A theoretical assessment of ASR-based pronunciation activities for classroom learning

Avaliação teórica de atividades de pronúncia baseadas em tecnologia de reconhecimento de fala para aprendizagem em sala de aula

William Gottardi (iD)[*1] **and Rosane Silveira** (iD)[†1]

[1]*Federal University of Santa Catarina, English: Linguistic and Literary Studies Program, Department of Foreign Languages and Literature, Florianópolis, SC, Brazil.*

Abstract

The use of digital technologies in the daily routine of second language (L2) classes became unavoidable when teachers needed to overcome mobility restrictions imposed by the COVID-19 pandemic. In countries where access to technological resources in regular classrooms is still limited, the use of free digital resources for pedagogical purposes is a valuable alternative. This pedagogical paper explores the use of Automatic Speech Recognition (ASR) for L2 pronunciation teaching and learning. Three ASR-based activities drawn from Gottardi (2023) are presented and discussed. The assessment is based on the criteria provided by two frameworks. From the pedagogical perspective, we adopted the communicative pronunciation teaching framework (Celce-Murcia; Brinton; Goodwin, 2010), which proposes five steps to teach pronunciation and design materials, including explicit instruction, perception, and production tasks, with different degrees of teacher guidance and feedback. Considering the use of technology and the design of pedagogical activities, we draw on Chapelle's (2001) six criteria to evaluate educational software and activities. We analyzed each ASR-based pronunciation activity in terms of language learning potential, learner fit, meaning focus, authenticity, positive impact, and practicality. Thus, the objective of this pedagogical paper is to 1) discuss how ASR technology can assist L2 pronunciation practice; 2) provide practical examples of ASR-based pronunciation activities designed to be used by L2 teachers; and 3) examine the pedagogical appropriateness of the proposed activities. Finally, pedagogical implications of using ASR for pronunciation practice are discussed.

Keywords: Computer-Assisted Language Learning. Pronunciation teaching. Pronunciation learning. Automatic Speech Recognition. Intelligibility.

Resumo

O uso de tecnologias digitais em aulas de uma segunda língua (L2) tornou-se inevitável quando os professores precisaram enfrentar as restrições de mobilidade impostas pela pandemia do COVID-19. Nos países onde o acesso a recursos tecnológicos nas salas de aula ainda é limitado, a utilização de recursos digitais gratuitos para fins pedagógicos é uma alternativa valiosa. Este artigo apresenta uma reflexão sobre o uso de Reconhecimento Automático da Fala (ASR – *Automatic Speech Recognition*) aplicado no ensino e aprendizagem da pronúncia de L2. Três atividades que utilizam a tecnologia de ASR (Gottardi, 2023) são apresentadas e discutidas. A avaliação se apoia em critérios fornecidos por dois modelos teóricos. Foi adotado o referencial de ensino comunicativo da pronúncia (Celce-Murcia; Brinton; Goodwin, 2010), o qual inclui cinco etapas para o ensino de pronúncia e elaboração de materiais, contendo momentos de instrução explícita e tarefas de percepção e produção, com diferentes graus de feedback e mediação do professor. Recorremos aos critérios de Chapelle (2001) para avaliar atividades e *softwares* educacionais, o qual nos permitiu analisar as atividades de pronúncia em termos de: potencial de aprendizagem da língua, adequação ao aprendiz, foco no significado, autenticidade, impacto positivo e praticidade. Assim, os objetivos deste artigo pedagógico são: 1) discutir como a tecnologia de ASR pode auxiliar na prática da pronúncia de uma segunda língua; 2) fornecer exemplos práticos de atividades de pronúncia que fazem uso de ASR, mediadas por professores de L2; e 3) examinar a adequação pedagógica das atividades propostas. Por fim, implicações pedagógicas são discutidas em relação ao uso de ASR para a prática de pronúncia.

Palavras-chave: Aprendizagem de Línguas Mediada por Computador. Ensino da pronúncia. Aprendizagem de pronúncia. Reconhecimento Automático da Fala. Inteligibilidade.

---

*Email: prof.williamgottardi@outlook.com
†Email: rosanesilduarte@gmail.com

1 Introduction

The term pronunciation encompasses "all aspects of how we employ speech sounds for communicating" (Burns; Seidlhofer, 2020, p. 247) not being limited to single sounds since it comprehends other speech features such as segments, prosody, and connected speech phenomena (Cardoso, 2017). Although pronunciation can be seen as an essential element of successful spoken communication (Baldissera; Tumolo, 2021; Pennington; Rogerson-Revell, 2019), it "has been historically under-represented and overlooked in comparison to other areas of second language acquisition research" (Foote; Trofimovich, 2017, p. 75).

Furthermore, pronunciation teaching is commonly neglected in second language[1] (L2) classrooms, which is not surprising since the field of second language acquisition (SLA) has historically marginalized pronunciation from both theoretical and pedagogical perspectives (Olson, 2023). Also, time dedicated to the teaching of L2 pronunciation is limited by an overloaded syllabus (Collins; Muñoz, 2016).

Bearing this scenario in mind, the use of digital technology – particularly speech technologies – can help mitigate this time constraint (Gottardi; Almeida; Tumolo, 2022; Silveira; Zanchet; Pereira, 2022). For example, Automatic Speech Recognition (ASR) is a speech technology with proven positive impact on L2 learning that can foster learners' autonomy by engaging them in extended pronunciation practice (Gottardi; Almeida; Tumolo, 2022). In short, the use of ASR technology might be an alternative to circumvent time constraints related to an overloaded curriculum and provide learners with more pronunciation practice opportunities.

Considering the new scenario brought about by the COVID-19 pandemic that "pushed virtually all teachers to teach via technological mediation" (Kern, 2024, p. 529), investigation into new digital technologies to assist language teaching and learning is welcome, especially the ones that can be applied into different teaching contexts as well as autonomous learning, as is the case with ASR technology. Thus, pedagogical reflections on the use of ASR as a teaching resource and suggestions on how to integrate it into teachers' practices are relevant if the guiding role played by the teacher is taken into account as well as their needs for support and assistance.

All in all, this pedagogical paper has the aim of 1) discussing how ASR technology can aid L2 pronunciation practice; 2) providing practical suggestions of ASR-based pronunciation activities designed to be used by L2 teachers; and 3) examining the appropriateness of these pronunciation activities from a pedagogical perspective.

In the following sections, we review relevant literature concerning L2 pronunciation teaching dimensions, priorities and a framework for designing pronunciation teaching materials. Then, we discuss important issues related to digital technologies and L2 pronunciation teaching. Section 4 brings a discussion about ASR for L2 pedagogical purposes followed by the analysis of three pronunciation activities designed to be implemented in L2 classes mediated by teachers.

2 Pronunciation teaching

Pronunciation is an essential component targeting successful spoken communication (Pennington; Rogerson-Revell, 2019) and is closely linked to social issues since it "cannot be ignored in view of globalization, the growing needs of cross-cultural communication, and issues of speaker identity and social integration" (Chun, 2020, p. 213).

In the field of L2 pronunciation, speech is often divided into three different dimensions, namely: *intelligibility*, *comprehensibility*, and *accentedness*. Following Derwing and Munro (2015, p. 5), intelligibility can be defined as "the degree of match between the speaker's intended message and listener's comprehension". Furthermore, the authors define comprehensibility as the degree of difficulty in understanding speech based on the listener's estimation. Lastly, accentedness refers to how different the speech is from the expected production patterns of a speech community.

Current views of L2 pronunciation teaching, based on the findings of L2 speech research, advocate a focus on enhancing intelligibility and comprehensibility, considering that accent is ubiquitous and striving for native-like pronunciation is considered an inadequate goal (O'Brien et al., 2018). Likewise, Levis (2020) contends that focusing pronunciation teaching on intelligibility aligns with realistic goals since accent should be seen as a variation to embrace instead of a problem to overcome.

In addition, communication is the primary purpose of language use. Hence, the main goal of L2 teaching should be using language to communicate (Celce-Murcia; Brinton; Goodwin, 2010). Following this principle, Celce-Murcia, Brinton, and Goodwin (2010) proposed a framework for pronunciation teaching based on the tenets of communicative language teaching (CLT). Table 1 summarizes this framework, which considers the learners' gradual development of new L2 pronunciation features.

Table 1. Communicative framework for teaching English pronunciation.

1- Description and analysis	Oral and written illustrations of how the feature is produced and when it occurs within spoken discourse.
2- Listening discrimination	Focused listening practice with feedback on learners' ability to correctly discriminate the feature.
3- Controlled practice	Oral reading of minimal-pair sentences, short dialogues, etc., with special attention paid to the highlighted feature in order to raise learner's consciousness.
4- Guided practice	Structured communication exercises, such as information-gap activities or cued dialogues, that enable the learner to monitor for the specified feature.
5- Communicative practice	Less structured, fluency-building activities (e.g., role play, problem solving) that require the learner to attend to both form and content of utterances.

Source: Celce-Murcia, Brinton, and Goodwin (2010, p. 45)

The above-mentioned framework suggests a 5-phase division of the pronunciation lesson. The first phase focuses on phonological awareness, the second focuses on perception skills while the last three phases address the learner's production. The framework recognizes that pronunciation learning is not a linear process. Therefore, teachers might need to revisit certain phases during their teaching process. The authors of this framework also advocate setting realistic goals for pronunciation teaching, "focusing on the targeted communicative needs of […] learners, not the language as a whole" (Celce-Murcia; Brinton; Goodwin, 2010, p. 283).

Considering the arguments presented in this section, pronunciation teaching and learning are of pivotal importance when it comes to achieving intelligible speech, aiming at successful communication. As Pennington and Rogerson-Revell (2019, p. 219) contend, "having a critical awareness of pedagogic resources and related research can help teachers make informed choices about what and how to teach". In addition, pronunciation learning is highly affected by learners' exposure to the L2 as well as the extent to which they actually use it (Lane; Brown, 2010). Therefore, digital resources may help learners to have more L2 exposure as well as teachers to achieve their pedagogical goals. The next section discusses the relationship between digital technologies and pronunciation teaching.

3 Contributions of digital technologies to pronunciation teaching

Digital technologies grant access for people to "speak or write either synchronously or asynchronously, with participants either at a distance or in close proximity" (Chun; Smith; Kern, 2016, p. 66). Nevertheless, the use of digital technology should not be considered the solution to all problems, but rather a way of supporting specific pedagogical goals (Chun; Smith; Kern, 2016), such as offering varied oral input and immediate feedback to learners (Baldissera; Tumolo, 2021).

Moreover, digital technologies may allow learners to be more autonomous toward their learning process, besides promoting learning beyond the classroom (Lai, 2018). In a similar vein, Thomson and Derwing (2015, p. 336) contend that "computer-based approaches may promote greater learner autonomy and, most importantly, afford the possibility of individualized instruction". Finally, technology can also offer "the learner limitless choice of what, when, and how to learn" (Rogerson-Revell, 2021, p. 201).

This complex nature of technology and pedagogy demanded the creation of a new subfield of applied linguistics called Computer-Assisted Language Learning (CALL), which studies the relationship between technology and L2 learning (Martins; Moreira, 2012). Simply put, CALL is "an approach to

language teaching and learning in which computer technology is used as an aid to the presentation, reinforcement, and assessment of material to be learned" (Davies, 2006, p. 261).

In recent decades, this field has advanced quickly (Pennington; Rogerson-Revell, 2019) and new digital resources are made available to teachers and learners all the time. Although having access to more CALL resources might be a positive asset, it can also represent an extra challenge for both educators and students to stay updated with the latest releases (Soleimani, 2021). Thus, evaluating CALL software and the activities designed to use CALL resources is relevant since it is what "teachers need to concentrate more on, in order to ensure that technology is not used for technology's sake" (Stanley, 2013, p. 9).

The evaluation of CALL learning activities can be challenging. However, it is necessary in order to assess how pedagogically adequate the digital resources are and how appropriate the CALL activity is to learners' needs. According to Chapelle (2017), a possible way of evaluating the appropriateness of L2 teaching materials and positive aspects of learning activities is by making a pedagogy-based argument. Following this line of argumentation, Chapelle (2001) suggests a set of SLA-grounded criteria (see Table 2 below) that can be applied to evaluating educational software and activities that teachers plan to use in their classes. Adopting such a framework can help "support claims about the value of technology-mediated learning for meeting pedagogical requirements" (Chapelle, 2017, p. 386).

Table 2. The six criteria for assessing CALL task appropriateness.

1- Language learning potential	The degree of opportunity present for beneficial focus on form.
2- Learner fit	The amount of opportunity for engagement with language under appropriate conditions given learner characteristics.
3- Meaning focus	The extent to which learners' attention is directed toward the meaning of the language.
4- Authenticity	The degree of correspondence between the CALL activity and target language activities of interest to learners out of the classroom.
5- Positive impact	The positive effects of the CALL activity on those who participate in it.
6- Practicality	The adequacy of resources to support the use of the CALL Activity.

Source: Chapelle (2001, p. 55)

The language learning potential criterion evaluates whether the activity is a proper language learning activity instead of a simple opportunity for language use, that is, the activity can promote a beneficial focus on linguistic forms. Although the importance of each criterion may differ according to the objective of the task, language learning potential "should be considered the most critical for CALL activities" (Chapelle, 2001, p. 58).

The learner fit criterion refers to how appropriate the activity is for the learner, considering the difficulty level, and the learner's individual differences and characteristics. The meaning focus criterion evaluates if the activity directs learners' primary attention to the meaning of the language required to accomplish the task. This criterion also takes into account activities that involve using language skills for the interpretation and construction of meaning.

The authenticity criterion analyses whether the learning task is similar to a task that learners would expect to see outside of the L2 classroom. As Chapelle (2001, p. 56) states, "the criterion of authenticity indicates the need to develop learners' willingness to communicate, but it also extends beyond the conditions believed important for acquisition"; therefore, activities should offer opportunity to interact with other language users and to generate more spoken output.

The positive impact criterion evaluates if the learning task teaches more than language, that is, supports learners to develop or improve their metacognitive strategies, allowing them to claim agency over their learning process. It also evaluates if both teachers and learners might have a positive learning experience during the activity.

Finally, the practicality criterion evaluates whether the learning task is easy for the teachers and

learners to implement considering the particular constraints of an L2 class or a language program. In addition, this last criterion takes into account the availability of software, hardware, and knowledgeable personnel needed for the proper execution of the learning task (Chapelle, 2001).

All things considered, the benefits of current digital technologies to pronunciation teaching are enormous (O'Brien *et al.*, 2018), and the advances in computer technology have considerably contributed to the improvement of L2 teaching and learning (Cardoso, 2022). Nonetheless, a particular speech technology, Automatic Speech Recognition (ASR), may be a suitable CALL resource as ASR-based pronunciation activities have the potential to fulfill the aforementioned criteria proposed by Chapelle (2001), as we shall see next.

3.1 Automatic speech recognition and L2 pronunciation

Simply stated, ASR is "an independent, machine-based process of decoding and transcribing oral speech" (Levis; Suvorov, 2020, p. 149); that is, it transcribes oral input (speech) into orthographic output. Although ASR technology is not recent, since the first ASR systems date back to the 1950s, recent years witnessed an accelerated improvement in its capabilities due to the development of more robust algorithms and new model techniques, the expanding demand for ASR integration with smart devices, and the improvement in noisy speech recognition (Jurafsky; Martin, 2024; Levis; Suvorov, 2020). Throughout these years, much progress has been made, allowing the current level of speech recognition accuracy exceeding 95% (Cardoso, 2022).

ASR technology has improved to the extent that it is suitable for many tasks. It can be embedded into smart devices containing Intelligent Personal Assistants (IPAs) (e.g., Google Assistant, Siri, Alexa), used to auto-generate captions, applied to human-computer interaction, or dictation tools (Jurafsky; Martin, 2024; Lieshout; Cardoso, 2022).

In recent years, two significant events have occurred that have attracted the interest of both the academic community and tech-companies. First, the impact of Covid-19 on educational contexts that "led to educational technology being used for learning out of school, at a pace and scale with no historical precedent" (Unesco, 2023, p. 13). Second, the emergence of generative Artificial Intelligence (AI)-powered chatbots, or AI-powered assistants, which allow users to have conversations with a virtual assistant while searching and generating content (Ibm, 2024).

When it comes to the affordances and limitations of ASR for L2 learning and teaching, especially for pronunciation development, much research has been conducted in the past decade[2]. First, teachers seem to hold a positive attitude towards ASR for pronunciation teaching (Gottardi; Silveira, 2024). Also, ASR's capacity to generate instant orthographic feedback may help learners to be more independent in their learning process and this corrective feedback might be useful for pronunciation learning (Liakina; Liakin, 2022). Finally, ASR tools may help learners to be more autonomous, working as a valuable self-directed language learning resource (Cardoso, 2022).

Furthermore, as stated by Kivistö-de Souza and Gottardi (2022, p. 778), ASR tools "can be a helpful additional pronunciation learning tool for L2 users", especially because they might assist learners in noticing pronunciation patterns that hinder communication. In a similar vein, Santos and Kivistö-de Souza (2024, p. 14) state that using ASR to practice isolated words "may still play a significant role in contributing with the noticing of specific phonological issues at the segmental level". ASR has also been shown to improve learners' pronunciation and that they tend to have a positive attitude toward ASR tools (Nickolai; Schaefer; Figueroa, 2024).

On the other hand, ASR can present pedagogical limitations and potential biases. ASR accuracy is significantly reduced if input speech contains distortions, is spoken far from the recording microphone, in a noisy environment, or the speaker's accent is too different from those used to train the ASR program; that is, native-accented speech may be more easily recognized by ASR systems (Jurafsky; Martin, 2024; O'Shaughnessy, 2024).

In summary, to consider ASR as a pedagogical resource, it is important to have an overview of all the recent technological advances and current limitations. Notwithstanding, ASR affordances for pronunciation teaching and learning are numerous. Thus, knowing how to implement ASR tools in an L2 classroom may be a valuable pedagogical skill for teachers. In the next section, ASR-based

pronunciation activities are presented and analyzed with the help of the two theoretical frameworks described in sections 2 and 3.
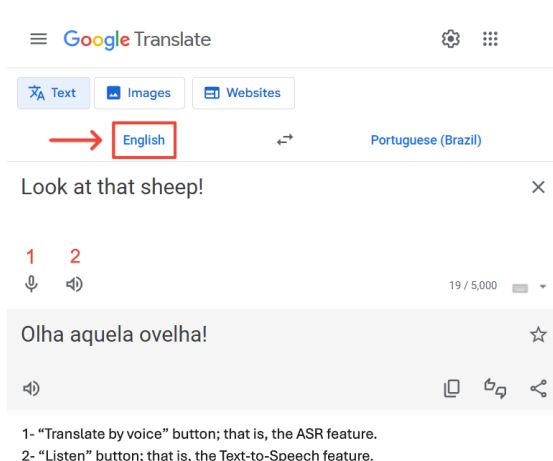
4 Assessing ASR-based activities designed for pronunciation teaching

In the following sub-sections, three ASR-based pronunciation activities are presented. The activities were part of a set of seven activities proposed by Gottardi (2023) and intended for implementation in L2 English classes. We selected three activities to analyze their appropriateness according to the criteria of the frameworks proposed by Celce-Murcia, Brinton, and Goodwin (2010) and Chapelle (2001).

The activities were designed following the same structure as the activities by Stanley (2013); that is, starting with a table which provides the L2 teacher with a summary of the main pedagogical aspects of the activity, followed by the procedures to implement the activities, possible variations and an appendix containing a handout to give to the students. Each activity aims at a specific Common European Framework of Reference for Languages (CEFR) proficiency level range. For readability improvement, only the summary table is displayed during the pedagogical analyses. The complete version of the activities can be found in the appendices.

Every activity suggests a specific dictation passage to be used to interact with the selected ASR tool. The ASR tool employed during the activities was Google Translate's ASR feature (hereafter, GT). GT is a highly accessible tool available for free as a website or a mobile app, particularly popular in Brazil (Google, 2016). In addition, it is constantly improving and updating its speech features (Cardoso, 2022), which makes GT an adequate source of the latest improvements in ASR technology.

Considering the technological requirements, all three activities need the same infrastructure for their implementation: a digital device (e.g., personal computer, laptop, a smartphone with Google Translate's app installed), stable internet connection, and an input device (e.g., microphone, headphone, cellphone's microphone). Figure 1 below shows Google Translate's speech features. Before using the web version of GT, however, it is important to allow the browser to use computer's microphone. Also, it is necessary to choose the language to translate to, in this case, English (Google Translate, 2024).

1- "Translate by voice" button; that is, the ASR feature.
2- "Listen" button; that is, the Text-to-Speech feature.

Figure 1. Google Translate's features on a browser.
Source: The authors.

From a pedagogical viewpoint, each pronunciation activity explores different teaching techniques for L2 pronunciation practice and addresses different phases of the Communicative Framework for Teaching Pronunciation from Celce-Murcia et al. (2010). Additionally, the activities address different pronunciation target features, chosen based on recurring learning difficulties found among L2 Brazilian learners (Zimmer; Silveira; Alves, 2009) and typical intelligibility issues in their speech (Gonçalves; Silveira, 2015; Silveira; Alves; Gottardi, 2017). Lastly, the activities were designed bearing in mind the six Criteria for CALL Task Appropriateness from (Chapelle, 2001).

In summary, the rationale informing the design of the ASR-based pronunciation activities was provided. In this research context, learners' intelligibility issues and pronunciation difficulties were prioritized when designing the activities. The following sections present a more detailed pedagogical analysis of the activities.

4.1 Activity 1 – pronunciation self-assessment

The Pronunciation Self-Assessment activity (see Table 3 and Section A) brings a classic communicative activity that involves completing a worksheet containing short sentences for self-introduction. It is an opportunity for guided pronunciation practice because the worksheet brings sentence templates that learners are supposed to complete with personal information. Furthermore, it can be used as an initial step for learners to feel more confident to prepare to interact with a classmate, sharing their personal information and getting to know each other.

Table 3. Summary of Activity 1.

Learning Focus	Pronunciation self-assessment
Level	A1 – A2
Age	Teenagers or adults
Time	30-60 minutes
Target Feature	Raising awareness about pronunciation patterns that might disrupt communication.
Communicative Framework Phase	Description and Analysis (1) and Guided Practice (4) with an information-gap exercise.
Dictation Passage	Example: I am _____ [number] years old. (see Section A)

Source: The authors, based on Gottardi (2023).

The activity provides an opportunity for self-directed learning (Song; Hill, 2007) and for students to use digital resources to identify and try to improve pronunciation patterns that might hinder successful communication. In order to locate their pronunciation problems, learners are required to first write their responses (mostly single words or phrases) to complete the sentences. Next, they are instructed to voice-type each sentence at a time, copy the transcription provided by GT, and paste it next to each English sentence in the worksheet. Then, learners should compare their original sentences with GT's transcription to locate passages or words that might have been mistranscribed and figure out what caused this communication breakdown. At this point, it is recommended that learners also make use of digital resources (GT's text-to-speech feature or online dictionaries with audio features) to check the pronunciation of words that might have been mistranscribed by GT.

The activity is rather controlled, providing sentence templates to make it easier for teachers to use it with low-proficiency learners. There is no clearly defined pronunciation target, meaning that each learner will be able to locate possible pronunciation difficulties they have. Furthermore, students can write personalized sentences and later use all the sentences from the worksheet to interact with their classmates in a getting-to-know each other activity. In this sense, we expect that the Pronunciation Self-Assessment activity should provide opportunities for both focus on form and meaning.

In order to complete the activity, learners have to be skilled in performing basic computer tasks such as copying and pasting short texts, given that they need to copy the transcriptions from GT and paste them in the worksheet document. Alternatively, the teacher can ask students to write down the transcriptions in the space provided in the worksheet, but this would increase the duration of the activity. Because the activity requires manipulation of written language and concentration to analyze the GT's transcriptions, comparing them with the original sentence, teenage and adult learners are expected to manage well these tasks.

Voice-typing, use of GT or online dictionaries, and copying and pasting texts are all activities that are likely to be part of the daily routine of students. Likewise, exchanging personal information and introducing themselves are typical uses of language within and outside the classroom. Thus, we can expect that both instructors and learners will perceive that this activity relates to real-world

communicative situations to a certain extent.

The activity requires learners to self-assess their pronunciation, developing L2 pronunciation awareness and learning strategies, either by monitoring one's speech intelligibility when interacting with a machine or by attempting to change one's pronunciation to produce intelligible speech. Most likely, learners will require teachers' help to identify what pronunciation issues are hindering intelligibility. For this reason, it is paramount that teachers feel confident to provide solid explanations about mispronunciations and ways to improve speech intelligibility.

Since there are no pre-established pronunciation traits in this lesson, both teachers and students might feel overwhelmed with the task of locating the sources for mistranscription. In this sense, a viable strategy is to ask students to work in pairs after they locate the mistranscriptions and help each other identify what is causing communication problems. Then, teachers' help could be limited to the cases that the pairs could not solve.

The teacher's role is also important to orchestrate a class discussion about the most recurrent mistranscriptions and possible reasons for the ASR output, thus helping students to systematize the pronunciation traits that need more attention. Overall, the activity provides a good opportunity to raise awareness about frequent mispronunciations that hinder intelligibility and comprehensibility and ways to improve pronunciation. Confident teachers that believe in the importance of pronunciation teaching and have the necessary skills to teach pronunciation should have positive experience with the activity.

An internet-connected computer simplifies the task, allowing learners to use GT, copy transcriptions, and paste them into a digital worksheet easily. However, the activity can be implemented using a smartphone if a printed version of the worksheet is used and students are asked to handwrite GT's transcriptions into the appropriate spaces on it.

4.2 Activity 2 – vowel contrast

The Vowel Contrast activity (see Table 4 and Section B) uses ASR to create opportunities for production practice and reflection on pronunciation accuracy and intelligibility. The activity provides learners with opportunities to notice that English has a vowel contrast that is absent from their L1 (Brazilian Portuguese) and to practice pronouncing these vowels multiple times, with the goal of identifying the phonetic cues that distinguish these English phonemes. In that sense, there is a clear focus on form. However, a clear understanding of the forms being contrasted (the vowel pair /i/ and /ɪ/) will depend on the teacher's ability to explain the articulatory and phonetic traits that are essential to distinguishing the vowel pair, given that studies have shown that Brazilians have difficulties in both perceiving and producing these contrasts (Gonçalves; Silveira, 2014).

Table 4. Summary of Activity 2.

Learning Focus	Raising awareness about acoustic features of each vowel and how to distinguish them in production.
Level	A2 and above
Age	Teenagers or adults
Time	30 – 45 minutes
Target Feature	Producing the vowel contrast /ɪ/ and /i/
Communicative Framework Phase	Description and analysis (1) and Controlled Practice (3) with sentence-level minimal pairs.
Dictation Passage	
	Example: Look at that **sheep**! / Look at that **ship**! (see Section B)

Source: The authors, based on Gottardi (2023).

The activity relies on short sentences containing frequent words. Learners are required to voice-type each sentence at a time. They need to keep an eye on the computer or portable device screen to see how each sentence is transcribed in English. This should be a familiar task for most learners nowadays, with the widespread use of online translators as a self-learning tool. Perhaps learners are

more used to typing than voice-typing. However, voice recording is also a familiar activity performed outside the classroom, especially when using messenger applications (e.g., WhatsApp).

The activity requires well-developed literacy skills and concentration to interpret GT's feedback and inspect the output to see what might be causing intelligibility problems for the machine. Therefore, it seems appropriate for adults and maybe teenage learners.

Moreover, the activity is clearly focused on form, as its central goal is to call attention to vowel pronunciation patterns that might hinder speech intelligibility. Thus, language meaning is not the central issue, but one can say that meaning is addressed too, given that problems producing certain sounds are likely to make learners' speech less intelligible to machines and, possibly, to human listeners. The extent to which the activity will be seen as authentic will vary depending on the learners' use of voice-typing and translation tools outside the classroom.

The activity is expected to have a positive impact on pronunciation learning, as it was designed to help learners to notice a phonological aspect of the L2 that is important to convey meaning, thus improving speech intelligibility and comprehensibility. Regarding the teachers' perspective, provided they recognize the importance of pronunciation learning and are aware of the pronunciation trait being focused, they will regard the activity as a valuable pedagogical resource. Instructors' knowledge of phonetics and phonology, as well as pronunciation pedagogy, is essential for the effective use of this activity. If teachers lack such knowledge, there is a risk of frustration, as learners still need teachers' guidance to work on their pronunciation by producing and perceiving the vowel contrast, and monitoring their pronunciation to be understood by the ASR tool.

Turning to the practicality criteria, the activity requires few steps and is easy to implement, either on a computer or a portable device equipped with a screen and a stable internet connection. Sentences with minimal pairs can be made available in a digital format, be projected or written on a classroom board, or even printed. In short, instructors' knowledge is key for the success of this activity. Well-prepared instructors can use additional resources and/or briefly explain how to produce the vowel contrast.

4.3 Activity 3 – role-play activity

Role-plays are cornerstones of the Communicative Approach to Language Teaching (Littlewood, 2011) and are suitable for communicative pronunciation practice (Celce-Murcia; Brinton; Goodwin, 2010). In the Role-Play activity (see Table 5 and Section C), students are asked to simulate a conversation, using a prompt provided by the teacher. In this activity, students are expected to simulate a job interview.

Table 5. Summary of Activity 3.

Learning Focus	Speech rehearsal for fluency and accuracy improvement
Level	B2 and above
Age	Adults or late teenagers
Time	1 hour – 2 hours
Target Feature	Raising awareness about pronunciation patterns that might disrupt communication in a real-life situation
Communicative Framework Phase	Communicative Practice (5) with a role-play activity
Dictation Passage	Common questions in a job interview (see Section C).

Source: The authors, based on Gottardi (2023).

This activity requires that students work in pairs, and they can alternate the roles of the interviewer and the candidate, while using GT to help them monitor the intelligibility level of the speech produced by the student playing the role of the candidate. Thus, the activity also creates an opportunity for peer feedback and for using strategies to improve speech intelligibility, while interacting with a classmate and an ASR tool. Considering Celce-Murcia, Brinton, and Goodwin's (2010) pronunciation teaching framework, this activity focuses on communicative practice, given that, except for the questions

provided for the job interview, students will use their linguistic and communicative resources to produce speech and are expected to use the target language to discuss GT's orthographic feedback and ways to improve intelligibility when interacting with an ASR tool.

Considering the dialogical nature of role-plays, the activity is expected to provide plenty of opportunities for students to analyze their pronunciation and propose strategies to improve speech intelligibility. Given that GT will be transcribing spontaneous speech samples, unintelligible pronunciation patterns as well as technology issues are likely to be brought into discussion. Many of the activity's procedures demand maturity, mutual respect, and commitment from both students working together. Therefore, the activity may be more suitable for adult learners, especially if the role-play topic is a job interview. It is possible to adapt the topic (see Variation 3) to implement the activity with teenagers, provided they are used to working with peer feedback.

The activity enables students to focus on meaning while monitoring their pronunciation and analyzing the feedback provided by GT and by a peer. In this sense, it integrates focus on both meaning and form. Because the activity simulates a real-world situation (job interview), the chances of seeing it as an opportunity for authentic use of the L2 are high for both teachers and students. However, one might argue that in a real job interview, most candidates would not have the opportunity to practice, get feedback, and repeat the answers multiple times. On the other hand, with ASR-mediated practice, students are likely to connect the Role-Play activity with real-life interactions with ASR tools, facing challenges in conveying their message, whether due to pronunciation difficulties or technological issues (e.g., bad internet connection, ASR limitations, noisy environment).

As the activity requires producing non-scripted speech, it allows students to use their full L2 repertoire to play the role of the candidate. Provided the learners interact in the L2 (which is sometimes challenging when they share the same L1), the moments of discussion involving the GT transcript should be a great opportunity for learners to develop their strategies to monitor and practice L2 pronunciation, ask teachers for explicit pronunciation instruction, improve their L2 pronunciation awareness and knowledge, and create strategies to produce speech with increased intelligibility.

Considering the role of teachers, it is expected that students with little familiarity with ASR tools or with pronunciation patterns that frequently hinder intelligibility will require substantial support to perform the activity. If working with large groups, it might become difficult to provide individualized feedback to each pair. Then, it is a good idea to ask each pair to select the transcript of a single answer to have teachers' feedback, so that neither students become frustrated, nor teachers are overwhelmed. In this sense, it is important to make it clear for students that there is no need to understand all of the pronunciation issues that are causing intelligibility problems, neither practice each response more than two or three times.

Among the three activities suggested in this article, this is the one that requires more attention to the equipment available to students, thus being less practical, especially if being used in regular classrooms. The ideal conditions are placing learners in a classroom equipped with a desktop computer or a laptop, a headset, stable internet connection, and a quiet environment. These are important conditions for better ASR performance, especially considering that students (especially job candidates) will be interacting and producing non-scripted speech. Having a large computer screen would also make it easier for the interviewer to share the GT's transcripts with the candidate and copy and paste them in another file to be shared with both candidate and teacher. Lastly, learners must be familiar with handling these computer tasks in order to perform this activity.

5 Discussion and final remarks

This pedagogical paper presented three ASR-based pronunciation activities based on Gottardi (2023) and offered an analysis of ASR use for L2 pronunciation teaching and learning. The activities were designed to have a clear learning focus and the possibility to focus on linguistic form, which aligns with Chapelle (2001) suggestion of prioritizing the *language learning potential* criterion while designing CALL tasks.

The activities were designed to offer a certain degree of adaptability. Each activity suggests variations for teachers and learners to offer a more appropriate learning experience, allowing the adap-

tation of the procedures according to the teaching environment and learners' individual characteristics. Therefore, different strategies to use ASR for pronunciation practice can be employed, fostering learners' autonomy. In addition, one advantage of ASR-based pronunciation practice is that learners can work at their own pace, adjusting their speech rate as needed. Lastly, by modifying the dictation passage, the activities can be used to practice different L2s as long as they are supported by the ASR program, and that the professional designing the material has knowledge about L2 pronunciation patterns that are likely to hinder intelligibility and comprehensibility. For example, for a group of Spanish learners who aim to improve their pronunciation when speaking Brazilian Portuguese, Activity 2 could include words containing the Portuguese phonemes /s/ and /z/, or even the contrast between /e/ and /ɛ/.

Moreover, the activities were designed to improve speech intelligibility aiming at successful communication, which benefit out-of-class spoken interactions. Overall, the activities provide Brazilian learners of English with opportunities to practice words and phrases that may cause intelligibility issues and hinder communication, offering extended opportunity for output practice and interaction during pair and group work.

However, ASR limitations can be challenging when used pedagogically, particularly regarding noise control, since ASR systems tend to perform better in a quiet room than in a noisy environment (Jurafsky; Martin, 2024; O'Shaughnessy, 2024). This is notably difficult when considering an L2 classroom context. Nevertheless, this issue can be mitigated by dividing the class into groups. While some groups are working on a different task that does not demand much oral interaction, the other groups can perform the ASR-based activities proposed in this paper. Then, with the mediation of the teacher, the groups take turns to perform different roles or tasks. This can reduce concurrent speech during the ASR practice. Lastly, to avoid reducing ASR accuracy due to noise interference, we also suggest considering a quieter environment available at school (e.g., a computer laboratory or a study hall).

According to the literature, other possible limitations include limited microphone access, limited and/or unstable internet, and difficulty interpreting ASR feedback for speech intelligibility improvement (Gottardi; Almeida; Tumolo, 2022; Gottardi, 2023). Therefore, as an attempt to circumvent those limitations and benefit from ASR's affordances, teacher's guidance still plays a pivotal role for the successful implementation of ASR for pronunciation teaching. It includes preparing the teaching environment, providing appropriate instructions, and designing and adapting pedagogically sound CALL activities (Gottardi, 2023; Gottardi; Almeida; Tumolo, 2022; Liakina; Liakin, 2022; Silveira; Zanchet; Pereira, 2022). Thus, additional studies proposing such activities and ways to mitigate ASR's limitations may help teachers to better leverage ASR for pronunciation teaching.

We acknowledge that our analysis would benefit from actual implementation of the activities in real classrooms, so that we could gather teachers' and students' perspectives about the suitability of the activities for pronunciation learning. This is certainly a very interesting avenue for further studies and vital for us to appraise the *language learning potential* criterion of ASR-based pronunciation activities and the affordances and limitations of ASR for L2 pronunciation practice.

Another promising line of investigation is the use of ASR embedded in AI-powered assistants. These virtual assistants might have the potential to add more interactive elements during the activities as well as simplify some procedures (during the role-play activity, for example). Therefore, a closer look at the affordances and limitations of these tools for pronunciation teaching should be taken, especially considering that "AI will become an increasingly present feature in our (and especially our students') lives; we cannot ignore it as an aspect of language use in society" (Kern, 2024, p. 530).

In conclusion, the activities not only suggest the use of ASR as a pronunciation teaching resource but also offer a practical way of implementing it in an L2 classroom. Thus, the activities demonstrate the use of a digital resource with the potential to assist L2 pronunciation practice, rather than activities that use a digital resource merely because it is trendy. In other words, pedagogical principles should come first when adopting and adapting CALL resources for L2 teaching and learning. This applies to ASR tools, or any other digital technology used for these purposes.

6 Acknowledgments

References

BALDISSERA, Luana Garbin; TUMOLO, Celso Henrique Soufen. Apps for developing pronunciation in English as an L2. *Revista X*, v. 16, n. 5, p. 1355, 2021.

BURNS, Anne; SEIDLHOFER, Barbara. Speaking and pronunciation. *In:* SCHMITT, Norbert (ed.). *An introduction to applied linguistics*. 3. ed. London: Routledge, 2020. p. 240–258.

CARDOSO, Walcir. English syllable structure. *In:* KANG, Okim; THOMSON, Ron; MURPHY, John (eds.). *The Routledge Handbook of Contemporary English Pronunciation*. 1. ed. New York: Routledge, 2017. p. 122–136.

CARDOSO, Walcir. Technology for speaking development. *In:* DERWING, Tracey M.; MUNRO, Murray J.; THOMSON, Ron I. (eds.). *The Routledge Handbook of Second Language Acquisition and Speaking*. 1. ed. London: Routledge, 2022. p. 299–313.

CELCE-MURCIA, Marianne; BRINTON, Donna M.; GOODWIN, Janet M. *Teaching pronunciation: A course book and reference guide*. 2. ed. Cambridge: Cambridge University Press, 2010. Hardback with audio CDs (2).

CHAPELLE, Carol. *Computer applications in second language acquisition*. Cambridge: Cambridge University Press, 2001.

CHAPELLE, Carol. Evaluation of technology and language learning. *In:* CHAPELLE, Carol; SAURO, Shannon (eds.). *The handbook of technology and second language teaching and learning*. 1. ed. Hoboken, NJ: John Wiley & Sons, Inc., 2017. p. 378–392.

CHUN, Dorothy; SMITH, Bryan; KERN, Richard. Technology in Language Use, Language Teaching, and Language Learning. *The Modern Language Journal*, v. 100, S1, p. 64–80, 2016. DOI: 10.1111/modl.12302.

CHUN, Dorothy M. Computer-assisted pronunciation teaching. *In:* CHAPELLE, Carol A. (ed.). *The Concise Encyclopedia of Applied Linguistics*. [*S. l.*]: Wiley Online Library, 2020. p. 213–223. DOI: 10.1002/9781119144038.ch24.

COLLINS, Laura; MUÑOZ, Carmen. The Foreign Language Classroom: Current Perspectives and Future Considerations. *The Modern Language Journal*, v. 100, S1, p. 133–147, 2016. DOI: 10.1111/modl.12305.

DAVIES, Graham. Computer-Assisted Language Education. *In:* BERN, Margie (ed.). *Concise encyclopedia of applied linguistics*. Amsterdam: Elsevier, 2006. p. 261–271.

DERWING, Tracey M.; MUNRO, Murray J. *Pronunciation Fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins Publishing Company, 2015. DOI: 10.1075/lllt.42.

FOOTE, Jennifer A.; TROFIMOVICH, Pavel. Second language pronunciation learning: An overview of theoretical perspectives. *In:* KANG, Okim; THOMSON, Ron; MURPHY, John (eds.). *The Routledge handbook of contemporary English pronunciation*. London: Routledge, 2017. p. 75–90.

GONÇALVES, Alison Roberto; SILVEIRA, Rosane. A produção das vogais altas anteriores do inglês por aprendizes brasileiros. *Fórum Linguístico*, v. 11, p. 9–22, 2014. DOI: 10.5007/1984-8412.2014v11n1p9.

GONÇALVES, Alison Roberto; SILVEIRA, Rosane. Intelligibility research in Brazil: empirical findings and methodological issues. *Revista Horizontes de Linguística Aplicada*, v. 14, n. 1, p. 51–81, 2015.

GOOGLE. *Ten years of Google Translate*. [*S. l.: s. n.*], 2016. Available from: https://blog.google/products/translate/ten-years-of-google-translate/. Visited on: 14 Mar. 2025.

GOOGLE TRANSLATE. *Translate by speech - Computer - Google Translate Help*. [*S. l.: s. n.*], 2024. Available from: https://support.google.com/translate/answer/6142468?hl=en&ref_topic=7011659. Visited on: 14 Mar. 2025.

GOTTARDI, William. *Automatic Speech Recognition as a Pronunciation Teaching Resource: What Are the In-Service Teachers' Perceptions of This Speech Technology?* Mar. 2023. MA thesis – Federal University of Santa Catarina (UFSC), Florianópolis.

GOTTARDI, William; ALMEIDA, Janaina Fernanda de; TUMOLO, Celso Henrique Soufen. Automatic speech recognition and text-to-speech technologies for L2 pronunciation improvement: reflections on their affordances. *Texto Livre*, v. 15, 2022. DOI: 10.35699/1983-3652.2022.37939.

GOTTARDI, William; SILVEIRA, Rosane. Automatic Speech Recognition as a Pronunciation Teaching Resource: In-Service Teachers' Perceptions. *In:* PSLLT 14 - The Practice of Pronunciation. Ames, IA: Iowa State University Digital Press, 2024. Available from: https://www.iastatedigitalpress.com/psllt/article/id/17567/. Visited on: 14 Mar. 2025.

IBM. *What Is a Chatbot?* [*S. l.: s. n.*], 2024. Available from: https://www.ibm.com/topics/chatbots. Visited on: 14 Mar. 2025.

JORDÃO, Clarissa M. ILA-ILF-ILE-ILG: Quem dá conta? EAL-ELF-EFL-EGL: Same Difference? *Revista Brasileira de Linguística Aplicada*, v. 14, p. 13–40, 2014. DOI: 10.1590/s1984-63982014005000002.

JURAFSKY, Daniel; MARTIN, James H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3. ed. [*S. l.: s. n.*], 2024. Available from: https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf. Visited on: 14 Mar. 2025.

KERN, Richard. Twenty-first century technologies and language education: Charting a path forward. *The Modern Language Journal*, v. 108, n. 2, p. 515–533, 2024. DOI: 10.1111/modl.12929.

KIVISTÖ-DE SOUZA, Hanna; GOTTARDI, William. How well can ASR technology understand foreign-accented speech? *Trabalhos em Linguística Aplicada*, v. 61, n. 3, p. 764–781, 2022. DOI: 10.1590/010318135124420221123.

LAI, Chun. *Autonomous language learning with technology: Beyond the classroom*. London: Bloomsbury Publishing, 2018.

LANE, Linda; BROWN, H. Douglas. *Tips for teaching pronunciation: A practical approach*. New York: Pearson Longman, 2010.

LEVIS, John. Revisiting the Intelligibility and Nativeness Principles. *Journal of Second Language Pronunciation*, v. 6, n. 3, p. 310–328, 2020. DOI: 10.1075/jslp.20038.lev.

LEVIS, John; SUVOROV, Ruslan. Automatic Speech Recognition. *In:* CHAPELLE, Carol A. (ed.). *The Concise Encyclopedia of Applied Linguistics*. [*S. l.*]: Wiley Online Library, 2020. p. 149–156. DOI: 10.1002/9781119144038.ch15.

LIAKINA, Natallia; LIAKIN, Denis. Speech technologies and pronunciation training: What is the potential for efficient corrective feedback? *In:* LEVIS, John M.; DERWING, Tracey M. (eds.). *Second Language Pronunciation: Different Approaches to Teaching and Training*. Berlin: De Gruyter Mouton, 2022. p. 287–312.

LIESHOUT, Catharina van; CARDOSO, Walcir. Google Translate as a Tool for Self-Directed Language Learning. *Language Learning & Technology*, v. 26, n. 1, p. 1–19, 2022.

LITTLEWOOD, William. Communicative language teaching: An expanding concept for a changing world. *In:* HINKEL, Eli (ed.). *Handbook of Research in Second Language Teaching and Learning, Volume II*. New York: Routledge, 2011. p. 541–547.

MARTINS, Claudia Beatriz Monte Jorge; MOREIRA, Herivelto. O campo CALL (Computer Assisted Language Learning): definições, escopo e abrangência. *Calidoscópio*, v. 10, n. 3, p. 247–255, 2012. DOI: 10.4013/cld.2012.103.05.

NICKOLAI, Dan; SCHAEFER, Emma; FIGUEROA, Paula. Aggregating the evidence of automatic speech recognition research claims in CALL. *System*, v. 121, 2024. DOI: 10.1016/j.system.2024.103243.

O'BRIEN, Mary Grantham; DERWING, Tracey M.; LEVIS, John M.; SARDEGNA, Veronica G. Directions for the future of technology in pronunciation research and teaching. *Journal of Second Language Pronunciation*, v. 4, n. 2, p. 182–207, 2018. DOI: 10.1075/jslp.17001.obr.

O'SHAUGHNESSY, Douglas. Trends and developments in automatic speech recognition research. *Computer Speech & Language*, v. 83, 2024. DOI: 10.1016/j.csl.2023.101558.

OLSON, Daniel J. Linking Linguistic Theory and Second Language Pedagogy: An Example from Second Language Pronunciation and Phonetics. *In:* SHEU, Vanessa; ZHOU, Alexis; WEIRICK, Joshua D. (eds.). *Variation in linguistics: second language acquisition, discourse studies, sociolinguistics, syntax*. Newcastle-upon-Tyne: Cambridge Scholars Publishing, 2023. p. 26–45.

PENNINGTON, Martha C.; ROGERSON-REVELL, Pamela Mary. *English pronunciation teaching and research*. London: Palgrave Macmillan, 2019. v. 10, p. 978–988.

ROGERSON-REVELL, Pamela Mary. Computer-Assisted Pronunciation Training (CAPT): Current Issues and Future Directions. *RELC Journal*, v. 52, n. 1, p. 189–205, 2021. DOI: 10.1177/0033688220977910.

SANTOS, Fhelippe Waltelon Souza dos; KIVISTÖ-DE SOUZA, Hanna. Examining Vowel Intelligibility in Brazilian Portuguese EFL Learners: A Study Using Automatic Speech Recognition. *Ilha do Desterro*, v. 77, 2024. DOI: 10.5007/2175-8026.2024.e95566.

SILVEIRA, Rosane; ALVES, Ubiratã Kickhöfel; GOTTARDI, William. Percepção, produção e inteligibilidade do inglês falado por usuários brasileiros. *In:* PERSPECTIVAS atuais de aprendizagem e ensino de línguas. Florianópolis: Lle/cce/ufsc, 2017. p. 237–283.

SILVEIRA, Rosane; ZANCHET, Crystoffer E.; PEREIRA, Manuella Hinchingr. Affordances of Digital Technology for English Pronunciation Teaching: The Perspective of Brazilian Teachers. *Veredas*, v. 26, p. 1–25, 2022. DOI: 10.34019/1982-2243.2022.v26.37960.

SOLEIMANI, Hassan. *Computer Assisted Language Learning: Theory and Practice*. Tehran: Payame Noor University, 2021.

SONG, Liyan; HILL, Janette. A Conceptual Model for Understanding Self-Directed Learning in Online Environments. *Journal of Interactive Online Learning*, v. 6, n. 3, p. 27–42, 2007.

STANLEY, Graham. *Language learning with technology: Ideas for integrating technology in the classroom*. Cambridge: Cambridge University Press, 2013.

THOMSON, Ron I.; DERWING, Tracey M. The Effectiveness of L2 Pronunciation Instruction: A Narrative Review. *Applied Linguistics*, v. 36, n. 3, p. 326–344, 2015. DOI: 10.1093/applin/amu076.

UNESCO. *Global Education Monitoring Report 2023: Technology in education – A tool on whose terms?* 1. ed. Paris: Unesco, 2023.

ZIMMER, Márcia; SILVEIRA, Rosane; ALVES, Ubiratã Kickhöfel. *Pronunciation Instruction for Brazilians: Student's Book*. Newcastle-upon-Tyne: Cambridge Scholars Publishing, 2009.

Author contributions

William Gottardi: Conceptualization, Investigation, Writing – original draft, Writing – review and editing; **Rosane Silveira**: Conceptualization, Formal analysis, Writing – original draft.

A Activity 1 – Pronunciation Self-assessment

Learning Focus	Pronunciation self-assessment
Level	A1 – A2
Age	Teenagers or adults
Time	30–60 minutes
Target Feature	Raising awareness about pronunciation patterns that might disrupt communication.
Communicative Framework Phase	Description and Analysis (1) and Guided Practice (4) with an information-gap exercise.
Dictation Passage	Activity's Appendix (Personal Information Sheet). Examples: • I am _____ [number] years old. • My favorite food is _____ [food]. • My favorite drink is _____ [drink].

Procedure:

1. Instruct students to access Google Translate (GT) and give them the Personal Information Sheet (Appendix – Personal Information Sheet).
2. Tell students to fill in the gaps with their personal information. A hint word is beside each gap to help students understand the kind of information that is required. Motivate students to ask questions regarding vocabulary at this step so they can focus on pronunciation next.
3. Then, ask students to voice type each sentence from the Personal Information Sheet at a time and write down the resulting transcription provided by GT in the second column of the sheet. They can repeat this procedure if some interference has happened during the voice typing.
4. Encourage students to underline and take notes of any words and phrases that GT transcribed differently from their original sentence.
    • During this process, students may require assistance in improving their pronunciation to be understood by the application. GT's text-to-speech feature or an online dictionary (www.dictionary.cambridge.org) can serve students as the target model for the words that were not accurately transcribed by GT.
5. After completing all the rows, students can practice the most difficult sentences or create new sentences about their personal information using the last two rows.

Variation: this activity can be used as homework. However, it may be necessary to suggest the usage of GT's text-to-speech feature or an online dictionary to serve students as a model for difficult words.

Activity's Appendix – Personal Information Sheet

Sentence	Google Translate Transcription
I am _____ [number] years old.	
I have _____ [number] brother(s). / I don't have a brother.	
I have _____ [number] sister(s). / I don't have a sister.	
My house is _____ [color].	
My favorite color is _____ [color].	
I have a pet. My pet is a(n) _____ [animal]. / I don't have a pet.	
My favorite food is _____ [food].	
My favorite drink is _____ [drink].	
?	
?	

B Activity 2 – Vowel Contrast

Learning Focus	Raising awareness about acoustic features of each vowel and how to distinguish them in production.
Level	A2 and above
Age	Teenagers or adults
Time	30–45 minutes
Target Feature	Producing the vowel contrast /ɪ/ and /i//
Communicative Framework Phase	Description and analysis (1) and Controlled Practice (3) with sentence-level minimal pairs.
Dictation Passage	1. Look at that sheep! / Look at that ship! 2. I may live. / I may leave. 3. It is a sin. / It is a scene. 4. Can you see the beach? / Can you see the bitch? 5. How can I spell "feel"? / How can I spell "fill"? 6. What is the definition of "sit"? / What is the definition of "seat"?

Procedure:
1. Instruct students to access Google Translate (GT) and show them the dictation passage. Ask students to make sure the input language is English. They can also set the output language to be English or their L1.
2. Ask students to voice type each sentence from the dictation passage and pay close attention to the words in bold (minimal pairs). Explicit instruction on how the two vowels differ acoustically and articulatorily may be necessary at this point to help students raise awareness regarding these contrasting sounds.
3. Encourage students to repeat the voice typing if GT transcribes the word in bold differently from the original sentence.
4. Tell students they can listen to the target words (in bold) using GT's text-to-speech feature or an online dictionary and try to imitate the sounds during their production.
5. The teacher asks students about difficulties they had with task and, if needed, provides explicit instruction on how to produce the vowel contrast.

  Variation 1: this activity can be used as homework. A video can be used to give explicit instructions regarding the vowel contrast. Example: https://youtu.be/FYI6Vt3uq7s.

Variation 2: different segmental contrasts can be practiced using these procedures. However, the sentences and the minimal pairs should be carefully tested using GT before suggesting them to the students.

C Activity 3 – Role-play Activity

Learning Focus	Speech rehearsal for fluency and accuracy improvement
Level	B2 and above
Age	Adults or late teenagers
Time	1 hour – 2 hours
Target Feature	Raising awareness about pronunciation patterns that might disrupt communication in a real-life situation
Communicative Framework Phase	Communicative Practice (5) with a role-play activity
Dictation Passage	Suggested questions for the job interview: 1. Could you tell me a little about yourself? 2. What is your greatest personal achievement? 3. What about your professional achievement? 4. What qualities make a reliable leader? 5. What are your biggest weaknesses? 6. What are your biggest strengths? 7. What happened last time a customer or co-worker got angry with you? 8. What would be your dream job? 9. What do you know about our industry? 10. Why should we hire you?

Procedure:

1. Introduce the topic of the activity by asking students to think about their dream job.
2. Tell them that they will perform a role-play activity in pairs. One student will be the *candidate* for a job application and the other will be the *interviewer*.
3. Divide the group into pairs and ask ONE student of the pair (interviewer) to access Google Translate (GT).
4. Then, give each pair the *Job Interview Questions*.
5. The pairs should practice each question at a time. The interviewer asks one question and, before the candidate starts answering it, the interviewer should click on the GT's voice typing feature. The interviewer should not show the screen to the candidate since it may disrupt his/ her attention to the activity.
6. After GT finish transcribing *the answer*, the interviewer should copy the resulting transcription and paste it to another document or send it to the candidate via an instant messaging program so students can check the transcription when the activity is over ,that is, all questions were answered.
7. Students should repeat the previous step until all the questions have been answered.
8. Finally, each pair will analyze GT's transcriptions and take notes about how intelligible the answers were and try to figure out what may have caused the mistranscriptions (ex.: technology issues, pronunciation patterns). Encourage the interviewer to give honest feedback to the candidate on whether he/she agrees with GT's mistranscribed passages (according to the candidate's communicative intention) or not.
9. For the unintelligible parts of the speech, the candidate can ask the teacher for explicit instruction on how to overcome the communication breakdown.
10. Finally, the candidate can practice the question that he/she found more difficult to answer by voice-recording it again.
11. Students can exchange roles and start the activity again.

12. Once students finish rehearsing, they can present the dialog to the class and discuss their answers with their classmates.

Variation 1: this activity can be used as homework individually. Students can practice their own answers to the job interview questions and take notes about any difficult words or unintelligible parts of their speech to GT. Then, students should share their notes with the teacher so he/ she can help them overcome possible communication breakdowns.

Variation 2: this activity can be used as homework in pairs using a video conference program, which would make it easier to share the screen with the transcriptions. Students should take notes and share them with the teacher so he/ she can help them overcome the communication breakdowns that have happened.

Variation 3: this role-play activity can be used for different contexts. The questions provided by the teacher can simulate a conversation between a salesperson and a customer, two people asking/giving directions, or ordering food in a restaurant, for example. In these variations, it would be easier to work with teenage students as well.