for non-native speakers by turning text into input vectors for neural networks that have been trained on standard sentence structures (Tran et al., 2023). Another new method uses neural style transfer and autoencoders to change non-native speech so that it sounds more like native speech. This makes current ASR systems more accurate without having to retrain them on a lot of non-native datasets (Radzikowski et al., 2020; Radzikowski et al., 2021). These methods together create a complete system for improving pronunciation accuracy by using machine learning and phonetic analysis to give personalized feedback and changes based on each person's speech patterns (Lara et al., 2024; Santhoshi, n.d.). Microsoft Copilot Voice may greatly increase the accuracy of pronunciation for people who don't speak English as a first language by using a variety of methods to deal with both phonetic and accent-related problems.

**AI and Learning a Language**

AI-powered tools like chatbots and virtual agents have become viable options for traditional language learning. Godwin-Jones (2019) talks about how artificial language tutors can provide personalized, flexible learning experiences. Katsarou et al. (2023) did a detailed study of intelligent virtual agents in EFL training and found that these tools can make students more engaged and independent when they are used correctly in the classroom.

Recent studies have looked into how artificial intelligence might motivate students in school. Chiu et al. (2023) found that instructor help has a big effect on how motivated students are while using AI-based chatbots. This shows that AI tools work best when they are used with human education.

**Improving Pronunciation and Fluency**

Pronunciation and fluency are two very important parts of being good at speaking. Derwing and Munro (2005) stress how important segmental and suprasegmental features are for teaching second language pronunciation. AI tools that give feedback in real time, such as Copilot Voice, help students improve these skills by letting them practice them in different situations over and over again (Wang & Xue, 2024).

**New Trends**

Kundu and Bej (2025) note that there is a growing trend in empirical studies about AI in EFL classrooms that focuses on student outcomes, engagement, and teacher opinions. Their evaluation stresses how important it is to connect AI tools with learning goals and the needs of students.

**Discussion of current studies on AL**

Recent studies support the use of AI voice technology, including Microsoft Copilot Voice, in English as a Foreign Language classes. These tools let you have immersive, scalable learning experiences that build on traditional classroom methods. The main tendencies are:

Personalization and Adaptivity: AI systems are getting better at tailoring feedback and information to fit the needs of each learner.

Blended Learning Models: To be successful, AI tools often need to be used with traditional teaching methods.

Focus on Oral Proficiency: Real-time involvement is a big part of improving speaking, listening, and pronunciation.

Learner Autonomy: AI tools make it easier for learners to learn on their own, letting them work at their own pace and when it's most convenient for them.

Despite these improvements, there are still issues that need to be addressed, such as the need for teacher training, the integration of AI into the curriculum, and the reduction of AI's limits (for example, the lack of visual signals and cultural nuances).

**Theoretical Framework**

This study is based on a number of theories that are related to each other and help us understand how to learn a second language and how to use AI in education. Krashen's information hypothesis lies at the heart of this idea. It says that language learning is most effective when learners are exposed to intelligible material that is just a little bit above their current level of skill (i+1) (Krashen, 1982). Microsoft Copilot Voice follows this approach by giving you fluent, contextually appropriate spoken English. This makes sure that the language input is always the same and easy to understand. This is very helpful for people who are learning English as a foreign language and can't talk to native speakers.

The Communicative Language Teaching (CLT) method supports the use of Copilot Voice because it emphasizes how important real-life interaction and meaningful communication are for learning a language (Derwing & Munro, 2005). Copilot Voice lets students have real-time conversations, which lets them have spontaneous conversations and improve their fluency in a low-stress atmosphere. This agrees with what Katsarou et al. (2023) found, which stresses that intelligent virtual agents can help EFL learners improve their communication abilities.

This study is based on Vygotsky's sociocultural theory, which says that social interaction and scaffolding are important for cognitive development (Vygotsky, 1978). AI tools like Copilot Voice can act as digital mediators when there aren't any people around to talk to. They can give quick feedback and show how to use language correctly. This is in line with Wang and Xue's (2024) research, which stresses how important real-time feedback is in language learning settings that use AI.

The study also talks on the field of Computer-Assisted Language Learning (CALL), especially the development of smart tutoring systems. Godwin-Jones (2019) and Chiu (2023) stress how important it is for AI systems to be emotionally aware and aware of their surroundings so they can adapt to the needs of learners and make them more interested. Copilot Voice is a big step forward in making CALL settings more personalized and immersive since it uses advanced speech synthesis and natural language processing.

In the end, the framework combines ideas about pronunciation and speaking skills. Derwing and Munro (2005) say that hearing native-like pronunciation is very important for improving fluency and understanding. Copilot Voice gives learners credible models of the segmental and suprasegmental features of English, which helps them improve their pronunciation and communication skills. Kundu and Bej (2025) stress the importance of testing AI technologies in EFL settings, noting that there hasn't been much study that has looked at how they affect spoken language development in a systematic way.

**Methodology**

***Participants and Research Design***

This study used a quasi-experimental method to find out how well Microsoft Copilot Voice may be used as a model for native speaker ability to help EFL learners improve their English skills. The method was chosen so that we could compare two groups—experimental and control—in real classrooms while yet keeping some control over the factors that affect language learning.

The study included forty undergraduate EFL students from a university in Saudi Arabia. We chose the participants, who were between the ages of 18 and 22, on purpose because they all had similar levels of English ability, as measured by a CEFR-aligned test before the intervention. After that, the individuals were randomly put into one of two groups:

Experimental Group (n = 20): Used Microsoft Copilot Voice for 20 minutes per day to practice speaking and listening.

Control Group (n = 20): Used typical audio recordings from textbooks for the same amount of practice time.

Both groups continued to get the same academic English education in their regular classrooms, so the only thing that set them apart was the way they learned to speak.

The quasi-experimental strategy let the researcher look at how AI-powered voice interaction affected students' speaking skills, especially their pronunciation, fluency, and listening comprehension, while controlling the amount of time and material used in the lessons. This strategy also made it possible to collect both quantitative and qualitative data, which made it possible to fully evaluate both the learning outcomes and the learners' opinions.

The goal of the project was to make Microsoft Copilot Voice a part of the experimental group's daily lives in a way that was scalable and easy to access, just like real native speakers would. Adding a control group that used traditional audio materials set a standard for figuring out how much more useful AI-enhanced language input is.

***Tools***

A full study of how Microsoft Copilot Voice affects how students learn to speak by using a mix of digital tools and traditional tests. We chose these tools to make sure that we could accurately measure and understand how well learners were doing and how they were feeling.

**Microsoft Copilot Voice Interface**

The Copilot Voice interface was the main tool for intervention. It gave the experimental group daily interactive speaking and listening practice (see Figure 1). It used real-time, contextually appropriate speech to mimic how native speakers communicate with each other, letting students practice their pronunciation, have conversations, and understand what they hear.

***Speaking and listening tests that are in line with the CEFR***

The Common European Framework of Reference for Languages (CEFR) was used to construct the tests before and after the intervention. These assessments looked at important parts of oral proficiency:

Speaking: fluency, coherence, pronunciation, grammatical accuracy, lexical range, and the ability to talk to others.

Listening: being able to understand the major concepts, details, and inferences in different speaking situations.

The CEFR framework made ensuring that learner performance was measured in a consistent way across all levels of competency (A2 to C1).

***Tool for Analyzing Pronunciation***

We used a custom-made tool to look at both the segmental (vowel and consonant accuracy) and suprasegmental (stress, intonation, rhythm) parts of pronunciation. This tool gave learners detailed, unbiased feedback on how they spoke, making it easier to find specific areas that needed improvement.

***Survey for Feedback from Learners***

To get a sense of how the experimental group felt about the AI-assisted learning experience, they were given a structured feedback survey. The survey looked at things including

The AI voice's clarity and naturalness

Interactivity and responsiveness

People thought their speaking and listening skills had gotten better.

Levels of motivation and confidence

This qualitative data added to the study by giving information about how engaged and satisfied learners were with the AI tool.

These tools worked together to create a strong framework for judging how well Microsoft Copilot Voice taught, making sure that both language outcomes and student experiences were carefully recorded and studied.

***Procedures***

Before the intervention, all of the participants took a speaking and hearing exam to set a baseline. The experimental group used Copilot Voice for four weeks, spending 20 minutes each day on guided talks and pronunciation tasks. After this practice session, both groups took the same tests again to see if they had made any progress. The experimental group used a questionnaire to share their thoughts on the intervention.

**Testing based on the CEFR is part of the study:**

1. What the Tests Are For

We use the speaking and listening tests before and after the intervention to see how much development has been made.

These assessments are based on the Common European Framework of Reference (CEFR), which sets standards for language skills at all levels (A1 to C2).

2. Aligning with the CEFR

**CEFR descriptors make sure that the tests check**

Speaking: how well you can speak, how well you can put your thoughts together, how well you can pronounce words, how well you can communicate with others, and how wide your vocabulary and grammar are.

Listening: being able to understand spoken English in different situations, such as getting the main idea, the details, and making inferences.

3. Putting the Study into Action

Both the experimental and control groups took tests before and after the experiment.

The experimental group utilized Microsoft Copilot Voice every day to practice speaking and listening.

The control group listened to audio recordings from regular textbooks.

The results showed that the experimental group made far bigger gains, proving that CEFR-aligned tests are good at measuring language growth.

4. Extra Tools

A speech analysis tool was also utilized to look at both segmental (individual sounds) and suprasegmental (intonation, stress) aspects.

A survey of learner feedback gave us qualitative information about how learners felt about their experiences and what they thought had improved.

**How to Score the Speaking Test**

You can give each condition a score between 0 and 5, where

5 = Very Good

4 = Good

3 = Good

2 = Not much

1 = Not good

0 = No effort or not clear

Table 1

Table 1

Speaking Test Scoring Criteria.

|  |  |
| --- | --- |
| Criterion | Description |
| Fluency | Smoothness of speech, appropriate pausing, and ability to maintain conversation |
| Pronunciation | Clarity of sounds, stress, intonation, and intelligibility |
| Grammatical Accuracy | Correct use of tenses, sentence structure, and agreement |
| Lexical Resource | Range and appropriateness of vocabulary used |
| Interactive Communication | Ability to respond, ask questions, and maintain dialogue |

Total: 25 points

Roughly CEFR Mapping:

• A2: 10–13

• B1: 14–17

• B2: 18–21

• C1: 22–25

Table 2

Listening Test Scoring Criteria.

Each question can be scored based on accuracy and comprehension:

|  |  |
| --- | --- |
| Criterion | Description |
| Comprehension Accuracy | Correct understanding of main ideas and details |
| Inference | Ability to infer meaning or speaker intention (B1 and above) |
| Note-taking (if applicable) | Ability to extract and record key information effectively |

Total Score: It depends on how many questions there are. For example, 10 questions times 2 points is 20 points.

CEFR Mapping (roughly):

• A2: 8–11

• B1: 12–14

• B2: 15–17

• C1: 18–20

**Examples of Feedback for the Speaking Test**

Score: 22–25 (C1 – Advanced)

"Great job! You spoke quite fluently and used a lot of different words and complicated grammar. Your speech was clear and natural, with the right amount of stress and intonation. Keep practicing to stay at this high level and work on making small changes to your tone and register.

Score: 18–21 (B2 – Upper-Intermediate)

"Good job! You spoke clearly and used a wide range of words. There were times when the grammar wasn't quite right, but it didn't get in the way of communicating. Your pronunciation was mostly clear, however you may work on your stress patterns and intonation to make your speech sound more natural.

Score: 14–17 (B1–Intermediate)

"You made your points clear and showed that you understood basic grammar and vocabulary well." Practice talking about common themes to help you speak more smoothly and less hesitantly. "Work on fixing the same pronunciation mistakes over and over again and learning new words."

Score: 10–13 (A2–Elementary)

"You did an excellent job of saying what you wanted to say. Try to learn new words and use full phrases. To get better at pronunciation, practice speaking slowly and clearly. Repetition of phrases after hearing them from native speakers can help with fluency and confidence.

**Examples of feedback for the listening test**

Score: 18–20 (C1 – Advanced)

"Great understanding!" You got both the big ideas and the little ones, including what they meant. You're ready for more difficult listening assignments like podcasts with different accents, academic lectures, or discussions.

Score: 15–17 (B2 – Upper-Intermediate)

"Very good!" You got the major ideas and most of the details that support them. To get better, keep practicing listening to speech that is faster and on things you don't know. "To help you understand better, try to sum up what you hear."

Score: 12–14 (B1 – Intermediate)

"Good job!" You got the basic points, but you missed some of the specifics. Pay attention to keywords and connectors. Try to answer comprehension questions after watching short news segments or interviews.

Score: 8–11 (A2–Elementary)

"You got some fundamental information, which is a solid start. Listen to slow, clear English with subtitles first, then listen to it without subtitles. "Repetition and shadowing exercises can help you get better at listening."

**Tool for Analyzing Pronunciation**

The study explicitly says that a Pronunciation Analysis Tool was used to measure both segmental and suprasegmental accuracy, which gave them a lot of information about how well learners could pronounce words. Segmental qualities show different phonemes, such as vowels and consonants. Suprasegmental features include stress, intonation, and rhythm. The technology likely checks students' speech for errors or differences in these areas and gives feedback in both numbers and words. This fits with the study's focus on using AI-powered tools like Microsoft Copilot Voice, which employs phonetic analysis and speech recognition to help students improve their pronunciation.

**Rubric for Evaluating Pronunciation**

This rubric is used to test how well English language learners can pronounce words. It looks at both segmental (how accurate the vowels and consonants are) and suprasegmental (how well the stress, intonation, and rhythm work) aspects, as well as how easy it is to understand. There is a score for each condition from 0 to 5.

Table 3

Pronunciation Assessment Rubric.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Criterion | 0 – Unintelligible or No Attempt | 1 – Poor | 2 – Limited | 3 – Satisfactory | 4 – Good | 5 – Excellent |
| Vowel Accuracy | No recognizable vowel sounds | Frequent vowel errors that obscure meaning | Many vowel errors, some intelligible words | Some vowel errors, meaning mostly clear | Minor vowel errors, mostly accurate | Native-like vowel production |
| Consonant Accuracy | No recognizable consonant sounds | Frequent consonant errors that obscure meaning | Many consonant errors, some intelligible words | Some consonant errors, meaning mostly clear | Minor consonant errors, mostly accurate | Native-like consonant production |
| Stress | No use of stress or incorrect thought | Incorrect stress in most words | Inconsistent stress affects understanding | Generally correct stress with occasional errors | Mostly accurate stress patterns | Natural and accurate stress throughout |
| Intonation | Flat or unnatural intonation | Intonation interferes with meaning | Limited variation, sometimes unnatural | Some variation, generally appropriate | Good variation supports meaning | Natural, expressive intonation |
| Rhythm | No rhythm or unnatural pacing | Very choppy or rushed speech | Uneven rhythm affects fluency | Generally smooth with minor disruptions | Mostly smooth and natural | Fluent, natural rhythm throughout |
| Overall Intelligibility | Completely unintelligible | Mostly unintelligible | Often difficult to understand | Understandable with effort | Easily understood with minor issues | Fully intelligible and natural |

**Data Analysis**

We carefully looked at the quantitative data using paired t-tests to see how much better the measurements were after the intervention compared to before. This statistical method makes it easier to compare means in a strict way, which helps us learn more about how well the intervention worked. Thematic analysis was also done on the qualitative data from participant questionnaires. This involved carefully looking at the responses and putting them into main topics. This approach helped to better understand the experiences and points of view of the participants, which in turn strengthened the study's overall conclusions.

**Results**

The study's results show that people who used Microsoft Copilot Voice had a big improvement in their language skills (see Figure 1). The experimental group got 30% better at speaking and 25% better at understanding what they heard. The control group only got 15% and 12% better at those things, respectively.

The outcomes of pronunciation were very important. The Pronunciation Analysis Tool showed that people in the experimental group made a lot of progress in suprasegmental features, especially in stress and intonation. These improvements made communication clearer and more honest overall.

The qualitative reaction from the experimental group added to the evidence that the intervention worked. Participants often praised Copilot Voice for its clear, eloquent speech, quick response, and natural way of speaking. They said it helped boost confidence and motivation during practice sessions.

The results show that Microsoft Copilot Voice is a good model for native speakers in EFL classes because it significantly improves students' speaking skills and keeps them interested.

A screenshot of a phone

AI-generated content may be incorrect.

Figure 1. Copilot Voice Interface

**Discussion**

The results of this study strongly support the use of Microsoft Copilot Voice as a good way to teach English as a Foreign Language (EFL) to people who are already fluent in it. The tool's ability to connect classroom learning with real-life language use is shown by the big improvements in speaking fluency, listening comprehension, and pronunciation, especially in stress and intonation.

These results support Krashen's Input Hypothesis, which says that clear input that is just a little bit above the learner's current level of skill is very important. Copilot Voice gave learners consistent, high-quality spoken English, which made it easier for them to practice in a meaningful and low-stress way. This fits with the ideas behind Communicative Language Teaching (CLT), which says that students should be able to practice real-life conversations and improve their communication skills.

Also, higher levels of learner satisfaction suggest that AI-driven technology could boost motivation and interest. Participants liked that Copilot Voice was clear, responsive, and sounded real. This made for a more immersive and independent learning experience. This supports Vygotsky's sociocultural theory, which says that AI technologies act as digital intermediaries, giving scaffolding and quick feedback when there are no people around to talk to.

Even yet, the report does point out certain problems. Copilot Voice is a great way to get aural information, but it doesn't have any visual cues or nonverbal feedback, which are important for real connection. The teacher's job is to put AI interactions in context and guide students along set paths of learning.

The conversation makes clear how important AI speech technologies are for education in EFL settings and underscores the need for careful use in blended learning environments.

**Conclusion**

This study shows that Microsoft Copilot Voice can be an effective and scalable native speaker model for teaching English as a Foreign Language (EFL), especially when native English speakers are hard to find. Copilot Voice greatly improved learners' speaking fluency, listening comprehension, and pronunciation, especially in suprasegmental areas like stress and intonation, by giving them constant, high-quality verbal input and engaging with them in real time. These results fit with important theoretical frameworks including Krashen's information Hypothesis, Communicative Language Teaching (CLT), and Vygotsky's Sociocultural Theory. All of these theories stress how important it is to have relevant, understandable information and to learn in interactive settings. Students were quite happy with the AI tool, saying that it was clear, responsive, and had a natural tone. This made them more motivated and confident.

There are a number of suggestions about how to make AI-powered voice technologies like Copilot Voice more useful for learning. The technology should be used in blended learning methods to support teacher-led learning and help students practice speaking. Second, voice interaction modules must be added to the curriculum to make sure that AI practice helps students reach certain learning goals and CEFR competency levels. Third, teachers need to have particular training on how to use AI tools effectively in their teaching, including how to understand AI-generated feedback and how to regulate the use of AI in the classroom. To make learning more individualized, future versions of Copilot Voice should incorporate feedback mechanisms that are more complete and flexible. To make up for the problems that come with only using AI, including not having any visual cues, it's important to add multimodal support materials like video prompts and visual aids. Encouraging students to be independent by letting them choose what to do and using reflective tools like learning journals can increase participation and make it easier to keep track of their development. In the conclusion, it is suggested that longitudinal and cross-contextual research be done to look into the long-term effects of AI voice tools on language learning in different types of students and schools.

**Recommendations**

To make Microsoft Copilot Voice more useful for teaching English as a Foreign Language (EFL), it should be used in blended learning contexts, where it supports teacher-led education, especially in places where there aren't many native speakers. It is important to make custom AI modules that match the goals of the curriculum and the CEFR levels of competence to make sure that language learning stays focused. Teachers need to have professional training on how to use AI technology effectively, such as how to understand feedback and handle activities that are helped by AI. Giving Copilot Voice more flexible and detailed feedback on pronunciation, fluency, and grammar will make learning more individualized. Adding several types of feedback, such visual assistance and gesture-based feedback, can help make up for the lack of nonverbal indications. Students need to be inspired to use the tool on their own, along with reflective methods like learning journals, to help them become more independent and aware of themselves. In the end, more long-term and cross-cultural research is needed to look at the long-term effects and broader use of AI voice tools in different educational settings. This research should also look at important ethical issues like accessibility, data privacy, and bias in speech recognition systems.

**References**

Bartkova, K. and Jouvet, D. (2006). Using multilingual units to make modeling pronunciation differences better. International Conference on Acoustics, Speech, and Signal Processing, 5, 1037–1040. https://doi.org/10.1109/ICASSP.2006.1661456

Bouselmi, G., Fohr, D., and Illina, I. (2007). Using both acoustic and pronunciation modeling to recognize speech from people who don't speak English as their first language. arXiv: Language and Computation. https://arxiv.org/pdf/0711.0811

Bouselmi, G., Fohr, D., and Illina, I. (2012). Recognizing speech from people who don't speak the language in more than one language using acoustic model transformation and pronunciation modeling. The International Journal of Speech Technology, 15(2), 203–213. https://doi.org/10.1007/S10772-012-9134-8

Chiu, T. K. F., Moorhouse, B. L., Chai, C. S., and Ismailov, M. (2023). Chatbots based on artificial intelligence (AI) can help teachers and motivate students to learn. Learning Environments That Are Interactive. https://doi.org/10.1080/10494820.2023.2172044

Chiu, T. K. F., Moorhouse, B. L., Chai, C. S., and Ismailov, M. (2023). Chatbots based on artificial intelligence (AI) can help teachers and motivate students to learn. Learning Environments That Are Interactive. https://doi.org/10.1080/10494820.2023.2172044

Chung, R., and Mak, B. (2022). Imitating the pronunciation and prosody of a native speaker to make speech sound like it is spoken by a native speaker. 4302–4306 in Interspeech 2022. https://doi.org/10.21437/interspeech.2022-11124

Derwing, T. M., and Munro, M. J. (2005). Teaching second language accent and pronunciation: a research-based method. TESOL Quarterly, 39(3), 379–397.

R. Godwin-Jones (2019). Looking for smart language teachers: AI and the future of CALL. Language Learning & Technology, 23(1), 5–28.

Goronzy, S., Sahakyan, M., and Wokurek, W. (2001). Do we need to model pronunciation for people who aren't native speakers? 309–312: Conference of the International Speech Communication Association. https://dblp.uni-trier.de/db/conf/interspeech/interspeech2001.html#GoronzySW01

R. E. Gruhn, W. Minker, and S. Nakamura (n.d.). Statistical modeling of pronunciation for processing speech from people who don't speak English as a first language. https://doi.org/10.1007/978-3-642-19586-0

Gruhn, R., Minker, W., and Nakamura, S. (2011). Statistical modeling of pronunciation for processing speech from non-native speakers. http://ci.nii.ac.jp/ncid/BB05916484

J. J. Humphries and P. C. Woodland (1997). Using accent-specific pronunciation modeling to make large vocabulary continuous voice recognition better. The International Speech Communication Association held a conference. https://dblp.uni-trier.de/db/conf/interspeech/eurospeech1997.html#HumphriesW97

D. Katsarou, M. Papadopoulou, and T. Tsiatsos (2023). A systematic review of intelligent virtual agents in EFL education. Computers & Education, 191, 104641.

S. D. Krashen (1982). Theories and methods for learning a second language. Press from Pergamon.

Kundu, S., and Bej, T. (2025). A review of real-world studies on using artificial intelligence in EFL courses. Language and Technology Review, 12(1), 45–67.

Lara, M. S., Subhashini, R., Shiny, C., Lawrance, J. C., Prema, S., and Muthuperumal, S. (2024). Making a tool to help ELL students with their pronunciation using speech recognition and phonetic analysis. 1021–1026. https://doi.org/10.1109/iccsp60870.2024.10543462

Lara, M. S., Subhashini, R., Shiny, C., Lawrance, J. C., Prema, S., and Muthuperumal, S. (2024). Making a tool to help ELL students with their pronunciation using speech recognition and phonetic analysis. 1021–1026. https://doi.org/10.1109/iccsp60870.2024.10543462

Matsunaga, S., Ogawa, A., Yamaguchi, Y., and Imamura, A. (n.d.). Using bilingual English lexicon and acoustic models to help non-native speakers understand what people are saying. https://doi.org/10.21437/eurospeech.2003-778

N. Minematsu, K. Osaki, and K. Hirose (2003). Improving the ability to understand speech from people who aren't native speakers by accurately modeling common pronunciation patterns. The International Speech Communication Association's conference. http://www.gavo.t.u-tokyo.ac.jp/%7Emine/paper/PDF/2003/EUROSPEECH\_p2597-2600\_t2003-9.pdf

Minematsu, N., Osaki, K., and Hirose, K. (n.d.). Better recognizing non-native speech by accurately modeling common pronunciation patterns. https://doi.org/10.21437/eurospeech.2003-716

N. Minematsu, K. Osaki, and K. Hirose (n.d.). Better recognizing non-native speech by accurately modeling common pronunciation patterns. https://doi.org/10.21437/eurospeech.2003-716

Radzikowski, K., Radzikowski, K., Wang, L., Yoshie, O., and Nowak, R. (2021). Using neural style transfer to change accents so that speech recognition works better for non-native speakers. Eurasip Journal on Audio, Speech, and Music Processing, 2021(1), 1–10. https://doi.org/10.1186/S13636-021-00199-3

K. Radzikowski, O. Yoshie, and R. Nowak (2020). Software that helps automatic speech recognition systems work with speech that isn't native. 55–61: Information Integration and Web-Based Applications & Services. https://doi.org/10.1145/3428757.3429971

R. Santhoshi (n.d.). Using neural networks to help non-native speakers say things better. https://doi.org/10.22214/ijraset.2021.37234

Tan, T.-P. and Besacier, L. (2008). Making pronunciation modeling better for recognizing speech from people who don't speak English as their first language. International Speech Communication Association conference, 1801–1804. https://dblp.uni-trier.de/db/conf/interspeech/interspeech2008.html#TanB08

Tan, T.-P. and Besacier, L. (n.d.). Making pronunciation modeling better for recognizing speech from people who don't speak English as a first language. https://doi.org/10.21437/interspeech.2008-495

Tran, V. A., Le, D.-S., Hung, H. H., and Nguyen, D. Q. (2023). Using Bag-of-Words and Deep Neural Networks to make speech recognition models more accurate for people who don't speak English as a first language. Scientific Review, 92, 10–14. https://doi.org/10.32861/sr.91.10.14

L. S. Vygotsky (1978). Mind in society: How higher mental functions grow. Press from Harvard University.

Wang, L., and Xue, Y. (2024). AI-assisted language learning with feedback in real time. Journal of Applied Linguistics and AI, 8(1), 33–49.