AUTOMATIC PRONUNCIATION ASSESSMENT IN A FLIPPED CLASSROOM CONTEXT

Hana Vancova
Faculty of Education, Department of English Language and Literature, University of Trnava, Trnava, Slovakia,
hana.vancova@truni.sk

Abstract

Automatic pronunciation assessment programs allow learners to practice pronunciation independently and receive immediate feedback based on acoustic analysis of their speech compared to a pronunciation model. The programs display a high level of reliability when assessing speakers' pronunciation accuracy. The paper aims to present the results of an investigation into a recent experimental application of an automatic pronunciation assessment online program used within the flipped classroom of English phonetics and phonology course taught to foreign language pre-service teachers of English. Flipped learning approach was selected to offload the preliminary activities and tasks that do not require the immediate attention of the teacher to an online learning management system. Thus, course participants' face-to-face learning could be enhanced and more effective. The participants could practice and receive tailored feedback from an online program before the face-to-face class. After the individual pronunciation practice, they could discuss more complex questions with the lecturer and, in

collaboration with fellow students. The paper will present the survey results among the course participants who evaluated their experience with automatic pronunciation evaluation. Special attention is paid to the potential use of such programs for developing linguistic competence.

Keywords

1. Introduction

Pronunciation is integral to linguistic competence, and spoken communication largely depends on this aspect of language. Its importance is globally recognized by EFL language learners (Huensch & Thompson, 2017; Vančová, 2020). On the other hand, teachers give it a relatively low degree of importance (Levis & Sonsaat, 2017; Vančová, 2020), even if good pronunciation positively impacts overall linguistic competence and the confidence of foreign learners (Aiello & Mongibello, 2019). Consequently, pronunciation training has two main objectives – achieving accuracy and/or comprehensibility. While comprehensibility (i.e. the general quality of foreign learner's pronunciation to be understood) is relatively subjective due to a range of factors influencing it on the part of the speaker as well as the listener (Saito, 2021), accuracy (i.e. the ability of the foreign learner to pronounce identical sounds to native speakers) is relatively easy to measure by digital tools. Learners train accent-free pronunciation to include themselves in their new language community (Derwing, 2019). On the other hand, they tend to keep the influence of their mother tongue in their foreign-language pronunciation to maintain their identity (Szyszka, 2022).

2. Literature Review

An inherent quality of digital tools for language learning and pronunciation is their ability to develop learner autonomy (McCrocklin, 2016). This concept has been promoted in language learning because it allows learners to train pronunciation independently, outside the formal setting of the classroom, to achieve the goals learners see as desirable for them. Providing learners with individualized feedback when learning pronunciation is a very effective way to improve their performance (Moxon, 2021; Ngo, 2023), but it requires the capacity many teachers do not have in a regular classroom. In addition, the pronunciation practice takes place in a "safe"

environment, where the learner is not exposed to the evaluation of their peers, and the feedback is individualized to their particular needs (Dai & Wu, 2021; Evers & Chen, 2022).

Automatic speech recognition technology has been one of the critical components of pronunciation assessment (Cheng et al., 2020) since the 1990s. According to the authors, pronunciation assessment takes place in three steps: segmentation of speech into smaller units – pronunciation analysis of the speech units – calculation of the score (comp. Sidgi & Shaari, 2017; Cámara-Arenas et al., 2023). The ASR systems are usually designed around a pronunciation model chosen by the program designers and using artificial intelligence (Pokrivčáková, 2019), which creates a basis for pronunciation comparison and assessment. The systems contain samples of individual words narrated in laboratory conditions, which deviates from naturally pronounced speech (Bogach et al., 2021). Thus, the quality of ASR-based tools must be inspected for possible systems limitations.

The accuracy of such systems has proven to exceed 90% (McCrocklin et al., 2019) and improve learners' motivation. The compared programs vary in their quality and ability to recognize the speech of native and non-native speakers using the program. In addition to pronunciation assessment tools based on ASR designed particularly for educational purposes, foreign language learners can benefit from authentic communication with ASR technology designed for non-educational, authentic purposes ("off the shelf", Tejedor-García et al., 2021, p. 2; Gottardi et al., 2022; John et al., 2022; Khademi & Cardoso, 2022; Papin & Cardoso, 2022). Therefore, besides the traditional drilling sequence in pronunciation training (listen and repeat), ASR also allows practicing open-ended tasks, e.g. asking users questions with predictable or expected answers (McCrocklin et al., 2019).

Based on the design of the program, the ASR technology can provide learners with explicit feedback (overall percentage of accuracy, analysis of individual segmental deviations from the model in the technology, comment) or implicit feedback (cooperation of the software after vocal input of the learner, rewritten text, etc.). Explicit feedback by ASR technology appears to be more beneficial for younger learners, while adults appreciate implicit feedback (Wang & Young, 2015).

Sigdi and Shaari (2017) confirmed a varying degree of pronunciation improvement in particular target sounds ranging from significantly significant improvement (mostly labial consonants) to less improvement (velar nasal) in Arabic users of ASR-based pronunciation training

software. Guskaroska (2020) found that learners enjoy using ASR for pronunciation training. The technology is sufficiently reliable to human raters, even if individual sounds may be interpreted differently. In dictation-based pronunciation training, participants used ASR technology to check correct pronunciation, along using other training strategies. It allowed them to see pronunciation mistakes more holistically across larger text units (McCrocklin et al., 2019). At the same time, participants would appreciate getting additional feedback on the dictation task.

Tejedor-García et al. (2021) proved significant improvement after using the ASR technology for Japanese learners of Spanish in a specifically designed tool. Wang and Young (2015) found that adult learners adapt to the learning environment better than younger learners, reflecting their intellectual development and learning strategies. In addition, younger learners preferred more explicit feedback to adult learners, who benefited from recasts. Inceoglu et al. (2020) found that ASR technology is limited to overcoming pronunciation problems resulting from adverse inferences from the learner's mother tongue. In addition, Artieda and Clements (2019) identified an overall positive attitude of learners towards ASR technology. However, the use of the system varies across different nationalities of learners.

Flipped learning as a concept is based on a reversed order of traditional learning sequence – theoretical introduction to a problem in class with a teacher and completing homework individually at home (Turan & Akdag-Cimen, 2020; Veres & Muntean, 2021). This standard sequence, however, does not consider that listening to explanations by learners in a video recording is relatively easy compared to homework which typically consists of more complex tasks. However, the flipped approach can also be applied without video lectures. Therefore, it might be more beneficial for learners to familiarize themselves with the theory individually at home and during the limited time in class with the teacher and other classmates to find answers to potential questions and complete more complex tasks in cooperation. This approach has been performed for a relatively long time but was formally elaborated by Bergman and Samms (2014) due to the high absence rate of learners and the subsequent inefficient use of time by revising the subject matter for previously absent students. Currently, flipped learning is combined with ASR for pronunciation training (Jiang et al., 2023; Sze & Nasri, 2022) in the context of other disciplines (e.g., Khanal, 2020; Linlin, 2021). Using the flipped approach ´, particularly within the course of phonetics and phonology, was studied by Sharipova and Kakhkhorova (2022), as well as Setter (2015) and Wahib and Tamer (2021).

3. Methodology

The following section of the paper will present the main objectives and research questions, the study's conditions, results and interpretations of the collected data.

3.1. Research Objectives

Kholis (2021) studied the use of ASR and confirmed that ASR could improve learners' pronunciation in English phonetics and phonology. However, the study does not report a flipped design. Therefore, the possible gap open for investigation has been identified. The presented study aims to investigate the perceived effectiveness and motivation of EFL learners' experience with using ASR technology within a flipped course of English phonetics and phonology using a questionnaire survey presented to the course participants after completion of the course. A qualitative research design was selected as the most suitable to conduct the study. As a result, the following research questions were formulated:

1. How effectively do the students find ASR to improve pronunciation and increase motivation?
2. Which skills of learners were improved by the ASR technology in the flipped course?
3. What are the learners' recommendations for using ASR in the flipped phonetics and phonology course?

3.2. Context of the Study

The study was carried out within a 12-week course of English phonetics and phonology taught to 130 pre-service English teachers and students of English philology in the first semester of their studies. The current course module comprises one 45-minute face-to-face lecture and one 90-minute seminar. The seminar is split into one face-to-face 45-minute seminar, and one 45-minute session is substituted by individual work in LMS Moodle with ASR-technology-based activities. All parts of the course are compulsory for all students. Other LMS Moodle tasks focus on practical pronunciation tasks and theoretical introduction to the course in the flipped design. The flipped design was chosen for the benefits related to class time management. An ASR technology for automatic pronunciation evaluation was included as a pilot project within the course, as the growing number of students in the course prevents participants from providing tailored and individualized feedback by the teacher.

As for the ASR tool, a free online program offering users practicing segmental and suprasegmental features was selected. The free version was used as part of a pilot study before committing to implementing the program's full version into the course. The aim of the paper is not to evaluate the quality and properties of the program itself but the users' experience with ASR technology in general. Therefore, the program will not be explicitly named in the paper.

This survey used two tools for measuring the participants' experience – a questionnaire containing Likert-scale, multiple-choice, and open items. The completion of the questionnaire for students was voluntary. The questionnaire was presented to course participants on survio.sk. The second data collection tool was an open-written item, where program users were asked to provide their opinions on their experience using the program. The data collection period was limited to the first week of the examination period, which could have contributed to a low questionnaire turnout (N=32). Due to the relatively low turnout of the questionnaire, the qualitative analysis design was selected.

3.3. Participants

Thirty-two students, in total, provided answers to a complete questionnaire. At the beginning of the questionnaire, they were asked six questions relevant to their background. Most were first-year students (N=28), and the remaining students repeated the class. None of the participants identified themselves as a native speaker of English.

Students assessed the level of their computer skills as intermediate (N=26), expert (N=4) or beginner (N=2). Thirteen participants have already used digital tools for pronunciation training. However, none of them provided a specific answer.

4. Results

After collecting the questionnaire data, they were analyzed and will be presented in the following sections.

The first questionnaire item related to questionnaire objectives is a list of 10 statements where the participants expressed their attitude to them on a 5-point Likert scale (1) Strongly agree; (2) Agree; (3) Neither agree nor disagree; (4) Disagree; (5) Strongly disagree.

Table 1.1 *Likert-scale item related to the use of ASR in the course of English phonetics and phonology*

Questionnaire item	Mean	Median	Mode
1. The automatic pronunciation feedback helped me learn how to pronounce correctly.	2.3	2	2
2. The exercises were appropriately difficult.	2.65	3	2, 3
3. The feedback was clear.	2.23	2	2
4. Using the program was easy.	1.56	1	1
5. Using the program encouraged me to participate in class.	3.06	3	3
6. I valued that the program provided the feedback anonymously	2.09	2	1
7. I would prefer more feedback from the teacher.	2.18	2	1, 2
8. The use of the program helped make some topics clearer.	3	3	3
9. Using the program motivated me.	2.84	3	2, 3
10. I think the program made the class better.	2.87	3	3

(Source: Author's Own Illustration)

The overall average scores for individual items suggest a partial agreement to the neutral position of participants to ASR technology in the flipped classroom of English phonetics and phonology with a pronunciation training section. The most positive was evaluated item 4, related to the relative ease of use of the program. The item with the lowest score was item 5, concerning developing theoretical knowledge on pronunciation features of English.

The items were divided into several categories:

a) Questions regarding the feedback provided (items 3, 6, 7)

The items related to the program feedback provide a comparable average score for all three items (ranging from 2.09 to 2.23). The results indicate that the clarity and maintenance of anonymity while receiving feedback were viewed partially positively by the participants of the course. However, a similar number of students prefer receiving feedback from the teacher, which establishes a more personal relationship between students and the lecturer, which is commonly recognized as an essential factor in implementing digital tools in language learning.

b) Questions regarding the class content, structure and impact on class performance (items 1, 8)

These two questionnaire items concentrated on two main facets of the course – presenting a theoretical introduction to pronunciation and providing space to practice the theory on a particular set of words using the ASR technology. The participants were more likely to evaluate the improvement in their pronunciation more positively (average 2.3). They could not see the impact of practical tasks on their theoretical understanding of pronunciation issues (average 3).

c) Questions regarding the motivational character of the classes (items 5, 9, 10)

The items related to the impact of ASR technology on classroom activity and the intrinsic motivation of learners indicate that technology did contribute to them in a significant way. On the contrary, the items scoring averages 3.0, 2.84 and 2.87, respectively, indicate learners' rather reserved attitude in this respect.

d) Questions regarding the ASR-based program itself (items 2, 4)

The score regarding the actual system use was the highest regarded feature of the program as seen by the study participants (average 1.56). The participants later confirmed that spontaneously written questionnaire items were presented later in the paper. On the contrary, the participants could not estimate the appropriateness level of exercises with the same confidence (average 2.65).

The following questionnaire item dealt with the issue of the future use of ASR technology in pronunciation classrooms. Out of all participants, 17 responded "yes", 11 responded "no", and four responded "other", but any of the respondents provided no explanation or suggestion.

The following item, item nine of the questionnaire, investigated the different aspects of the course that students identified as positive.
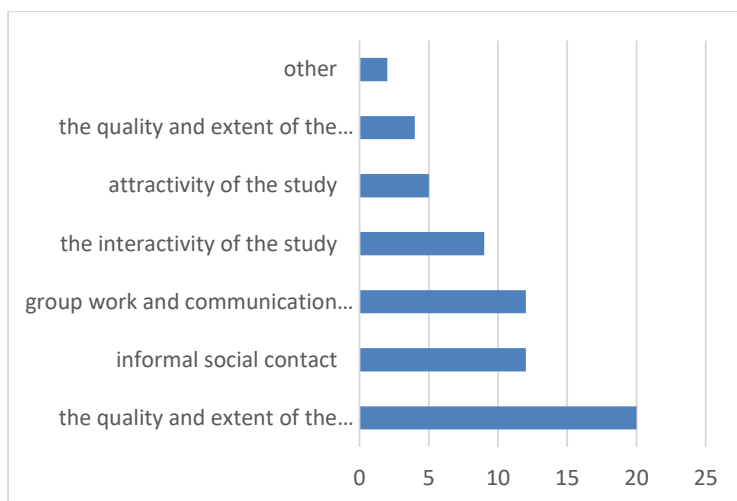
Figure 1: *The aspects of the course of phonetics and phonology that the learners appreciated*
(Source: Author's Own Illustration)

The collected data indicate that the learners see the feedback provided within the course as the sixth feature they appreciated – their focus was primarily on the actual course content, followed by various forms of social interaction (informal, in-class, general interaction). In this case, learners could choose more than one option.

Finally, participants were asked to identify the areas of their improvement in the course. This item also allowed participants to choose more than one answer. The results indicate that the ASR-based program helped learners improve their pronunciation, which was the most frequently selected item, and improve their computer skills (the third most frequent answer).
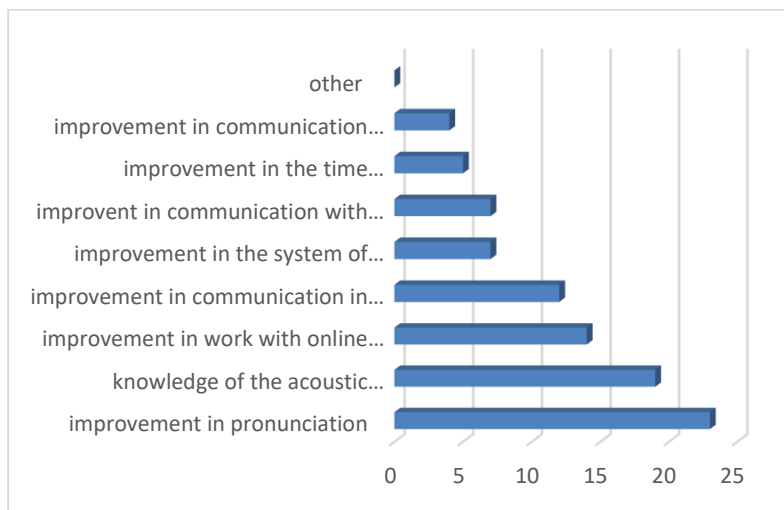

Figure 2: *Areas of Improvement Identified by the Participants*
(Source: Author's Own Illustration)

The second part of the data collection is based on the open answers of participants to the ASR-based technology used within a flipped classroom.

The attractivity of the digital tool interface and the perceived ease of use are among the main criteria for the success of e-learning. In terms of the interface, the program appeared to be friendly to the users *("I did not find using the program difficult; actually, it was quite easy"; "very easy to use, even for a beginner").*

The effectiveness of the program appeared to be proven by the results the learners could receive *("I found these exercises way more useful and also I had fun doing them", "Doing these exercises improved my pronunciation, mostly because I was able to check original pronunciation from just clicking on one button [...] I was able to see how % my pronunciation is correct."; "When I opened the link, which our teacher gave us, for the first time, I thought like 'Oh, what is this? It seems really easy!' I tried some words and found out it is quite difficult to pronounce certain words the right way.")*

The evaluation of the program's effectiveness as perceived by the users depends on the ability of users to self-evaluate their own skills realistically. The literature proves the accommodation tendency in communication in the pronunciation sphere *("The thing that was happening to me all the time was that even though I said a word/sentence correctly, it still gave me 95 per cent, and when I said the same word/sentence again, it gave me 99% whereas I didn't change anything in my pronouncing.").*

However, one of the limitations of digital tools for pronunciation training is their relative difficulty in distinguishing a model added to the system narrated by a professional speaker in laboratory settings from authentic, natural, fluent speech. In some cases, the participants could observe this shortcoming of the program *("I don't think it helped much, to be honest. The best way to get a high score on the words and sentences is to say them in a very monotone robotic voice. In everyday speech, we don't really speak that way").*

Each digital tool must be implemented for a specific purpose. In pronunciation training, instructors typically provide exercises addressing specific learners' issues. Exercises provided to learners focused on specific segmental as well as suprasegmental features. The learners identified the following segmental mistakes they were able to notice in their production – final consonant sounds *("I struggled the most with pronouncing the ending letters. And this program helped me with that a lot")* or rhoticity of English *("I think it was effective. It certainly did help me improve*

with my pronunciation of "R" sound in particular and some more.") or generally incorrect acquired word pronunciation *("I didn't know about so many words which I was pronouncing incorrectly and thanks to this website I learnt it right. I didn't realize there are so many ways how you can pronounce the word before.").* Regarding suprasegmental features, learners could correct their misplacement of word stress *("I found out a lot of words that I was pronouncing wrong or that I was emphasizing wrong syllables").*

In addition, the program allowed learners to systematize the pronunciation of words across various accents of English *("The program helped me with my pronunciation a lot, I used to pronounce almost every word in a sentence with a different accent").*

When students asked about the possibility of using ASR technology in the future, those participants who were in favor of the program would use it in the future if they observed reoccurring problems by self-monitoring *("I would use it in the future if I would feel that I am struggling with some types of nouns or verbs.")* or if they taught in the future *("I would definitely use it in the future, perhaps while teaching my own class someday. Everyone should try it, there is nothing to lose", "I have already planned to include this program into my future learning sessions."),* or have already introduced it in their teaching practice or for entertainment among their friends *("I would like to. Some tests I shared with my friends to see how good they are.").* In addition to practical use, learners would use it to make lessons more enjoyable *("It is also quite fun. I recommend this program and others like this because it is not a boring way how to learn something and improve your skills."),* and acknowledge the irreplaceable place of technology in education *("In my opinion, I see using digital tools as a future in learning. We are living in a fast and technological world, and we can't omit technology in studying.").*

As reported by the participants, the ASR technology appears to provide an instant applicable input for learners *("I was surprised how many words I was not able to pronounce correctly, but I followed the example and got better immediately. The effectiveness of this program is astounding because, in my head, I have always realized the correct pronunciation.").*

However, participants who self-identified themselves as more proficient learners of English expressed a rather reserved attitude to the implemented tool *("Maybe. I don't think it's all that useful for me personally.", "I would recommend it to people who really struggle with pronunciation or beginners.").*

The participants were also asked to suggest improvements for implementing ASR in the context of the English phonetics and phonology course. Students appreciated systematicity, regularity, predictability of lessons, using authentic tools *("I'm pretty happy with its current use in this course, but there is always room for improvement.", "Actually, I do not have anything that I would improve. I think once a week it was effective.").* Other students would suggest daily lessons *("Maybe make it like an everyday task if it is possible. But only one lesson (15 words). I think if we can do this every day, it can be really helpful to stay in that mindset of focusing on our pronunciation.")*

Several students provided a more complex suggestion:

"I would say that the lectures could be divided into topics (not just randomly chosen sentences or words) […] If I look directly into its use in the lessons, I would maybe recommend it to practice also with different speech recognition tools like Siri or some online shopping site or so."

This student focused on the fact that ASR technology is also available for commercial use in various systems, and pronunciation can be practiced in authentic real-life contexts. In addition, the student would similarly practically focus on thematically organized vocabulary instead of the focused practice of target sounds.

One of the typical features of using ASR technology is the focus on pronunciation accuracy based on comparing the model pronounced in ideal conditions with the actual pronunciation of the speaker or learner. However, accuracy is only one of the ultimate goals of current pronunciation training, as comprehensibility is viewed as a more effective training goal in real-life communication. Students who recognize it, therefore, would not use practice pronunciation nor ASR technology *("In the future, unless I was going to work as a teacher or in some profession where I need to have pure British English, I probably would not use it.").*

One of the participants pointed out the fact that learners can practice independently at home, saving valuable classroom time for more complex tasks *("It was a good idea to use it because in the class is a lot of students and it saves time, that they can do it alone at home.").*

One of the participants needed to retake the class because he failed the previous academic year, therefore could compare the class in the previous academic year, drilling the target sounds without the automatic feedback *("I would say even liked it more than last year when I had to do recordings of my own voice.").*

However, one of the participants confirmed one previous finding carried out in the course of phonetics and phonology during the previous academic year (Vančová, 2021). The participants in the study identified that computer-based tasks could be overwhelming to learners if they are overused. They benefit from traditional pen-and-paper tasks, too, especially in this period of learning in the online domain.

"Neither ASR nor flipped learning was uninteresting. On the contrary, both of the methods were intriguing. However, due to the amount of homework, tasks and assignments also from other academic subjects, they quickly became overwhelming. Therefore, decreasing the regularity of these tasks would certainly make them more enjoyable."

To sum up participants' comments, they spontaneously pointed out commonly recognized features of ASR technology in pronunciation training without any prompt. They pointed out the positive and negative features of ASR technology, and their answers allowed the formulation of the following recommendations.

5. Discussion, Conclusion and Future Directions

The presented study investigated the use of ASR technology in a flipped classroom of English phonetics and phonology courses. In particular, it investigated its efficiency and motivational nature as perceived by the study participants.

The study results confirmed the findings of McCrocklin et al. (2019) regarding the increased motivation of study participants as well as Tejedor-García et al. (2021) and Kholis' (2021) conclusions that ASR improves learners' pronunciation, therefore can be perceived as an effective tool for pronunciation practice (research question 1).

The results also revealed that the most beneficial areas of learners' improvement were pronunciation (comp. Moxon, 2021; Ngo, 2023) as well as the theoretical awareness of the issues of phonetics and phonology Setter, 2015; Wahib & Tamer, 2021; Sharipova & Kakhkhorova (2022). Learners also developed greater autonomy and study skills while participating in the course (comp. McCrocklin, 2016), as well as communication and computing skills (research question 2).

Finally, the participating students expressed a positive attitude towards regularity and predictability in the course structure and would recommend greater systematicity. The systematicity in learners' view should cover the frequency of the exercise and its content. A more concrete, topic-based approach to choosing lexically related vocabulary items seems more

acceptable for specific learners than focusing on abstract target sounds of English. In addition, although several participants expressed their satisfaction with the feedback provided by the system, some students wished to receive the teacher's feedback as well. When evaluating two comparable pronunciations, a human rater could pinpoint the spheres where the technology reaches its limits for learners who believe the system is unstable or unreliable (research question 3).

As suggested by the research, the learners need to see the benefits of using a digital tool; the innovative nature of a program should not be the only advantage of a class. In the presented study, learners with self-reported higher foreign language proficiency (due to the anonymous nature of the questionnaire) addressed the program's limitations in further developing their proficiency.

ASR used in the context of a flipped phonetics and phonology course has proven to be a viable tool for improving the general competencies of learners; therefore, in the future, the research should focus on a more quantitative analysis of the correlation between the two variables within the course. Other areas of future research should include, for instance, the use of ASR to evaluate the acquired theoretical knowledge or implement the flipped course design into other courses.

5.1. Study Limitations

Due to the relatively low number of participants, the results presented in the study have a limited character to the participating group of students. However, they can confirm the existing assumptions of implementation of ASR technologies in language learning, especially in flipped classroom contexts.

ACKNOWLEDGEMENT

REFERENCES

Aiello, J. & Mongibello, A. (2019). Supporting EFL learners with a virtual environment: A focus on L2 pronunciation. *Journal of E-Learning and Knowledge Society*, 15(1), 95-108. https://www.learntechlib.org/p/207521/

Artieda, G. & Clements, B. (2019). A comparison of learner characteristics, beliefs, and usage of ASR-CALL systems. *CALL and complexity*, 19, 19-25. https://doi.org/10.14705/rpnet.2019.38.980

Bogach, N., Boitsova, E., Chernonog, S., Lamtev, A., Lesnichaya, M., Lezhenin, I., Novopashenny, A., Svechnikov, R., Tsikach, D., Vasiliev, K., Pyshkin, E. & Blake, J. (2021). Speech processing for language learning: A practical approach to computer-assisted pronunciation teaching. *Electronics*, 10(3), 1-22. https://doi.org/10.3390/electronics10030235

Cámara-Arenas, E., Tejedor-García, C., Tomas-Vázquez, C. J. & Escudero-Mancebo, D. (2023). Automatic pronunciation assessment vs. automatic speech recognition: A study of conflicting conditions for L2-English. *Language Learning & Technology*, 27(1), 1-19. https://hdl.handle.net/10125/73512

Cheng, S., Liu, Z., Li, L., Tang, Z., Wang, D. & Zheng, T. F. (2020). ASR-free pronunciation assessment. https://doi.org/10.21437/Interspeech.2020-2623

Dai, Y. & Wu, Z. (2021). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. *Computer Assisted Language Learning*, 1-24. https://doi.org/10.1080/09588221.2021.1952272

Derwing, T. M. (2019). Utopian goals for pronunciation research revisited. *Pronunciation in Second Language Learning and Teaching Proceedings*, 10(1), 27-35. https://www.iastatedigitalpress.com/psllt/article/15363/galley/13562/view/

Evers, K. & Chen, S. (2022). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning,* 35(8), 1869-1889. https://doi.org/10.1080/09588221.2020.1839504

Gottardi, W., Almeida, J. F. D. & Tumolo, C. H. S. (2022). Automatic speech recognition and text-to-speech technologies for L2 pronunciation improvement: reflections on their affordances. *Texto livre*, 15, e36736. https://doi.org/10.35699/1983-3652.2022.36736

Guskaroska, A. (2020). ASR-dictation on smartphones for vowel pronunciation practice. *Journal of Contemporary Philology*, 3(2), 45-61. https://doi.org/10.37834/JCP2020045g

Huensch, A. & Thompson, A. S. (2017). Contextualizing attitudes toward pronunciation: Foreign language learners in the United States. *Foreign language annals*, 50(2), 410-432. https://doi.org/10.1111/flan.12259

Inceoglu, S., Lim, H. & Chen, W. H. (2020). ASR for EFL Pronunciation Practice: Segmental Development and Learners' Beliefs. *Journal of Asia TEFL*, 17(3), 824. https://doi.org/10.18823/asiatefl.2020.17.3.5.824

Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S. & Wu, N. (2023). Exploring the effects of automatic speech recognition technology on oral accuracy and fluency in a flipped classroom. *Journal of Computer Assisted Learning*, 39(1), 125-140. https://doi.org/10.1111/jcal.12732

John, P., Cardoso, W. & Johnson, C. (2022). Evaluating automatic speech recognition for L2 pronunciation feedback: a focus on Google Translate. *Intelligent CALL, granular systems and learner data: short papers from EUROCALL 2022*, 197-202. https://doi.org/10.14705/rpnet.2022.61.1458

Khademi, H. & Cardoso, W. (2022). Learning L2 pronunciation with Google Translate. *Intelligent CALL, granular systems and learner data: short papers from EUROCALL 2022*, 228-233. https://doi.org/10.14705/rpnet.2022.61.1463

Khanal, R. (2020). An investigation of the effectiveness of flipped classroom teaching in project management course: A case study of Australian higher education. *PUPIL: International Journal of Teaching, Education and Learning*, 4(2), 348-368. https://doi.org/10.20319/pijtel.2020.42.348368

Kholis, A. (2021). Elsa speak app: automatic speech recognition (ASR) for supplementing English pronunciation skills. *Pedagogy: Journal of English Language Teaching*, 9(1), 01-14. https://doi.org/10.32332/joelt.v9i1.2723

Levis, J., & Sonsaat, S. (2017). Pronunciation teaching in the early CLT era. *The Routledge handbook of contemporary English pronunciation,* 267-283. https://doi.org/10.4324/9781315145006-17

Linlin, S. (2021). Blended Teaching of Foreign Language in China from The Perspective of Ecology. *PUPIL: International Journal of Teaching, Education and Learning*, 5(2), 44-62. https://doi.org/10.20319/pijtel.2021.52.4462

McCrocklin, 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

McCrocklin, S., Humaidan, A. & e Edalatishams, I. (2019). ASR dictation program accuracy: Have current programs improved? *Pronunciation in Second Language Learning and*

Teaching Proceedings, 10(1), 191-200.
https://www.iastatedigitalpress.com/psllt/article/15376/galley/13549/view/

Moxon, S. (2021). Exploring the Effects of Automated Pronunciation Evaluation on L2 Students in Thailand. *IAFOR Journal of Education*, 9(3), 41-56.
https://doi.org/10.22492/ije.9.3.03

Ngo, T. T. N., Chen, H. H. J. & Lai, K. K. W. (2023). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 1-18.
https://doi.org/10.1017/S0958344023000113

Papin, K. & Cardoso, W. (2022). Pronunciation practice in Google Translate: focus on French liaison. *Intelligent CALL, granular systems and learner data: short papers from EUROCALL 2022*, 322-327. https://doi.org/10.14705/rpnet.2022.61.1478

Pokrivčáková, S. (2019). Preparing teachers for the application of AI-powered technologies in foreign language education. *Journal of Language and Cultural Education*, 7(3), 135-153. https://doi.org/10.2478/jolace-2019-0025

Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *Tesol Quarterly*, 55(3), 866-900.
https://doi.org/10.1002/tesq.3027

Setter, J. (2015). Using flipped learning to support the development of transcription skills among L1 and L2 English speaking students. *Phonetics teaching and learning conference proceedings,* 77-82. https://doi.org/10.1080/02699206.2017.1350882

Sharipova, N. K. & Kakhkhorova, N. A. (2022). Using the" Flipped Classroom" Mixed Learning Model for the Formation of Phonetic and Phonological Competence. *International Journal on Orange Technologies,* 4(5), 67-70.
https://researchparks.innovativeacademicjournals.com/index.php/IJOT/article/view/4214/4369

Sidgi, L. F. S. & Shaari, A. J. (2017). The effect of automatic speech recognition EyeSpeak software on Iraqi students' English pronunciation: a pilot study. *Advances in Language and Literary Studies*, 8(2), 48-54.
https://journals.aiac.org.au/index.php/alls/article/view/3321

Sze, C. A. & Nasri, N. M. (2022). The Role of the Flipped Learning in the English Language Learning and Students' Motivation: A Systematic Review. *International Journal of Academic Research in Progressive Education and Development,* 11(2), 1122-1154. https://doi.org/10.6007/IJARPED/v11-i2/13932

Szyszka, M. (2022). Learners' attitudes to first, second and third languages pronunciation in structuring multilingual identity. *Applied Linguistics Review*, 13(6), 1127-1147. https://doi.org/10.1515/applirev-2019-0061

Tejedor-García, C., Cardeñoso-Payo, V. & Escudero-Mancebo, D. (2021). Automatic speech recognition (ASR) systems applied to pronunciation assessment of L2 Spanish for Japanese speakers. *Applied Sciences*, 11(15), 6695. https://doi.org/10.3390/app11156695

Turan, Z. & Akdag-Cimen, B. (2020). Flipped classroom in English language teaching: a systematic review. *Computer Assisted Language Learning*, 33(5-6), 590-606. https://doi.org/10.1080/09588221.2019.1584117

Vančová, H. (2020). *Pronunciation Practices in EFL Teaching and Learning*. University of Hradec Králové, Gaudeamus Publishing House. https://pdf.truni.sk/sites/default/files/katedry/kaj/vancova_pronunciation-practices.pdf

Vančová, H. (2021). *Teaching English pronunciation using technology*. Kirsch-Verlag.

Veres, S. & Muntean, A. D. (2021). The flipped classroom as an instructional model. *Romanian Review of Geographical Education*, 10(1), 56-67. https://doi.org/10.23741/RRGE120214

Wahib, M. & Tamer, Y. (2021). Exploring the effect of the flipped classroom model on EFL phonology students' academic achievement. *International Journal of Language and Literary Studies*, 3(2), 37-53. https://doi.org/10.36892/ijlls.v3i2.581

Wang, Y. H. & Young, S. C. (2015). Effectiveness of feedback for enhancing English pronunciation in an ASR-based CALL system. *Journal of Computer Assisted Learning,* 31(6), 493-504. https://doi.org/10.1111/jcal.12079