INNOVATION IN LANGUAGE LEARNING International Conference

Effective Ways of Integrating Speech Recognition Software into Speaking and Listening Activities

Liliia Kuchmarenko

ITMO University, Russian Federation

Abstract

The paper will demonstrate various strategies of integrating speech recognition software into speaking and listening activities. The strategies include using the digital tools that provide immediate feedback on pronunciation and fluency, encouragement of self-evaluation (the students will be able to hear and evaluate their own performance based on given criteria). The usage of speech recognition software will enable the educators to create a learning environment where language acquisition will become more engaging and effective.

Keywords: *Automated speech recognition, ASR, ELT, automated assessment, speech-to-speech recognition*

1. Introduction

Automatic speech recognition is a technology that enables a computer or a smartphone to recognize and process human speech into text. ASR is still imperfect but improving rapidly in terms of its accuracy in recognising spoken discourse and transcribing it into written text. ASR is based on big data-searching language corpora and finding matching patterns in data in order to convert the audio into written text. However, it does not analyse the audio semantically. It merely transcribes speech and turns spoken language into written language – using complex statistical and language analysis models.

Before proceeding with describing possible ways of integrating ASR technology into ELT, it is worth mentioning that ASR has already become a part of everyday life. This function is used by billions of users on their smartphones every day. Sending and receiving voice messages, transcribing the voice recordings have already become integral functions of the most popular messengers. It is hard to imagine any digital dialogue, especially an informal one, without using the voice messages.

It is worth emphasizing that ASR technology is able to make life easier not only for average smartphones' users, but also for English educators. According to N. Sutomo, an author of the study dedicated to the Implementation of Automated Speech Recognition (ASR) in ELT Classroom, "ASR technology, which converts speech audio into text, offers significant advantages in teaching pronunciation and evaluating learners' spoken language. It opens avenues for interactive aural engagements between students and computers. <…> It enables instructors to effectively assess spoken English in classrooms, bridging the gap between technology and teaching. The technology's impact on pedagogy extends to how educators perceive its benefits and limitations in second language (L2) instruction. ASR's capability to offer direct feedback and pronunciation assessment simplifies the teaching process, making it an invaluable tool in modern language education" [2].

2. Benefits of integrating ASR into ELT

First of all, ASR can facilitate new ways to work on phonology and accent – various applications are able to 'listen' to a learner's pronunciation and provide formative assessment and feedback on the accuracy of the pronunciation. The applications and the chatbots which were developed for working on accent and pronunciation will be described further.

Secondly, ASR can also facilitate responses to communicative interactions in the classroom, where students can use their devices (in pairs or groups) to speak or write responses to a task and get instant correction or formative assessment of their pronunciation or comprehensibility.

Moreover, ASR technology allows the educators to make the language learning process less intimidating as soon as it provides the teachers and the students with an opportunity to substitute a real speaker with a digital analogue that will not judge the student or make him or her feel embarassed. According to Rotimi and Cisneros, the authoes of the research dedicated to improving oral communication skills of English learners with voice messages and short videos, "Lack of communication skills is one of the problems students face when learning English as a second

language. Therefore, finding a solution to this problem is significant". [2]. ASR technology happens to be a suitable solution. The authors of the mentioned research claimed the following: "It may be argued that the opportunity to repeat and listen to individual voices both inside and outside the classroom increased confidence in using the language. The results show that most of the participants agree that using voice messages and short videos on WhatsApp would improve their Grammar and pronunciation in their oral communication skills" [2]. Therefore, it may be assumed that the use of ASR may increase the learners' confidence and fluency.

Additionally, various applications and chatbots based on the ASR technology are able to facilitate the assesment process - both written and spoken exams, with an accuracy approaching that of human assessors. According to N.Sutomo, "It enables instructors to effectively assess spoken English in classrooms, bridging the gap between technology and teaching. The technology's impact on pedagogy extends to how educators perceive its benefits and limitations in second language (L2) instruction. ASR's capability to offer direct feedback and pronunciation assessment simplifies the teaching process, making it an invaluable tool in modern language education [2].

Automated speech recognition can also facilitate automatic translation. The students will be provided with a tool that can make a real-life conversation with a native speaker less confusing. It may also help with facing some new vocabulary while completing reading or even listening activities. As M. Carrier points it out, "Speech-enabled translation software and apps allow students to speak into a phone or tablet and instantly hear the spoken translation in the target language (L2)" [1].

Another benefit of making the students familiar with automated speech recognition tools is the fact that they can provide the students with a flexible study option. Interacting with ASR applications may serve as a part of homework, additional self-study activity or any type of freer practice.

3. ASR tools and possible ways of integrating them into speaking and listening activities

In this part of the research the reader will become familiarized with different types and examples of the tools based on automated speech recognition software and various options of using them at the ELT classroom. According to the analysis of existing ASR tools, they can be divided into three groups:

1) Chat-bots in messengers (e.g. Telegram): these are the AI-powered software applications that stimulates the users' conversation to understand and respond the users' queries via text or voice. ASR chatbots are able to analyze the user's pronunciation, vocabulary and grammar and to provide the student with feedback on it.
2) ASR mobile applications: these are the mobile applications for learning English, based on the use of automated speech recognition software. The applications allow the users to receive detailed feedback on their accent, pronunciation, grammar and vocabulary, and even intonation. ASR applocations provide the users with the most individualized experience of diving into a foreign language, since the apps build up the studying plan, conversations with an AI-assistant and suggested exercises according to the user's personal information and studying needs.
3) ASR websites: these are the online-platforms offering a variety of activities, based on ASR technology. Unlike ASR mobile applications, they cannot be used from the smartphones, the user will need a laptop or a computer, which makes them a bit less convenient in terms of individual or freer practice. Nevertheless, it may serve as a useful tool to be applied at the full-class or groups activities.

Now, we will describe the examples of each type of ASR tools and suggest some effective ways of integrating them into speaking and listening activities.

3.1 ASR chatbots in messengers

Before we present a useful chatbot for leraning English, based on ASR technology, it should be mentioned that despite the fact that there are thousands of chatbots for learning a foreign language, only a few of them include a function of speech recognition for practicing speaking and listening. Most of the chatbots are aimed at improving vocabulary, translating, or reading skills.

AI Tutor for Speaking English – FalaBola Bot is an AI-driven chatbot that was programmed to conduct friendly conversation based on the student's interests. It helps the user to improve his or her English by talking with AI using voice messaged and practicing phrases where the mistakes are made. It also provides the user with detailed feedback on pronunciation, vocabulary and grammar mistakes by highlighting them and suggesting the student to correct themselves. As soon as the users start the bot, it opens the main menu which includes the following rubrics:

- Free talk;

- Role plays;
- Discuss post (this is a new function which allows the user to resend the photo or post from another telegram-chanel or any other platform in order to discuss it with an AI-assistant);
- Practice phrases.

Before starting the first conversation with an AI-assistant the chatbot asks you to choose at least three categories that interest the user (e.g. art, literature, technology, education, environment, fashion, etc.). Then, a personalized virtual assistant called Fala invites the user to introduce themselves in a very friendly manner (for instance, it uses the phrase "I can't wait to hear your story!"). It continues the conversation via voice messages, depending on what was said in the first user's message. Such a conversation with a virtual assistant can last for a limitless time. However, in case the user wants to receive some feedback on their speech, they should ask the chatbot to provide them with it. This quire makes the chatbot to switch from the speaking-and-listening to analyzing mode. Then, it stops replying with the voice messages and switches to sending texts in which it demonstrates the user's mistakes and suggests possible ways to correct them. Such text messages containing feedback are divided into two parts:

1) Advice: it repeats the fragment of the user's speech where the mistake was made, transcribing it from oral to written speech, and crosses out the incorrect words.
2) "Why it's better" part: in this part of the feedback the chatbot suggests possible ways of correcting the mistakes, explaining why it is better to say it this way. It is worth mentioning that the chatbot pays attention to stylistic nuances as well, explaining what sounds formal/informal.

That was a description of possible integrating FalaBola Bot into self-work or freer practice. The students can be given with target topics or vocabulary to discuss them with a virtual FalaBola assistant and to share the received feedback with their teacher.

Further, we will suggest a way of applying FalaBola chatbot to in-class speaking practice.

The target group for trying the following activity was a class of B1.2 (upper-intermediate) students. The topic of the lesson was called "Behind the scenes" (based on the unit 6a from English File Upper-Intermediate Student's book, 4th edition). Target vocabulary of the lesson was cinema terms and vocabulary, target grammar - passive voice.

Step 1: The students were divided into two groups given the following instructions: each group had to ask the chatbot to ask them a question about cinema using passive voice.

Step 2: Each group had to answer the question generated by AI-assistant. In they answer they had to use the words from target vocabulary (extras, scene, science fiction film, sequel, audience, soundtrack, subtitles, special effects, set, plot,trailer, etc.) and passive voice.

Step 3: Each group had to ask the virtual assistant to provide them with feedback on their answer.

Step 4: Each group was asked to count the mistakes they made and share this number (only the number, not the mistakes) with another group.

Step 5: The groups exchange the voice messages they have sent to the chatbot, listen to the voice message of another group and look for the mistakes they made and suggest the ways of correction.

Step 5: The groups share the analysis of each other's mistakes and compare it with the chatbot's feedback to make sure they found the required mistakes and suggested the right ways of correction.

As the experiment showed, such way of integrating FalaBola Bot into in-class speaking practice encourages the students' enthusiasm to implement target vocabulary and grammar into free speech, to work on mistakes without feeling pressure, since the teacher does not take part in correcting and assessing process.

3.2 ASR mobile applications

As an example of a useful mobile application, based on automated speech recognition technology, we will present an application named Roobbie.

This is a free app which includes an AI chat-bot that helps you to practice speaking based on your level and your interests. It analyses and counts the mistakes you make and gives you points and feedback after each conversation. It also gives you the exercises to work on your mistakes after each conversation.

Roobbie is the name of the AI-assistant which is always ready to support a conversation based on the user's needs and interests. The user can write text messages and record voice messages on any

topic at any time, while Robbie is always there to help to find the errors and to suggest some exercises to work on them.
These are the key features of the app:

1. Communication practice: the user is allowed to choose a topic he or she is interested in or to ask the AI-companion to start a conversation on a random topic.
2. Text and audio: as we have already mentioned it in the previous part of the article, despite the fact that there are hundreds of applications for practicing English via smartphone, there are not so many tools supporting the function of recording and receiving voice messages. Roobbie App provides the users with such an opportunity.
3. Dictionary and Translator: the app provides instant access to translation of the new words right in the chat.

In the next part of the article, we will demonstrate a way of applying the app in the classroom.
The suggested activity can be organized as work in pairs.

Step 1: the students are given the QR-code which allows them to download the app.
Step 2: before starting their work in pairs each student interacts with the app "face-to-face". (The topic of the lesson was social media). Each student asks the AI-assistant to start a conversation on the given topic. The app requires the user to choose his or her level and suggests a question. Here is the question suggested by the app for an advanced speaker: "Social media tools have revolutionized communication, but what do you think are the biggest drawbacks they present?"
Each student keeps the conversation with an AI-assistant via voice-messages by answering its three following questions (the number of the questions is unlimited, a teacher can regulate it according to the lesson's timing).
Step 3: the app analyzes the student's messages and prepares the feedback on the user's grammar/pronunciation mistakes. Each student has to push the button "native speaker's version" in front of their voice-messages. The button allows the user to see what mistakes he or she has made and to receive the corrected version of the native speaker. Even if the message was absolutely correct in terms of grammar and vocabulary, the "native speaker's" button will provide an improved version of the same statement. It paraphrases the student's statement, making it sound closer to real-life speech. (Here is the example of such grammatically correct student's response: "In my opinion, the biggest drawback of social media tools is that they make people communicate face-to-face less than they used to, so we don't really have any access to real-lofe communication anymore". Here is the native speaker version generated by the app: "In my opinion, the biggest drawback of social media tools is that they make people communicate face-to-face less than they used to, so we no longer have real-life communication").
Step 4: the students start working in pairs. Each student copies the transcript of his or her voice-message and transcript of the "native speaker's version" and resends them to the partner without letting them know what the transcript of the first message is and where the improved version is. The student's partner is supposed to identify his classmate's speech from the improved version and identify where it was improved.
Step 5: the partners show their texting history to each other and see whether they identified everything correctly.
The experiment demonstrated that such way of integrating Robbie app into in-class speaking practice stimulates the students' ability to reflect on different speakers' performance and nuances of real-life speech.

3.3 ASR websites

As an example of the website based on automated speech recognition technology we will suggest a platform called **"Yoodli: Interactive role-plays"**. This digital tool is useful for teaching ESP (English for Special Purposers) courses. The platform generates the interactive role-plays for pitch certification, sales onboarding, partner enablement, GTM (go-to-market) enablement, and manager training. The platform allows the users to try ready-made role-plays (buying committee product pitch, hiring manager interview, prime-time earnings interview, post-demo networking, etc.) and to build your own role-play based on your own goals and interests. It also provides a quick role-play to practice anything by just letting the AI know what you want to talk about.
We sill descrive a ready-made role-play "Hiring manager interview". At this role-play the users are supposed to discuss their skills and fit for a company and role. The participant of this role-play is required to demonstrate active-listening by responding thoughtfully and acknowledging key points, as well as demonstrating good interview skills through the effective use of the STAR (situation, task, action, result) method.

Step 1: the student allows access to the camera and microphone as it simulates a situation of an online-interview. Each role-play includes a character simulating a real person. In our case the character is Agness Potts,hiring manager.
Step 2: the student chooses a company where he is planning to work (the platform suggests a long list of real existing companies e.g. Gooogle, Amazon, Coca0Cola, etc.) and a role he is applying for (product manager, CEO, technical program manager, etc.)
Step 3: the students interacts with a role-play hiring manager by answering the questions about his or her working experience and academic background. Moreover, the manager provides the applicant with some imaginary situations asking how he or she would act in the given scenarios. The manager asks the participant to share not only professional, but also networking experience by asking about some challenging situations of interaction with the colleagues or administration from the past experience).
Step 4: as soon as the user feels ready to end the conversation, the button "summarize" leads him to the step of assessing his or her performance and getting some feedback from the platform. The platform allows the user to rewatch the whole interview (video recording), to relisten each response and to see the transcript of everything that was said during the conversation. It provides detailed feedback on each response and assesses the performance according to the following rubrics: active listening, the use of STAR method). It also shows the user the page called "Growth Area" where he or she may get acknowledged with the characteristics that require improvement.
"Yoodli: **Interactive Role-plays"** allows the users to get ready for real-life professional situations by taking part in simulated video-calls dedicated to various topics. It also provides the students with an ability to reflect on their performance and to receive useful and practically applicable recommendations that could improve their language, professional and soft skills.

4. Conclusion

This study aimed to investigate the effectiveness of applying automated speech recognition technology into speaking and listening activities. Various ways of applying the apps, websites and chat-bots based on ASR technology demonstrated that it can provide the users with an ability to overcome lack of real-life possibilities to train their communication skills. Interacting with the ASR chat-bots encourages the students' enthusiasm to implement target vocabulary and grammar into free speech, to work on mistakes without feeling pressure, since the teacher does not take part in correcting and assessing process. Integrating the ASR mobile apps into in-class speaking practice stimulates the students' ability to reflect on different speakers' performance and nuances of real-life speech. By getting familiar with the ASR websites the students receive an opportunity to get ready for real-life professional situations and to become aware of the forthcoming professional challenges.
Further research could be done on other chat-bots, platforms and applications by characterizing and categorizing them. According to R.P. Oye and C.Salvador-Cisneros, "Using those tools in the EFL classroom is sought to improve oral communication skills and the understanding of complex topics in language teaching. Creativity in using technology in the classroom is always recommended for teaching EFL" [3].

REFERENCES

[1] Carrier, M. "Automated Speech Recognition in language learning: Potential models, benefits and impact", Training, Language and Culture, 1(1), 2017, 46-61.
[2] Sutomo, N. (2024). "The Implementation of Automated Speech Recognition (ASR) in ELT Classroom: A Systematic Literature Review from 2012-2023", VELES (Voices of English Language Education Society),7, 2024, 816-828.
[3] Oye, R. & Salvador-Cisneros, K. "Improving oral communication skills of English learners with voice messages and short videos", Revista Tecnológica - Espol, 34(2), 2022, 155-164.