

Web-based speech-to-text tools to enhance english language pronunciation: a descriptive study

Herramientas de conversión de voz a texto basadas en la web para mejorar la pronunciación del idioma inglés: un estudio descriptivo

Xavier Sulca-Guale¹, Adriana Nicole Lozano Celleri¹, Marbella Cumanda Escalante Gamazo¹, Rafael Ricardo Arias Huertas²

¹ Universidad Técnica de Ambato Facultad de Ciencias Humanas y de la Educación, Ambato – Ecuador

² Instituto Tecnológico Superior Universitario ISTE, Ambato – Ecuador

Correo de correspondencia: manuelxsulcag@uta.edu.ec, alozano6373@uta.edu.ec, ma.escalante@uta.edu.ec
 rafaehr.ariash@iste.edu.ec

Información del artículo

Tipo de artículo:
Artículo original

Recibido:
10/07/2023

Aceptado:
30/09/2023

Aceptado:
31/10/2023

Revista:
DATEH



Resumen

El presente estudio determinó la importancia de las herramientas web de reconocimiento de voz a texto para mejorar la pronunciación del idioma inglés. Un total de 73 estudiantes universitarios (33 hombres y 40 mujeres) participaron en esta investigación descriptiva y no experimental. Los datos fueron recolectados a través de una encuesta con 29 ítems en escala Likert y 3 preguntas abiertas. El instrumento fue validado por expertos y con el coeficiente Alfa de Cronbach (0,856). Además, dicha encuesta se basó en tres preguntas de investigación. Los resultados revelaron que las herramientas web de reconocimiento de voz a texto son un buen medio para practicar la pronunciación y mejorarla puesto que, son de acceso gratuito, las voces de los hablantes se adaptan sin problemas a estos dispositivos, además de convertir las frases y palabras habladas en texto. Los alumnos consideraron que estas herramientas tienen varios beneficios, como alentarlos a mejorar su pronunciación y sus habilidades de habla y comunicación oral. Así como también, ayudan a los estudiantes a desarrollar características de pronunciación y promueven el trabajo autónomo. Por otro lado, los participantes manifestaron que existen diversas estrategias para poner en práctica la mejora de la pronunciación. La mayoría de los estudiantes prefirieron el uso de diferentes medios como; ver videos y escuchar música o podcasts en inglés y de esta manera repiten la pronunciación de sonidos y palabras. Sin embargo, el aprendizaje de las reglas de pronunciación es una estrategia poco frecuente porque los alumnos no tienen suficiente conocimiento de ellas y no se implementan en el plan de estudios o en la planificación de lecciones.

Palabras clave: Herramientas web de reconocimiento de voz a texto, Beneficios, Características de pronunciación, Estrategias de pronunciación.

Abstract

The current study determined the importance of pronunciation through the web-based speech-to-text tools to enhance the English language pronunciation. A total of 73 university students (33 males and 40 females) participated in this descriptive and non-experimental research. The data was gathered through a survey with 29 items on a Likert scale and, 3 open-ended questions. It was validated by experts and with the coefficient Cronbach's Alpha (0,856). Moreover, it was based on three research questions. The results revealed that web-based speech-to-text tools are a good means of practicing pronunciation because it has free access, the speaker's voice adapts smoothly, and it converts the spoken phrases and words into text. Learners pointed out that web-based speech-to-text tools have several benefits such as encouraging them to improve their pronunciation and speaking and oral communication skills. They assist students to develop pronunciation features, and they promote autonomous work. Furthermore, there are many strategies to put into practice the improvement of pronunciation. Most of the students preferred the use of media by watching videos or listening to music or podcasts in English and they usually repeat the pronunciation of sounds and words. However, pronunciation rules are not considered as a principal strategy because learners do not have sufficient knowledge of them, and they are not implemented in the curriculum or lesson planning.

Keywords: Web-based speech-to-text tools, Benefits, Pronunciation features, Pronunciation strategies.

Forma sugerida de citar (APA): López-Rodríguez, C. E., Sotelo-Muñoz, J. K., Muñoz-Venegas, I. J. y López-Aguas, N. F. (2024). Análisis de la multidimensionalidad del brand equity para el sector bancario: un estudio en la generación Z. Retos Revista de Ciencias de la Administración y Economía, 14(27), 9-20. <https://doi.org/10.17163/ret.n27.2024.01>

INTRODUCTION

In recent years some changes have occurred in our world. Grand-Clement (2017) stated that technological tools are a great educational resource. However, technology will not reduce the role of teachers in education, but educators will use it to make education more flexible and accessible. Twining et al. (2013) agreed that to effectively educate people, some changes must occur. They also stated that the educational community must see technology as an opportunity to introduce new goals, structures, and roles that support these changes.

Web-based tools for speech could be used for many different purposes such as education, business, translation, and document classification (Phung et al., 2021). Along with web-based speech tools there are some applications that take advantage of speech interface such as dictations tools (Tebelskis, 1995). Furthermore, there are plenty of spoken language materials such as audios, videos, etc. Nevertheless, in order to access to that content, it would be necessary to transcribe. That is why an automatic speech recognition (ASR) systems are required (Denisov et al., 2019).

The main purpose of this research work is to determine the importance of web-based speech to text tools. Phung et al., (2021) added that it is useful to use web-based speech tools with educational purposes to help people who want to learn any new target language and practice it. For the most part it is aimed to analyze how useful is it for learning another language through (ASR) to text.

Automatic Speech recognition is used to describe the ability to understand spoken languages (Levis & Suvorov, 2012). ASR is a technology based on developed speech to text recognition, which, with proper use, can be useful in pronunciation training. Therefore, the learning process becomes more realistic and interesting thanks to the ability of the computer to interpret the student's voice and response, it also provides feedback on the learner's pronunciation (Liu et al., 2019).

Speech recognition system: speech-to-text is the process of converting an acoustic signal which is captured using a microphone to a set of words. The recorded data can be used for document preparation among many other uses (Prasad, 2015).

Phung et al., (2021) concluded that the benefits of speech to text are as follows.

Ease of communication: it will provide an opportunity to easily communicate with others via text message as it can be dictated by the user and converted to text to be sent to the receiver.

Linguistic preservation: it can be used as a tool to encourage the use of the language as a medium of communication. It is therefore important to exclude features like code-mixing and code-switching and

incorporate linguistic features to facilitate the process of linguistic preservation by the community.

Time saved with increased efficiency and less paperwork: When traditional method is replaced with a mobile transcription app to speak into, one is able to boost their writing speed by nearly 4 times, an average of 150 words per minutes using a speech-to-text app.

Multitasking: Dictation on the go - eliminating the need to perform dictation tasks on larger and more cumbersome devices such as laptops or personal computers.

Accessibility: Devices such as mobile phones, tablets, and personal computers (PC) can be easily handled using the developed system.

The use of automatic speech recognition (ASR) or web-based speech to text tools can help learners sound more comfortable when speaking. For skills training, speaking with ASR was helpful for students in EFL settings. Although recording and monitoring were not unfamiliar to the learners, they responded favorably when they were linked with technology. Learners demonstrated good attitudes towards the use of ASR in learning English and it made the speaking activity more dynamic and motivated student performance (Ahn & Lee, 2016).

This approach promises a stress-free environment, which motivates students to contribute more as independent learners. It is highly important since it may have a big impact on how well learners pronounce words (Yaniafari & Olivia, 2022). For teaching the new abilities, the teacher employs many techniques. The teaching-learning process has recently concentrated on the classroom as well as teaching and evaluation methods. ASR technology can be used to assist different speaking practice methods and provide real-time feedback on many aspects of language competency, such as pronunciation and usage of the target language, to help learners improve their speaking abilities (Jiang et al., 2021).

On the other hand, pronunciation is one of the language skills that international students must learn to improve their communication skills. Pronunciation is the process of producing speech sounds to convey ideas (Yoshida, 2016). Tergujeff (2013) emphasized that a key component of pronunciation instruction is what to teach about it. Teachers need to take into account certain important aspects of pronunciation. First of all, they need to be aware that there will be as many different pronunciation problems as there are students. Teachers should also be aware of the main language phonetic framework impedances. Lastly, they should be concerned that they are expected to have accurate hypothetical articulation information as well as palatable knowledge of variations and contrasts.

Additionally, speaking clearly requires good pronunciation. In this sense, segmental features and suprasegmental features both contribute to the creation of

sounds in English (Gilakjani, 2012). The suprasegmental qualities are connected to units, like stress and tone, that span many sounds in an utterance. Learners should be aware of suprasegmental qualities in order to improve their pronunciation. Understanding the meaning of speakers is more important than just perfecting one's pronunciation (Ahmadi, 2018). In contrast, vocal-tract gestures are the units of segmental representation for particular vowels and consonants, which are segmental elements of the sound system (Hosseini et al., 2013).

Also, there are several strategies to enhance the English pronunciation. The covert rehearsal, is known as private practice, is the practice of speaking English aloud while the learner is alone. Anyone can auto evaluate the pronunciation if it is correct or incorrect by selecting words that people want to learn. The self-monitoring strategy involves focusing on a particular element of grammar or speech, for instance, the past tense of verbs. First, the student selects the language feature he wants to learn. Then learners start self-monitoring during a conversation and evaluates language progress. Finally, learners improve the harder aspect of his speech (Jensen, 2011). Apart from this, Pawlak (2018) mentioned the six language learning strategies related to pronunciation:

Cognitive strategy: Consist in using media, telling stories, repeating or talking with foreign people in order to practice the pronunciation.

Memory strategy: The learner repeats a statement many times, relating the manner a word or sound is pronounced to a context in which it was heard.

Compensation strategy: It refers to guess the pronunciation of a new word and try to avoid limitations.

Metacognitive strategy: It involves taking charge of own learning by centering, planning, and evaluating oneself. For example, taking note of mouth movements or lip movements, read up on the rules of the target language and put into practice the pronunciation.

Social strategy: It entails asking for help, cooperate, interact and empathizing with others.

Affective strategy: Being amusing when people mispronounce words and then ask the tutor to pronounce the word well or to look for the correct pronunciation.

MATERIALS AND METHODS

The investigation was carried out based on the mixed approach, qualitative and quantitative. A survey was used to gather qualitative data. Furthermore, the textual data collected through the survey was described and compared with the data previously evaluated. A total of 73 university students (33 males and 40 females) were enrolled in this study. The data was collected through a survey with 31 items on a Likert scale and, 3 open-ended questions. It was validated by experts as well as the coefficient Cronbach's Alpha, which resulted in 0,856. In this sense, Macnaughton (1996) stated that the qualitative approach

allows the study things in their natural environment, and the interpretation of the phenomena in terms of the meanings that individuals give them. This includes field notes, surveys, conversations, recordings, and private notes. In contrast, this research work had a quantitative approach due to the fact that the variables were susceptible to being observed and measured. The results obtained from the survey applied to the students were examined and compared through numbers. This approach basically involves gathering numerical data to describe a specific phenomenon, and certain questions appear to be naturally adapted to being addressed with quantitative methods. Eventually, it provides statistical analysis via statistical comparison of the data collection groupings.

Descriptive research was applied because the data and results obtained were analyzed and described to determine the precision and the point of view of the students. It took place in the classroom, in an environment where students perform naturally. It consisted of collecting the data from the survey and describe the real information that students completed according to their own experience and knowledge. Kothari (2004) pointed out that the goal of a descriptive research study is to describe the qualities of a certain person or group. This research is focused on making specific predictions and narrating information about facts and features related to people, organizations, or circumstances. The purpose of this research was to gather accurate and reliable data, hence the method to be utilized must be properly designed. It is essential to clearly define the objectives of descriptive research in order to guarantee that the information gathered is pertinent. The researcher may then employ one or more of these techniques to obtain the data: surveys, observation, questionnaires, interviewing, examination of records, etc. Lastly, descriptive research focuses on describing the features of a certain person or group are known as descriptive research studies. The researcher should be able to determine precisely what researchers want to assess, discover appropriate ways to use it, and define precisely the population he wants to examine.

RESULTS AND DISCUSSION

The results of this current study were based on the three research questions. The three research questions were:

1. What are learners' perspectives on web-based speech-to-text tools?
2. To what extent do web-based speech-to-text tools benefit the English pronunciation?
3. What are the strategies that learners use to improve their pronunciation?

Analysis and interpretation

Research question number one: *1. What are learners' perspectives on web-based speech-to-text tools?* The

results showed that the students consider that web-based speech-to-text tools are quite useful due to the fact that the accessibility, free access without any registration are the things that they like the most. These characteristics allow students to work freely with the applications. Consecutively, the applications adapt to the students' voice while they are using the apps. This indicates that the application adjusts the voice of a new speaker and recognizes the words easily. Moreover, when learners practice their pronunciation by saying different words, dialogues, etc., web-based speech-to-text tools write what they understand and change them into text. Meanwhile, when they mispronounce a word, the apps write other words instead. If there is an inaccurate pronunciation, the software will not write the correct phrase. In this regard, when the app does not write the word the students pronounced, they practice until the correct word is written. Most of the times teachers just teach different rules to improve pronunciation. Yet, these tools will provide learners stress-free environment, which motivate them learning meaningfully.

| Item | Mean |
|---|------|
| Web-based speech-to-text tools recognize what I say. | 3,88 |
| When I speak, the web-based speech-to-text tools adapt to my voice. | 4,03 |
| When I say a phrase, the web-based speech-to-text tools recognize and writes it. | 3,96 |
| When I practice with web-based speech-to-text tools, they have a high range of vocabulary. | 3,92 |
| I have free access to the application, without the need to register or create an account. | 4,66 |
| When I practice my pronunciation by saying different words, dialogues, etc., the web-based speech-to-text tools write them into a text. | 4,25 |
| When I mispronounce a word, the web-based speech-to-text tools write another word instead. | 4,32 |
| When the web-based speech-to-text tools do not write the word I pronounced, I practice | 4,40 |

until the correct word is written.

Table 1: *Perspectives of web-based speech-to-text tools*

Analysis and interpretation

Research question number two: *To what extent do web-based speech-to-text tools benefit the English pronunciation?* Results showed that when students use web-based speech-to-text tools, they have the opportunity to focus on their specific pronunciation difficulties. As a result of this, they concentrate on their individual learning challenges. Most students think that web-based speech-to-text tools are an effective means of teaching and learning the English pronunciation. Therefore, it assists students in developing their oral communication abilities. Also, web-based speech-to-text tools help learners to improve their pronunciation and their speaking skill. Thus, students would use web-based speech-to-text tools to improve their pronunciation features such as intonation, stress, etc to practice the segmental and suprasegmental features of pronunciation. In contrast, when students practice their pronunciation with web-based speech-to-text tools, they sometimes prefer to work independently the most.

| Item | Mean |
|---|------|
| When I use web-based speech-to-text tools , I have the opportunity to focus on my specific pronunciation difficulties. | 4,29 |
| I think that working with app promises a stress-free environment, which motivates me a better learning. | 4,11 |
| I think, web-based speech-to-text tools could be an effective means of teaching and learning the English pronunciation. | 4,15 |
| When I practice my pronunciation with web-based speech-to-text tools , I prefer to work with the whole class. | 3,08 |
| Web-based speech-to-text tools help me to improve my pronunciation and my speaking skill. | 4,30 |
| I would use web-based speech-to-text tools to improve my pronunciation features (intonation, stress, etc). | 4,14 |

Table 2: *Benefits of web-based speech-to-text tools*

Analysis and interpretation

Research question number three: *What are the strategies that learners use to improve their pronunciation?* According to the results, the majority of the students use

the cognitive strategy because they practice their pronunciation by using media. For example, with audio, music, watching series, and movies on Netflix or listening to podcasts on YouTube or Spotify. Nevertheless, they rarely practice their pronunciation by talking to foreigners. Another strategy that students use the most is self-monitoring because they pay close attention to their pronunciation during a conversation and they pay close attention to one specific aspect of pronunciation, for instance, verb tenses. In addition, the students practice their pronunciation by talking to themselves and it indicates that they often use the covert rehearsal strategy. On the other hand, learners hardly ever work with metacognitive strategy since they seldom take notes of their mouth or lips movements to check their pronunciation; and they not often study the pronunciation rules of words to pronounce them correctly.

| Item | Mean |
|--|------|
| I practice my pronunciation by talking to myself. | 4,16 |
| I pay close attention to my pronunciation during a conversation. | 4,16 |
| I pay close attention to one specific aspect of pronunciation, for example, past tense verbs. | 3,81 |
| I practice my pronunciation by telling stories. | 3,66 |
| I practice my pronunciation by using media. For example, with audio, music, watching series, and movies on Netflix or listening to podcasts on YouTube or Spotify. | 4,37 |
| I practice my pronunciation by talking to foreigners. | 2,82 |
| I practice my pronunciation by listening and repeating many times the words and sentences. | 4,01 |
| I repeat the pronunciation of a word when I remember a situation in which I heard it. | 4,12 |
| I guess the pronunciation of the words when I see new vocabulary. | 4,10 |
| I take notes of my mouth movements or lips movements to check my pronunciation. | 2,59 |
| I study the pronunciation rules of words to pronounce them correctly. | 3,21 |
| I ask for help if I cannot pronounce a word correctly. | 3,78 |
| When I practice pronunciation, I prefer to work with my teacher. | 3,10 |
| When I mispronounce a word, I do not pay attention to the mistake. | 2,64 |

| | |
|--|------|
| When I mispronounce a word, I make fun of it, and I correct the mistake immediately. | 3,58 |
|--|------|

Table 3: *Strategies to improve students' pronunciation*

Additionally, students use the memory strategy when they practice their pronunciation by listening and repeating the words and sentences many times. Learners repeat the pronunciation of a word when they remember a situation in which they heard that word. Moreover, students guess the pronunciation of words when they see new vocabulary. Thus, this way they practice with the compensation strategy. However, the participants sometimes choose the social strategy, when they occasionally work with their teacher. They use the affective strategy, when they mispronounce a word, they do not pay attention to the mistake.

| Question 1 | Answer | Total |
|--|---|-----------|
| What kind of applications do you use to improve your pronunciation? | Cambridge dictionary | 2 |
| | My English lab | 2 |
| | Youtube | 2 |
| | Windows 10 and Microsoft Word speech recognition. | 4 |
| | None | 4 |
| | Google: dictation, translation, voice. | 5 |
| | Other applications (Tandem, BoldVoice, Cake, HelloTalk, Reverso Context, Speak 6 English Pronunciation) | 6 |
| | Elsa Speak | 10 |
| | Speechnotes (web-based speech to text tool) | 10 |
| | Duolingo | 28 |
| Total | | 73 |
| Question 2 | Answer | Total |
| What are the benefits of using Web-based speech-to-text tools to improve your pronunciation? | I can practice more my speech delivery. | 1 |
| | I can recognize my pronunciation mistakes. | 1 |
| | I laugh with my partner. | 1 |

| | | |
|---|---|--------------|
| | Improve my accuracy in pronunciation. | 1 |
| | I could practice stressed words and intonation. | 3 |
| | I can improve my speaking and communicative skills. | 8 |
| | I improve my fluency. | 11 |
| | I practice my pronunciation with motivation. | 12 |
| | I improve my pronunciation. | 35 |
| | Total | 73 |
| Question 3 | Answer | Total |
| | By playing videogames. | 1 |
| | By singing songs. | 1 |
| | None. | 1 |
| | Roleplaying with my friends. | 1 |
| | Singing karaoke. | 1 |
| | Practicing with foreigners. | 6 |
| What kind of strategies do you use to improve your pronunciation? | Studying pronunciation rules. | 6 |
| | Dictating. | 7 |
| | Listening to podcasts or Music. | 9 |
| | Watching videos in English. | 9 |
| | Working with an application. | 9 |
| | Repeating sounds and words. | 22 |
| | Total | 73 |

Table 4: *Open-ended questions*

Analysis and interpretation

Table 4 presents the results of the open-ended questions used to reinforce the research questions:

The first open-question, *what kind of applications do you use to improve your pronunciation?* The principal application that a big part of students (28) prefer to use to improve their pronunciation is Duolingo. This free application is famous for the different activities they offer to develop the English skills (listening, writing, reading and speaking) followed by grammar and pronunciation. Another popular application that 10 learners use is Elsa

Speak. It is focused only on pronunciation activities. Speechnotes, which is a web-based speech-to-text tool was given 10 marks, so that students said they would like to use it to practice their pronunciation. A small number of participants (2) make use of the MyEnglishlab platform to train their pronunciation despite the fact that it is a platform they manage during their college studies. Additionally, the rest of the interviewees work with other applications for pronunciation, for instance, HelloTalk, Tandem, Google dictation, Google translation, Google voice, YouTube and among others.

The second open-question, *what are the benefits of using an application like web-based speech-to-text tools to improve your pronunciation?* Learners realize many benefits of web-based speech-to-text tools application. A large number of students (35) stated that they can improve their pronunciation while they are training with web-based speech-to-text tools. Some other benefits, that 11 students considered important, were the opportunity to work with these tools to improve their fluency, their oral communication abilities, as well as correct speech. Moreover, 12 learners indicated that they practice their pronunciation motivation, and it encourages students to perform pronunciation tasks in an effective and interactive learning environment. In addition, 8 learners could improve their speaking and communicative skills and 3 could practice stressed words and intonation which corresponds to the segmental and suprasegmental features. The third open-ended question, *what kind of strategies do you use to improve your pronunciation?* According to the results, 22 interviewees claimed that the principal strategy they use to improve their pronunciation is by repeating sounds and words. It is common that students listen and repeat the pronunciation of new words either listening to audios in English or to their teacher. Consequently, 27 learners pointed out that the strategies that they use most of the time are by using media: 9 of them prefer listening to podcasts or music; the other 9 would rather watch videos in English, and the last 9 learn to work with an application. This means that the use of technology it is quite effective to improve pronunciation by using various applications, listening to podcasts or music, or simply watching videos in English is a useful strategy. Hence, most of them have easy access to it, which can develop interactive activities in a context closer to English pronunciation. However, 7 learners practice their pronunciation by dictating and 6 study pronunciation rules. These results suggest that a great deal of students do not know much about these tools. Finally, it is possible to remark that only 2 students sing songs in order to practice their pronunciation. This indicates that even though teachers carry out tasks with songs in their classes, it is not the main method to practice pronunciation, the way many teachers assume it happens.

Discussion

Question 1: What are learners' perspectives on web-based speech-to-text tools? participants observed that the web-based speech-to-text tools adapt to their voice when they speak. Nevertheless, they realized that it occasionally fulfills with the speech continuity and vocabulary size dimensions for the reason that it seldom recognizes the spoken words or phrases; and when they practice with the application, it regularly has a high range of vocabulary. In this sense, Levis and Suvorov (2012) described the principal dimensions that web-based speech-to-text tools application have are: speaker independent, when the system adapts to the voice of a new speaker; speech continuity is related to the recognition of spoken words, and phrases, and converts them into text. Finally, the vocabulary size represents the variety of vocabulary that the system has.

The results showed that web-based speech-to-text tools respects the speech recognition phase when learners practice their pronunciation by saying different words, dialogues, etc., it writes them into a text. Also, the tools comply the error diagnosis phase is when the application does not write the word they pronounced, they practice until the words or phrases are written correctly. In addition, it follows the feedback presentation phase because the learners ask for help if a word is mispronounced. Neri et al. (2003) explained the principal phases that web-based speech-to-text tools application contains: error diagnosis phase and feedback presentation phase. The speech recognition phase which the system recognizes the incoming speech signal and writes it into text. The error diagnosis phase is when the participants speak too quickly or incorrectly, the application will write another word instead (Nurjanah et al., 2019). Finally, feedback appears to be efficient when the teacher constantly monitors the learners' progress to guide the progress of speaking skills pronunciation (Yaniafari & Olivia, 2022).

Question 2: *To what extent do web-based speech-to-text tools benefit the English pronunciation?* It can be shown that while the participants were practicing their pronunciation with web-based speech-to-text tools, it was possible to realize some benefits that Gottardi et al. (2022) mentioned: these tools facilitate the extensive practice of segmental and suprasegmental features of the language, from minimal pair to mirroring famous speeches or rehearsing presentations. It would help learners to improve students' pronunciation and oral communication skills, speaking fluency, and accuracy. These tools allow learners to focus on their specific difficulties and work independently. In this wise, the results showed that the implementation of the web-based speech-to-text tools application in class would help learners to improve their pronunciation and their speaking skill. Learners could

work with words, for example, minimal pairs, and following phrases that students consider are difficult for them to pronounce. Also, when they used web-based speech-to-text tools, they had the opportunity to focus on their specific pronunciation difficulties, as a result, they concentrated on their learning challenges.

Furthermore, the speech to text tools would assist students in developing their oral communication abilities because they think, they could be an effective means of teaching and learning English pronunciation. Nevertheless, the students would sometimes prefer to train their pronunciation with the whole class or with a classmate, which means that they would rather work on their own. In addition, learners would use web-based speech-to-text tools to improve their pronunciation features, this way, they could realize their pronunciation errors and improve them. In this regard, Pourhosein and Sabouri, (2017) concluded that computer technology can be an effective means of teaching English pronunciation especially on pronunciation features. Some of them involve speech rate, fluency, and liveliness, intonation, pronunciation quality of individual words. Therefore, teachers can use computer in their pronunciation classes to improve these features. It can provide them an interactive learning environment in different modes such as whole class, small group or pair, and teacher to student.

Question 3: *What are the strategies that learners use to improve the English pronunciation?* The principal strategies that students use are the covert rehearsal strategy when they practice their pronunciation by talking to themselves. They make emphasize individual pronunciation activities. The self-monitoring strategy is drawn upon by students when, during a conversation, they pay close attention to their pronunciation, and one specific aspect of it, for example, verb tenses. These two learning approaches reflects that they enjoy the autonomous work. Contrasted with Jensen (2011), the covert rehearsal strategy allows learners to focus on practicing their pronunciation on their own by talking with themselves. Subsequently, self-monitoring entails paying attention to specific aspect of grammar or speech and students assesses their advancement.

Students hardly ever use the metacognitive strategy because few of them pay close attention to the pronunciation rules to check their pronunciation or their mouth movements. Furthermore, most of the time, learners not frequently prefer the social strategy because they work alone when they practice their pronunciation, and they sometimes request the advice of a teacher for a mispronounced word or phrase. Also, the affective strategy is occasionally used by the participants. They sometimes make jokes or fun with the wrong pronunciation of words. Consequently, Oxford (1989) explained that students often get confused when the rules

and new vocabulary are presented, and they feel suffocated if they make mistakes with the metacognitive strategy. In regard of the social strategy, it is possible to contrast that it is important and recommended for learners to practice the communicative skills with a partner or cooperate with others, and if they have doubts, ask for help from the teacher. Moreover, the affective strategy is related to the attitude's emotions, a positive attitude. In this vein, students have an enjoyable and effective atmosphere because of the good sense of humor with their mispronunciation.

Additionally, the majority of the students prefer the cognitive strategy the most because they usually make the use of media, for instance, they listen to music or podcasts, and watch videos in English for practicing their pronunciation, as well, they occasionally read aloud or tell stories. On the other hand, they almost never practice their pronunciation by talking with foreign people. In a somewhat similar vein, the memory strategy is the second strategy that learners employ the most. In fact, they frequently practice their pronunciation by listening and repeating several times the words and sentences, and when they associate them with a situation in which they heard them. Considering the compensation strategy, it is another one that students prefer to use. When they see new vocabulary or words, they use to guess their pronunciation. Pawlak (2018) explained that the cognitive strategy is more used by students because of the variety of resources they choose to enhance their pronunciation. however, they do not naturally practice with foreigners since they are not in a native speaking environment. Otherwise, memory is the kind of strategy that benefits students in remembering the pronunciation and sounds of words while they remember different situations, they were exposed. Finally, it is common for learners when they produce sounds from words as they think their pronunciation is, they apply the compensation strategy.

CONCLUSIONS

According to the interviewees, the most important dimension and phases that web-based speech-to-text tools have are speaker dependence when speakers say words through the application, it adapts to their voice works without the need to register. The speech continuity is a dimension which the application recognizes and writes a word or phrase. The participants asserted that web-based speech-to-text tools have the advantage of having free access. This is evident when they downloaded the application, they used it instantly. Furthermore, web-based speech-to-text tools comply with the phases of speech recognition because converts the spoken words into text. Meanwhile, the error diagnosis phase is associated with a mispronounced word or phrase that is written instead of the correct one in the web-based speech-to-text tools application. Finally, if the speaker asks for help or keep

practicing until the word is written correctly, the feedback phase is applied.

It was possible to identify the benefits of the web-based speech-to-text tools. They reside in the following pillars. It promotes concentration on the individual learning challenges and brings an effective and stress-free environment with motivation as it encourages students to state their points of view according to their preferences. Secondly, these tools allow the practice of segmental and suprasegmental features, and it promotes the improvement of pronunciation and speaking skill. Last but not least, it assists students to develop their oral communication abilities since it could be an effective means of teaching and learning English pronunciation. Web-based speech-to-text tools promote autonomy in L2 students, as a result, they prefer to perform the pronunciation activities on their own.

The principal strategies the students frequently use to improve their pronunciation are the covered rehearsal, indeed, learners practice their pronunciation by themselves in the target language, talking, or listening carefully to their utterances while speaking. Moreover, they showed a particular interest in the memory strategy, in view of the fact that, repeat the sounds and words that associate with situations that they have heard. Regarding the cognitive strategy, learners preferred the use of different resources especially social media. Most of the time, they make use of it to practice their pronunciation by listening to music, podcasts, or watching videos or movies in or out of classes. However, they almost never practice talking to foreigners. It means that learners are not exposed to an English language environment with native speakers. Also, it is worth mentioning that the social strategy is occasionally used by the participants which are related to practicing with a partner or asking for help if it is needed. They frequently prefer to work independently. Finally, they rarely take advantage of the metacognitive strategy because they hardly ever study the pronunciation rules which reveals that they employ the compensation strategy because they often guess the pronunciation of the words if new vocabulary is presented.

AUTHORS CONTRIBUTION

Xavier Sulca-Guale and Adriana Nicole Lozano Celleri wrote, revised the article as well as created the survey and surveyed all participants. Marbella Cumanda Escalante Gamazo revised and provide feedback to the article. Rafael Ricardo Arias Huertas wrote a part the introduction and revised the article.

REFERENCES:

Ahmadi, D., & Reza, M. (2018). The use of technology in English language learning: A literature review. *International Journal of Research in English*

- Education, 3(2), 115- 125. DOI 10.29252/ijree.3.2.115
- Ahn, T. Y., & Lee, S. M. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology*, 47(4), 778-786. DOI:10.1111/bjet.12354
- Denisov, P., & Vu, N. T. (2019). IMS-speech: A speech to text tool. arXiv preprint arXiv:1908.04743
- Grand-Clement, S. (2017). *Digital Learning: Education and Skills in the Digital Age*. RAND Europe.
- Gilakjani, A. P., & Ahmadi, M. R. (2011). Why is pronunciation so difficult to learn? *English language teaching*, 4(3), 74-83. DOI 10.5539/elt.v4n3p74
- Gottardi, W., Almeida, J. F. D., & Tumolo, C. H. S. (2022). Automatic speech recognition and text-to-speech technologies for L2 pronunciation improvement: reflections on their affordances. *Texto livre*, 15. <https://doi.org/10.35699/19833652.2022.3676>
- Hosseini, S. B., & Pourmandnia, D. (2013). Language learners' attitudes and beliefs: Brief review of the related literature and frameworks. *International Journal on New Trends in Education and Their Implications*, 4(4), 63-74.
- Jensen, B. S. (2011). *The Michigan guide to the TOEIC(R) speaking test* (1 ed.). University of Michigan Press ELT. doi:10.5926/ijep.60.92
- Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., & Wu, N. (2021). Using automatic speech recognition technology to enhance EFL learners' oral language complexity in a flipped classroom. *Australasian journal of educational technology*, 37(2), 110-131. <https://doi.org/10.14742/ajet.6798>
- Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International. 9788122424881
- Levis, J., & Suvorov, R. (2012). Automatic speech recognition. *The encyclopedia of applied linguistics*. DOI 10.1007/978-981-15-0595-9_2
- Liu, X., Xu, M., Li, M., Han, M., Chen, Z., Mo, Y., ... & Liu, M. (2019). Improving English pronunciation via automatic speech recognition technology. *International Journal of Innovation and Learning*, 25(2), 126-140. <https://doi.org/10.1504/IJIL.2019.097674>
- Macnaughton, R. J. (1996). Numbers, scales, and qualitative research. *The Lancet*, 347(9008), 1099-1100.
- Neri, A., Cucchiarini, C., & Strik, H. (2003, August). Automatic speech recognition for second language learning: How and why it actually works. In *Proc. ICPhS* (pp. 1157- 1160). https://www.researchgate.net/publication/228604457_Automatic_speech_recognition_for_second_language_learning_How_and_why_it_actually_works
- Nurjanah, S. E. L., Ifadah, M., & Mulyadi, D. (2019a). Enhancing students' pronunciation accuracy through Web-based speech-to-text tools application at MAN 1 Semarang. In *Prosiding Seminar Nasional Mahasiswa Unimus* (Vol. 2). <https://prosiding.unimus.ac.id/index.php/mahasiswa/article/view/490>
- Oxford, R. L., Lavine, R. Z., & Crookall, D. (1989). Language learning strategies, the communicative approach, and their classroom implications. *Foreign Language Annals*, 22(1), 29-39. <https://doi.org/10.1111/j.19449720.1989.tb03139.x>
- Pawlak, M., & Szyszka, M. (2018). Researching pronunciation learning strategies: An overview and a critical look. *Studies in Second Language Learning and Teaching*, 8(2), 293-323. DOI 10.14746/ssl.2018.8.2.6
- Phung, K., Ramachandran, R., & Ogunshile, E. (2021). Exploring a Web-Based Application to Convert Tamil and Vietnamese Speech to Text without the Effect of Code-Switching and Code-Mixing. *Programming and Computer Software*, 47, 757-764.
- Pourhosein Gilakjani, A., & Sabouri, N. B. (2017). Advantages of using computer in teaching English pronunciation. *International Journal of Research in English Education*, 2(3), 78-85. DOI 10.18869/acadpub.ijree.2.3.78
- Prasad, V., *Voice recognition system: speech-to-text*, J.Appl. Fundam. Sci., 2015, vol. 1, no. 2, p. 191.
- Tebelskis, J., *Speech recognition using neural networks*, PhD Dissertation, Carnegie Mellon Univ., 1995
- Tergujeff, E. (2013). English pronunciation teaching in Finland. *Jyvaskyla studies in humanities*, (207). https://jyx.jyu.fi/bitstream/handle/123456789/41900/1/978-951-3953225_vaitos03082013.pdf
- Twining, P., Raffaghelli, J., Albion, P., & Knezek, D. (2013). Moving education into the digital age: the contribution of teachers' professional development. *Journal of computer assisted learning*, 29(5), 426-437.
- Yanifari, R. P., & Olivia, V. (2022). The Potential of ASR for Improving English Pronunciation: A Review. *KnE Social Sciences*, 281-289. <https://doi.org/10.18502/kss.v7i7.10670>
- Yoshida, M. T. (2016). *Beyond Repeat after Me: Teaching Pronunciation to English Learners*. TESOL Press. Available from: TESOL International Association. 1925 Ballenger Avenue Suite 550,

Alexandria, VA
22314.[http://www.tesol.org/docs/defaultsource/
books/14038_sam.pdf?sfvrsn=2](http://www.tesol.org/docs/defaultsource/books/14038_sam.pdf?sfvrsn=2)