

AUTOMATIC SPEECH RECOGNITION AS A PRONUNCIATION TEACHING RESOURCE: IN-SERVICE TEACHERS' PERCEPTIONS

William Gottardi, Federal University of Santa Catarina – Brazil / Coordination for the Improvement of Higher Education Personnel (CAPES)

Rosane Silveira, Federal University of Santa Catarina – Brazil / National Council for Scientific and Technological Development (CNPq)

Digital technologies may aid pronunciation teaching, especially Automatic Speech Recognition (ASR). This study sought to explore the affordances of ASR for pronunciation teaching from the perspective of in-service English teachers and investigate teachers' appraisal of ASR-based pronunciation activities designed to be implemented in second language English classes. To achieve these objectives, this research followed a mixed-method approach. An online workshop session was delivered to two different groups of participants in order to gather participants' perceptions of ASR technology for pronunciation teaching. In addition, seven ASR-based pronunciation activities were made available for their appraisal. A total of twelve teachers participated in this study. Data from an online background questionnaire and an online survey were analyzed quantitatively. The overall perceptions of the participant-teachers indicate that ASR can 1) be used as a tool for self-studying; 2) help encourage learners to produce more output outside the classroom; 3) provide students with relevant orthographic feedback; 4) serve as a supplementary resource beyond the classroom; and 5) be an adequate auxiliary resource for pronunciation teaching in regular classes, especially, in a hybrid environment.

INTRODUCTION

Digital technologies demonstrate great potential for both pronunciation teaching and learning (Revell-Rogerson, 2021). Automatic Speech Recognition (ASR) technology, for example, can facilitate pronunciation improvement and provide learners with instant feedback (Golonka et al., 2014). This technology is now available for free as a built-in feature in varied websites and programs. Simply put, ASR can transcribe speech (oral input) into word sequences (written output) (Yu & Deng, 2015). An exemplary application of ASR technology is its implementation in dictation tools that offer flexibility in usage (McCrocklin, 2019). This flexibility may represent an opportunity to circumvent the time constraint and the lack of resources that may hinder pronunciation teaching in second language (L2) English classes.

ASR programs are especially suitable for learners with little access to the target language outside the classroom since they offer an opportunity for learners to produce more output with instant orthographic feedback (Liakin et al., 2017) and can foster learner autonomy (Liakin et al., 2017; McCrocklin, 2016; Mroz, 2018). Moreover, ASR technology is particularly important due to its potential to aid pronunciation learning and teaching as resource both in and out of the classroom

(Cardoso, 2022). All in all, much research has been conducted trying to comprehend the affordances of ASR for pronunciation teaching (see Gottardi et al., 2022).

Regarding pronunciation teaching, Pennington and Rogerson-Revell (2019) point out that research studies usually offer insufficient information related to how teachers can adopt the study's suggestions into their practices, considering their teaching context. In addition, previous studies indicate that pronunciation teaching is a major gap in teacher education, indicating that there might be a lack of knowledge about what to teach and how to teach pronunciation (Baker, 2014; Buss, 2013, 2016; Costa, 2016; Silveira et al., 2022). Taking this into consideration, there might be a mismatch between teacher education, research, and pronunciation teaching. This research sought to contribute towards understanding this mismatch and propose pedagogical uses of ASR for L2 pronunciation teaching.

Bearing in mind the abovementioned considerations, this study endeavored to explore the affordances of ASR for pronunciation teaching from the perspective of in-service English teachers and to investigate teachers' appraisal of ASR-based pronunciation activities designed to be implemented in L2 English classes. Thus, this study intends to provide teachers with relevant information regarding ASR technology applied to pronunciation teaching so they can decide what path they want to follow to achieve their pedagogical goals.

Research Questions

RQ1: What are the in-service English teachers' perceptions of ASR for pronunciation teaching, and how does a workshop on how to use this speech technology influence their perceptions regarding the following constructs: teacher development needs, ASR accessibility, and ASR affordances?

RQ2: How do in-service teachers appraise the ASR-based pronunciation activities designed to be implemented in L2 English classes?

METHODS

Participants

Participants were recruited online via e-mail and social media. A total of twelve in-service English teachers participated in this research. Participants' ages ranged from 24 to 42 ($M = 32.2$, $SD = 5.6$); 41.6% male ($n = 5$) and 58.3% female ($n = 7$), and their experience teaching English, in years, ranged from 4 to 16 ($M = 10.6$, $SD = 3.89$). With respect to sectors where the participant-teachers worked, 50% worked in the private sector ($n = 6$); 33.3% in the public sector ($n = 4$), and 16.6% in both sectors ($n = 2$). All participants held an undergraduate degree; 83.3% ($n = 10$) related to English teaching, and the other 16.6% ($n = 2$) in a different area.

Procedures

This research followed a mixed-method approach. In order to answer the research questions, an online workshop session was designed and delivered to the participant-teachers presenting an

overview of the literature on pronunciation teaching and ASR, along with seven original ASR-based pronunciation activities.

The background information of the participants was collected via an online background questionnaire. Participants' perceptions of the affordances and limitations of the ASR technology as well as their appraisal of the ASR-based pronunciation activities were collected via an online survey. Data from the background questionnaire and survey were mostly analyzed quantitatively except for one open-ended question from the survey. Two separate workshop sessions were conducted online via Zoom Meetings. Five participants joined the first session and seven joined the second. The duration of each session varied from two to three hours. Both workshop sessions were recorded and transcribed, and the chat logs saved. These resources were used for qualitative analysis.

Background Questionnaire

A background questionnaire was shared online with the participants to support the data analyses and to guarantee that participants matched the criteria of this research. The questionnaire was created using Google Forms. It was designed to obtain the demographic characteristics of the participants (e.g., gender, age, residential location, occupation) and find out about the participants' experience with digital technologies and pronunciation teaching through 10-point Likert-scale behavioral questions (see Table 2 in Results and Discussion section).

Survey

An online survey was created using Google Forms and shared with the participants at the end of the workshop session. Each survey item was carefully designed to address a particular research question. The items were divided into four sections. Sections 1, 2, and 3 addressed a particular construct from RQ1. Section 1 (Teacher Development Needs) was composed of 3 items; section 2 (ASR Accessibility), 4 items; and section 3 (ASR Affordances), 6 items. Section 4, composed of 14 items, addressed RQ2.

Most of the items were Likert-scale questions ranging from 0 (strongly disagree) to 10 (strongly agree). There were also some multiple-choice questions and a single open-ended question that inquired about the workshop session itself. The survey items from section 4 that inquired about the design, suitability, and adaptability of the ASR-based pronunciation activities presented during the workshop session were designed based on Jamieson and Chapelle's (2010) survey items which were "logically linked to the six criteria for CALL evaluation" (p. 362) from Chapelle (2001).

Workshop Design

The workshop sessions followed a specific program. During the workshop, an overview of relevant topics was provided to the participants, namely, pronunciation teaching issues, Computer-Assisted Language Learning (CALL), ASR technology, and the rationale behind the ASR-based pronunciation activities designed by the authors.

Then, instructions on how to use Google Translate's (GT) ASR feature were provided to the participants. GT was the ASR tool chosen to be used during the workshop mainly because it is available for free as a website or a mobile app. In addition, Brazilian teachers may have a better understanding of Google products for English teaching and learning compared to other digital tools (Junior et al., 2022). Finally, participant-teachers were able to use GT on their own by performing the ASR-based pronunciation activities during the hands-on portion of the session.

ASR-based Pronunciation Activities

Seven ASR-based pronunciation activities were presented to the participants during the workshop. The activities followed the *Criteria for CALL Task Appropriateness* (CCTA) from Chapelle (2001). The six criteria are 1) language learning potential (beneficial focus on form); 2) learner fit (respect learner's individual characteristics); 3) meaning focus (learners' attention is guided to the meaning of the language); 4) authenticity (correspondence between the activity and what learners might see out of the classroom); 5) positive impact (positive effects of the activity on the learner); and 6) practicality (adequacy of resources to implement the activity considering the particular constraints of the class).

In addition, the activities addressed different phases (3, 4, and 5) of the *Communicative Framework for Teaching Pronunciation* (CFPT) from Celce-Murcia et al. (2010). Activity 1 focused on guided practice (4th phase), activities 2-6 focused on controlled practice (3rd phase), while activity 7 focused on communicative practice (5th phase). Lastly, the pronunciation target features that each activity addressed were based on common intelligibility issues presented on L2 Brazilian learners' speech (Gonçalves & Silveira, 2015; Silveira et al., 2017) and on different learning difficulties faced by L2 Brazilian learners (Zimmer et al., 2009). Table 1 summarizes all the activities by presenting the expected proficiency level of the learners, the CFPT stage, and the learning focus of the activity.

Table 1

Summary of the Seven ASR-Based Pronunciation Activities

Activity Name	CEFR Level	CFTP Phase	Learning Focus
1 – Pronunciation Self-assessment	A1 – A2	Guided Practice	Pronunciation self-assessment
2 – Vowel Contrast	A2 and above	Controlled Practice	Raising awareness about the acoustic features of each vowel and how to distinguish them in production
3 – Shadow Reading	B2 and above	Controlled Practice	Practicing speech fluency and accuracy
4 – Paragraph-reading Task	B1 – B2	Controlled Practice	Perceiving and producing the different -ed pronunciations of regular verbs in the simple past tense.
5 – Tongue Twisters	All levels	Controlled Practice	Pronunciation accuracy and fluency
6 – Monitoring Worksheet	All level	Controlled Practice	Pronunciation self-assessment and monitoring skills development
7 – Role-play Activity	B2 and above	Communicative Practice	Speech rehearsal for fluency and accuracy improvement

As Table 1 demonstrates, most of the activities are categorized as “Controlled Practiced”, which is easier to implement with ASR technology (Gottardi et al., 2022; Souza & Gottardi, 2022).

Data Analyses

In order to answer both RQs, data were gathered from the background questionnaire and the online survey and then analyzed quantitatively. The open-ended question from the online survey was analyzed qualitatively. For the quantitative analysis, descriptive statistics were run, and the standard deviation (SD) and the means (M) were calculated. After the data were cleaned and grouped, they were coded into four different categories: 1) Teacher Development Needs; 2) ASR

Accessibility; 3) ASR Affordances; and 4) Activities Appraisal. The results of these data analyses procedures are presented and discussed next.

RESULTS AND DISCUSSION

Teacher Development Needs

Data regarding the participants' experience with digital technologies and pronunciation teaching were gathered from the background questionnaire before the workshop session. Answers ranged from 0 (not much) to 10 (very much). As Table 2 displays, the lower scores were related to the use of Google Translate as an English teaching resource and the use of ASR tools as a pronunciation teaching resource.

Table 2

Summary of the Responses about Participants' Teaching Practices

Background Questionnaire Question	M (SD)
How confident are you in teaching pronunciation?	7.9 (1.2)
How often do you teach pronunciation in your classes?	8.0 (2.0)
How confident are you in using digital resources (computer, cell phone, projector, websites, apps) in your pedagogical practices?	8.3 (2.1)
How often do you use digital resources in your classes?	8.9 (1.4)
How often do you use Google Translate in your classes?	2.8 (3.2)
How often do you use any Automatic Speech Recognition (ASR) tool as a resource for teaching pronunciation?	2.0 (3.0)

Data from the online survey administered after the workshop session show that participants wish for more professional development opportunities related to teaching techniques and digital resources for pronunciation teaching as Figure 1 shows. These results suggest that teachers would benefit more from practical suggestions than theory on teaching pronunciation.

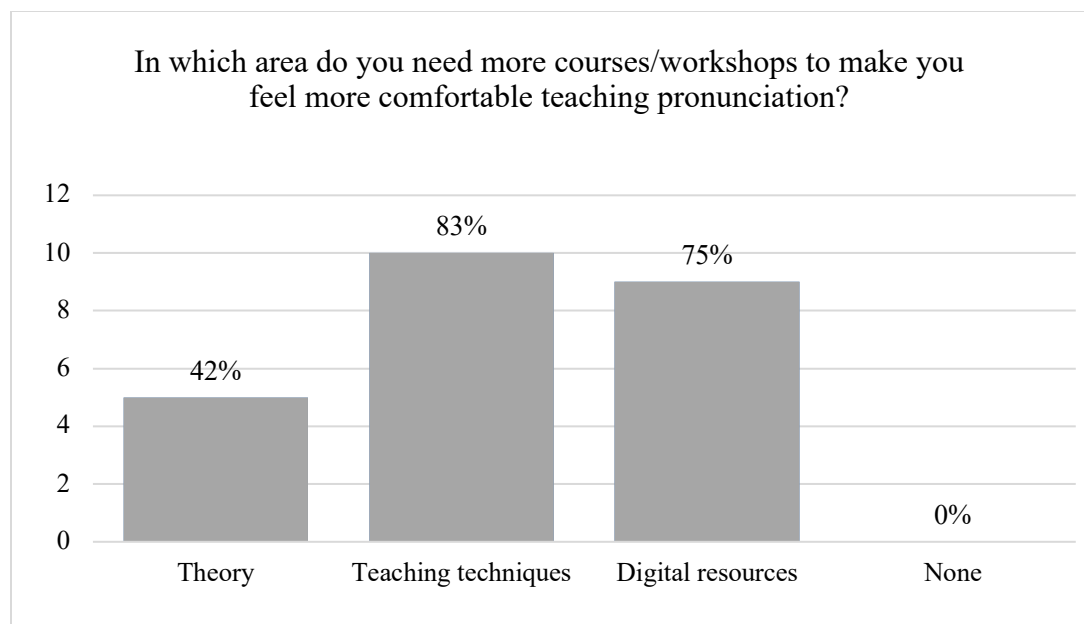


Figure 1. Participant-teachers' development needs.

Moreover, participant-teachers were asked about the use of ASR tools. Most participants answered that they feel comfortable using ASR to teach pronunciation after attending the workshop ($M = 8.6$, $SD = 1.4$). Their answers ranged from 7 to 10. In addition, most of the participant-teachers would use ASR to improve their own pronunciation ($M = 8.7$, $SD = 1.6$), and their answers ranged from 5 to 10.

ASR Accessibility

According to the data gathered from the online survey, when participants were asked about their perceptions of Google Translate's ASR feature, nearly all participants answered that it is easy to use ($M = 9.3$, $SD = 1.1$). Most reported that their students would have access to GT to practice pronunciation ($M = 9.3$, $SD = 1.5$). Participants were also asked about which teaching context would be more appropriate to use ASR to teach pronunciation (online, face-to-face, or hybrid). Figure 2 displays the percentage of responses.

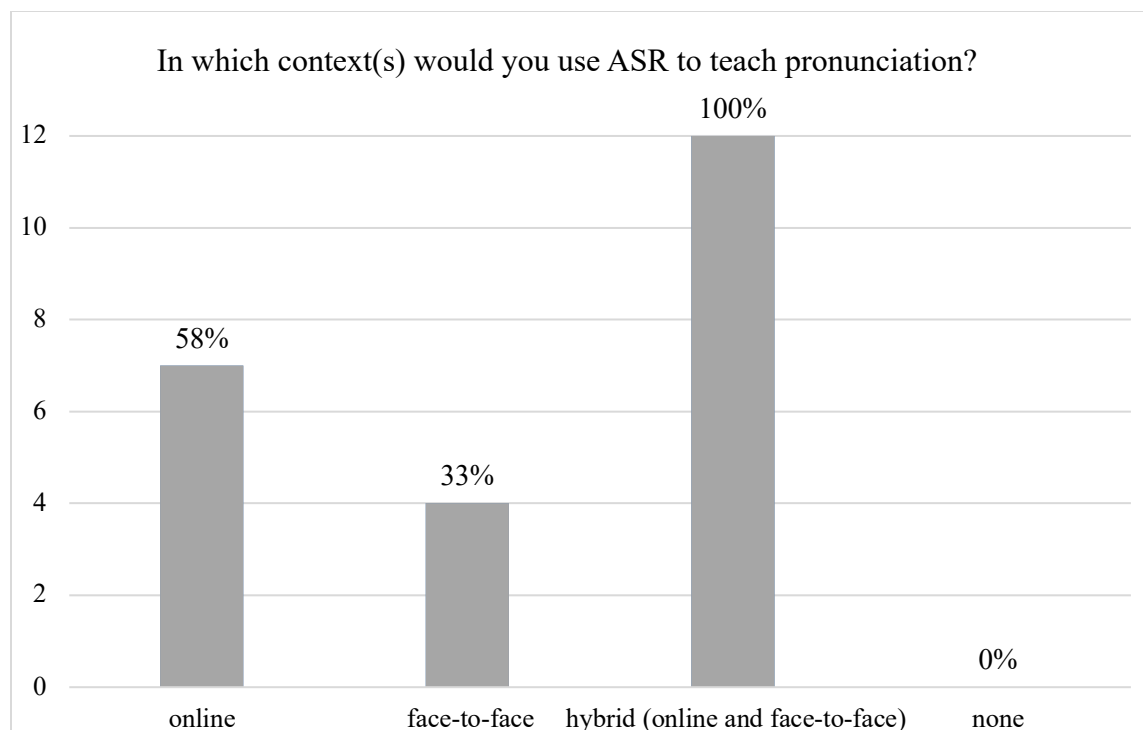


Figure 2. Participants’ preferred teaching context for ASR use.

A hybrid environment was unanimously chosen as appropriate by the participants (100%), followed by online (58%), and then, lastly, face-to-face (33%). Notice, however, that no participant marked the option “none”. Therefore, according to the participants of this study, a hybrid environment, that is, online and face-to-face classes, represents the most appropriate teaching context to use ASR for pronunciation teaching.

ASR Affordances

As an attempt to comprehend the English teachers’ perceptions of the affordances of ASR for pronunciation teaching, the online survey contained a third section composed of six questions. The means and standard deviations can be seen in Table 3.

Table 3

Participants' Perceptions of the ASR Affordances

Survey's Item	<i>M (SD)</i>
1- ASR facilitates the teaching of pronunciation.	9.0 (1.0)
2- My students would be interested in using ASR to improve their own pronunciation.	7.7 (1.9)
3- I would use ASR as a complementary tool for teaching pronunciation in my classes.	9.1 (1.2)
4- ASR can encourage/motivate learners to produce more output outside the classroom.	8.7 (1.6)
5- ASR transcription (orthographic feedback) can be beneficial to the development of learner's pronunciation.	8.8 (1.1)

Most participants agree that ASR indeed facilitates the teaching of pronunciation. However, when they were asked if their students would be interested in using ASR to improve their pronunciation, the results presented some variations ($M = 7.7, SD = 1.9$). Question 3 inquired whether participants would use ASR as a complementary tool for teaching pronunciation in their classes. The answers to this question ranged from 7 to 10 ($M = 9.1, SD = 1.2$). Most reported, thus, that ASR can be a complementary resource for pronunciation teaching, and that the participant-teachers hold a positive attitude towards using ASR for pronunciation teaching.

Participants were also asked whether they believe that ASR can encourage/motivate learners to produce more output outside the classroom. Participants' answers to this question ranged from 5 to 10 ($M = 8.7, SD = 1.6$). Only one participant gave a rating that fell below 7. The conjecture is that this rating happened due to the educational level at which this participant was currently teaching (young learners). Consequently, the participant's students might have limited autonomy in utilizing ASR independently. Furthermore, for young learners, certain digital resources may be distracting and demand teacher's mediation (Silveira et al., 2022). The overall perceptions of the participant-teachers indicate that ASR can help by encouraging learners to produce more output outside the classroom. Question 5 asked participants whether ASR's orthographic feedback can be beneficial to the development of learner's pronunciation. Their answers ranged from 6 to 10. Most participants believe that ASR transcription can be beneficial for learners' pronunciation development ($M = 8.8, SD = 1.1$).

The last question of this section elicited information on how participants would use ASR for pronunciation teaching. As displayed in Figure 3, participants would unanimously use ASR as a learning tool for self-study (100%). Many participants reported that they would also use ASR as an out-of-class supplement/homework (83%). In addition, a considerable number of participants reported that they would use ASR as an auxiliary resource for regular classes (75%) and as a pronunciation self-assessment tool (67%).

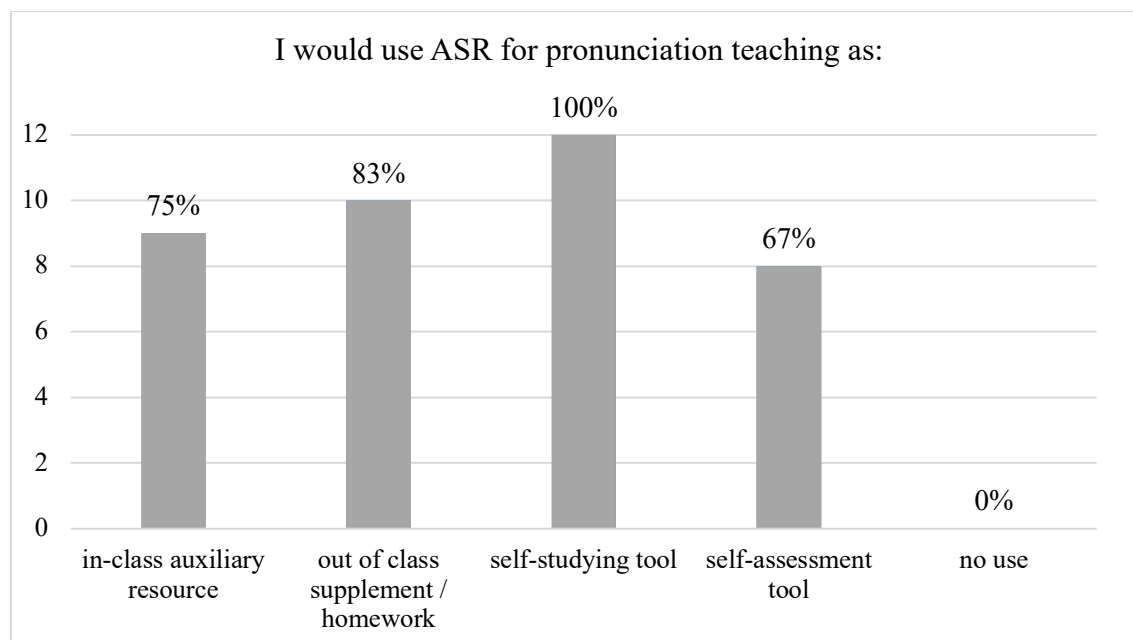


Figure 3. Participants' perceptions of ASR as a teaching resource.

Activities Appraisal

Aimed at gathering participant-teachers' appraisal of the seven ASR-based pronunciation activities, a fourth section of the online survey was designed. This section contained 13 questions addressing the six CCTA from Chapelle (2001). Table 4 shows the mean rates for each criterion.

Table 4

Summary of participants' answers to the questions addressing the CCTA from Chapelle (2001)

CALL Task Appropriateness criteria	<i>M (SD)</i>
1- Language learning potential	9.5 (.8)
2- Learner fit	9.1 (1.2)
3- Meaning focus	8.2 (2.1)
4- Authenticity	8 (1.9)
5- Positive impact	9 (2.4)
6- Practicality	8.9 (.9)

The first criterion (Language learning potential) obtained the best rates. These results indicate that the participant-teachers could clearly see the language learning potential (beneficial focus on form) of the activities. On the other hand, the fourth criterion obtained the lowest rates (Authenticity). The fourth criterion inquired about whether the activities showed a strong

correspondence between the activity and what learners would see outside the classroom. It is hypothesized that some controlled practice techniques (e.g., minimal pair and tongue-twisters) did not represent authentic language use to the participants of this research. Nevertheless, such activities supply learners with speech correction strategies that aid successful communication. Hence, it is emphasized that a clearer link between the intelligibility issues addressed by the activities and their implications for successful communication should be drawn.

All in all, the overall mean considering all the questions that address the six CCTA was high ($M = 8.8, SD = 0.7$). These results suggest that the activities suit most of the participant-teachers' realities and that the participants demonstrate an overall positive attitude towards the ASR-based pronunciation activities presented in the workshop.

After the questions addressing the six CCTA, participant-teachers were asked which one(s) of the seven ASR-based pronunciation activities presented during the workshop they would use with their students. Figure 4 illustrates the distribution of the responses to this multiple-choice question. Participant-teachers could select multiple answers.

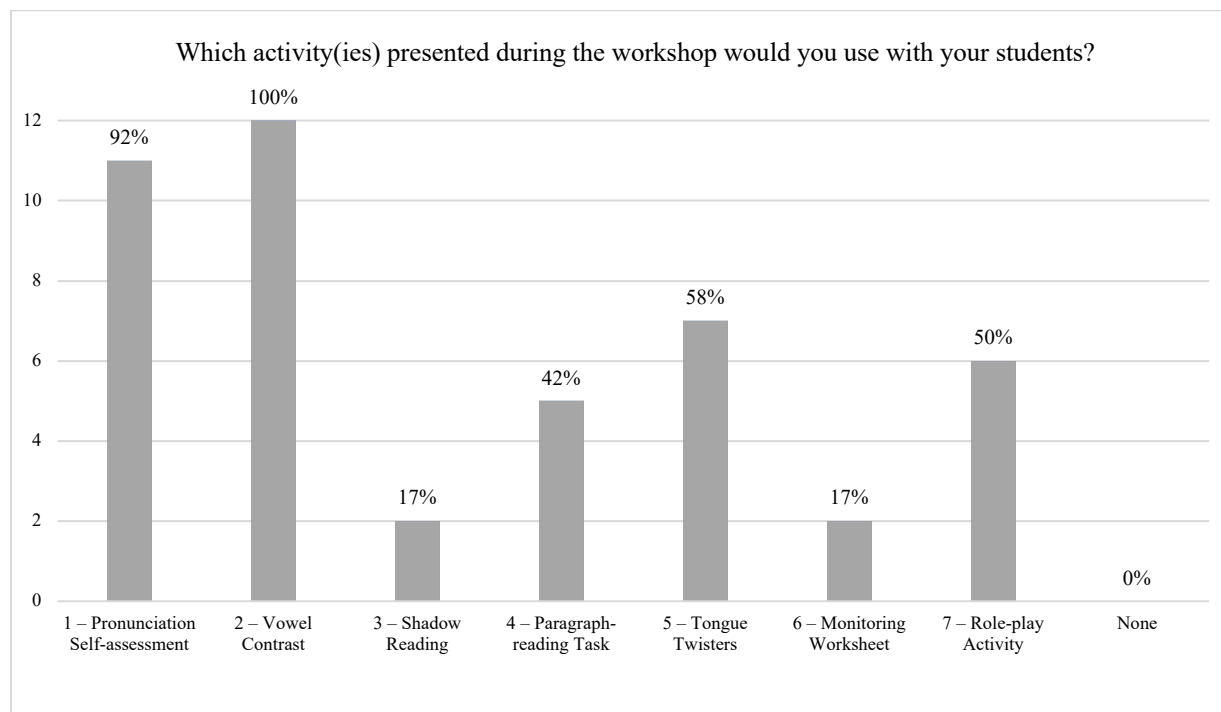


Figure 4. ASR-based pronunciation activities evaluations.

As can be seen in Figure 4, all participants marked Activity 2 (Vowel Contrast). The second most marked option (92%) was Activity 1 (Pronunciation self-assessment). Activity 5 (Tongue-twisters), 58%, and Activity 7 (Role-play Activity), 50%, were the other two activities that at least half of the participants marked as a viable option for their students. Therefore, it can be concluded that these activities may best fit the participants' teaching needs.

This study sought to explore the affordances of ASR for pronunciation teaching from the perspective of in-service English teachers and investigate teachers' appraisal of ASR-based pronunciation activities designed to be implemented in L2 English classes. Figure 5 displays a summary of the results presented in this section.

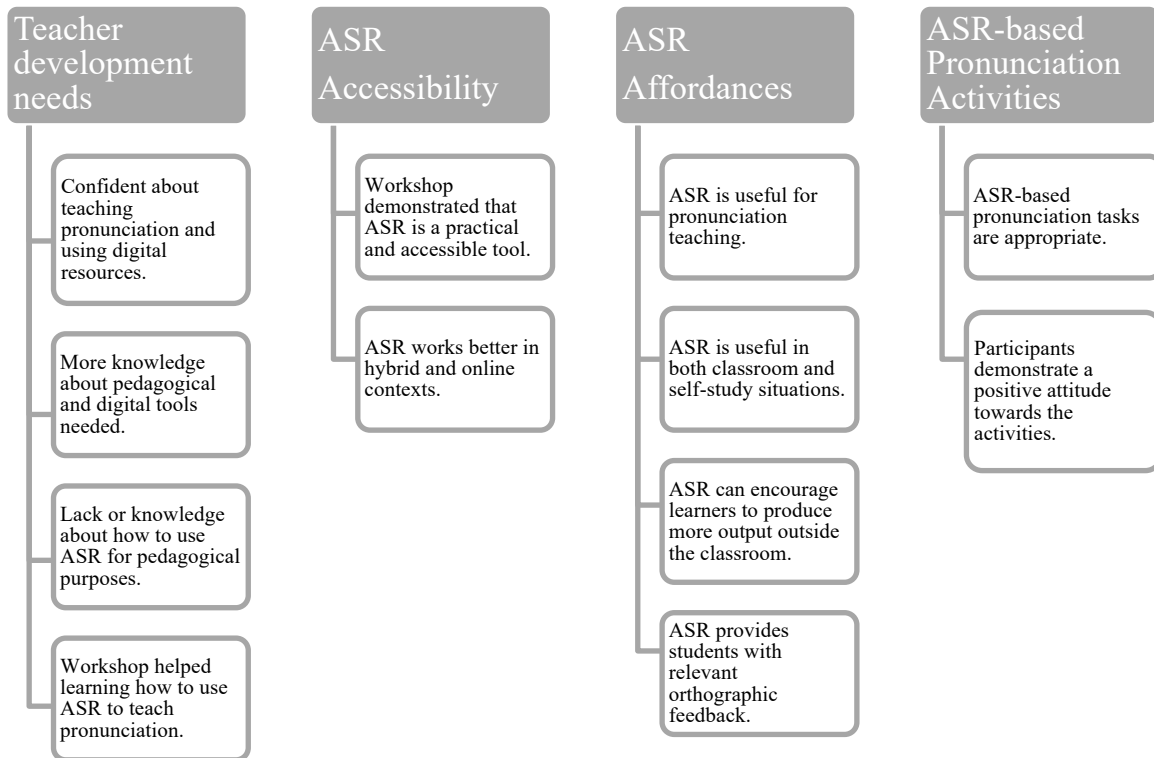


Figure 5. Teachers' perceptions of ASR for pronunciation teaching and learning.

CONCLUSIONS

Overall, participant-teachers reported positive attitudes towards the ASR-based pronunciation activities and how they were presented during the workshop session. They also perceive the ASR-based pronunciation activities as an adequate auxiliary resource for pronunciation teaching in regular classes, especially in a hybrid environment, for segmental features, for self-assessment, and for reading-aloud practice.

Moreover, the affordances of ASR for pronunciation teaching are numerous. Yet, teachers' guidance is of paramount importance for an optimal result (based on workshop comments). Although the number of participants in this research was quite limited (n=12), the results suggest that the proposed activities can provide teachers with practical ideas on how to integrate ASR technology into pronunciation teaching.

ACKNOWLEDGMENTS

The authors thank the Brazilian funding agencies CAPES and CNPq for the grants that allowed us to conduct this research and presenting it at the PSLLT2023 event.

ABOUT THE AUTHORS

William Gottardi is a doctoral candidate at the English Graduate Program at the Federal University of Santa Catarina (Brazil). He also holds a graduate scholarship granted by the National Council for Scientific and Technological Development (CAPES). His current research interests include digital technologies for pronunciation teaching and learning, CALL, and autonomous learning.

Contact Information:

Federal University of Santa Catarina - Brazil
Department of Foreign Languages and Literature
Centro de Comunicação e Expressão – CCE “B” – Room 313
Florianópolis – SC – 88.040-900
prof.williamgottardi@outlook.com

Rosane Silveira is an associate professor at the Department of Foreign Languages and Literature Federal University of Santa Catarina (Brazil). She also holds a research fellowship granted by the National Council for Scientific and Technological Development (CNPq). She conducts and supervises graduate and undergraduate research in the area of Phonetics and Phonology Applied to Second Language. Contact information: rosanesilduarte@gmail.com.

Contact Information:

Federal University of Santa Catarina - Brazil
Department of Foreign Languages and Literature
Centro de Comunicação e Expressão – CCE “B” – Room 313
Florianópolis – SC – 88.040-900
rosanesilduarte@gmail.com

REFERENCES

- Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly*, 48(1), 136–163. <https://doi.org/10.1002/tesq.99>
- Buss, L. (2013). Pronunciation from the perspective of pre-service EFL teachers: An analysis of internship reports. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th PSLLT* (pp. 255-264). Ames, IA: Iowa State University.
- Buss, L. (2016). Beliefs and practices of Brazilian EFL teachers regarding pronunciation. *Language Teaching Research*, 20(5), 619–637. <https://doi.org/10.1177/1362168815574145>
- Cardoso, W. (2022). Technology for speaking development. In T. Derwing, M. Munro, & R. Thomson (Eds.), *The Routledge handbook of second language acquisition and speaking* (pp. 299–313). Routledge.
- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation hardback with audio CDs (2): A course book and reference guide* (2nd ed.). Cambridge University Press.

- Chapelle, C. A. (2001). *Computer applications in second language acquisition*. Cambridge University Press.
- Costa, B. C. de A. (2016). *Pronunciation teaching is not a one-size-fits-all endeavor: EFL teachers' beliefs and classroom practices* (Publication No. 341497) [Master's thesis, Federal University of Santa Catarina]
<https://repositorio.ufsc.br/handle/123456789/168110>
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27(1), 70–105.
<https://doi.org/10.1080/09588221.2012.700315>
- Gonçalves, A. R., & Silveira, R. (2015). Intelligibility research in Brazil: Empirical findings and methodological issues. *Revista Horizontes de Linguística Aplicada*, 14(1), 51–81.
<https://doi.org/10.26512/rhla.v14i1.1384>
- Gottardi, W., Almeida, J. F. de, & 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/10.35699/1983-3652.2022.36736>
- Jamieson, J., & Chapelle, C. A. (2010). Evaluating CALL use across multiple contexts. *System (Linköping)*, 38(3), 357–369.
- Junior, R. C. G., Silva, L. de O., & Paiva, V. L. M. de O. e. (2022). Tecnologias digitais para aprender e ensinar inglês no Brasil. *Texto Livre*, 15, e38008.
<https://doi.org/10.35699/1983-3652.2022.38008>
- Liakin, D., Cardoso, W., & Liakina, N. (2017). Mobilizing instruction in a second-language context: Learners' perceptions of two speech technologies. *Languages*, 2(3), 11.
<https://doi.org/10.3390/languages2030011>
- McCrooklin, S. (2019). Dictation programs for second language pronunciation learning: Perceptions of the transcript, strategy use and improvement. *Konin Language Studies*, 7(2), 137–157. <https://doi.org/10.30438/ksj.2019.7.2.2>
- McCrooklin, S. M. (2016). Pronunciation learner autonomy: The potential of Automatic Speech Recognition. *System*, 57, 25–42. <https://doi.org/10.1016/j.system.2015.12.013>
- Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals*, 51(3), 617–637.
<https://doi.org/10.1111/flan.12348>
- Pennington, M. C., & Rogerson-Revell, P. (2019). *English pronunciation teaching and research: Contemporary Perspectives*. Palgrave Macmillan UK.
- Revell-Rogerson, P. M. (2021). Computer-Assisted Pronunciation Training (CAPT): Current issues and future directions. *RELC Journal*, 52(1), 189–205.
<https://doi.org/10.1177/0033688220977406>
- Silveira, R., Delatorre, F., Reis, L. P., & Goncalves, A. R. (2017). Percepção, produção e inteligibilidade do inglês falado por usuários brasileiros. *Perspectivas Atuais de Aprendizagem e Ensino de Línguas*. Florianópolis: LLE/CCE/UFSC, 237–283.
- Silveira, R., Zanchet, C. E., & Pereira, M. H. (2022). Affordances of digital technology for English pronunciation teaching: The perspective of Brazilian teachers. *Veredas - Revista de Estudos Linguísticos*, 26(2). <https://doi.org/10.34019/1982-2243.2022.v26.38449>
- Souza, H. K. de, & Gottardi, W. (2022). How well can ASR technology understand foreign-accented speech? *Trabalhos Em Linguística Aplicada*, 61(3), 764–781.
<https://doi.org/10.1590/010318138668782v61n32022>

- Yu, D., & Deng, L. (2015). *Automatic speech recognition: A deep learning approach*. Springer.
<https://doi.org/10.1007978-1-4471-5779-3>
- Zimmer, M., Silveira, R., & Alves, U. K. (2009). *Pronunciation instruction for Brazilians: Student's book*. Cambridge Scholars Publishing.