Review Journal of Social Psychology & Social Works

http://socialworksreview.com

ISSN-E: 3006-4724 Volume: 3
ISSN-P: 3006-4716 Issue: 3 (July - September, 2025)

Improving Pronunciation Accuracy, Stress, and Intelligibility with Digital Tools: Evidence from Speech Texter Intervention

Muhammad Moazzam Khan[1], Gull-e-Nayab[2], Zia Ullah[3], Rahat Munir[4]

DOI: https://doi.org/10.71145/rjsp.v3i3.404

Abstract

This study investigates the effectiveness of speech-to-text technology in enhancing students' pronunciation skills, with a specific focus on the Speech Texter application. The research aims to investigate whether the use of this digital application can produce measurable improvements in pronunciation accuracy, stress, rhythm, and overall intelligibility among English language learners at Undergraduate level studying in 2nd Semester of BS Accounting and Finance at Hailey College of Commerce, Punjab University Lahore. 20 students were selected following the purposive sampling technique. A quasi-experimental design was used, involving 20 students who participated in a structured intervention using Speech Texter over a period of four weeks. Pre-test and post-test assessments were conducted to evaluate changes in their pronunciation performance. The results revealed a significant improvement in students' pronunciation scores, indicating that the tool positively impacted their articulation, stress patterns, and clarity of speech. These findings suggest that integrating speech-to-text applications into language learning can provide immediate feedback, enhance learner autonomy, and support the development of more intelligible spoken English. The study recommends incorporating such digital applications into pronunciation practice as a supplementary method in ESL/EFL classrooms.

Keywords: Speech-To-Text, Pronunciation Accuracy, Stress, Rhythm, Intelligibility, Articulation

Introduction

To achieve effective communication and linguistic competence, pronunciation is a crucial element and thus holds the status of being one of the core elements in the process of learning and teaching English (Gilakjani, 2012). As the ultimate goal of learning language is the achievement of communicative competence, poor pronunciation may hinder intelligibility, limit social interactions and reduce the confidence of learners to communicate through the oral language (Munro and Derwing, 2015). These communication breakdowns can be caused by lack of intelligibility, which can lower the self-esteem of the learners and reduce their academic as well as career prospects. As compared to grammar and vocabulary, pronunciation needs learners to both identify phonology patterns and to pronounce them with proper stress and rhythm and intonation. It makes pronunciation one of the most difficult skills that English as a Second Language (ESL) and English as a Foreign Language (EFL) learners need to be skilled at. The traditional way of the pronunciation training is the listen-and-repeat type of the training, and the training is concerned with the suprasegmental and with the segmental one. These methods however lack systematic corrective feedback and rely on the low level of self-monitoring and self-evaluation capabilities of the learners as far as speech is concerned. Nevertheless, new technological systems have appeared in modern times, which can be used to teach pronunciation (Saragih, 2022). Such systems unite the application of hi-tech elements such as data gathering, data analysis devices and data feedback therefore offering the learners the chance of being more precise in their pronunciation, more effective at the same time. The

availability of new opportunities in the sphere of language teaching and learning has been introduced because of the increased availability of online-based tools. One of such tools is the speech-to-text application that offers special advantages of visual feedback on the speech output to present learners. The provided study focuses on the question of how Undergraduate Pakistani students may develop pronunciation accuracy, stress, rhythm, and intelligibility by using Speech Texter as a speech to text software.

Specifically, this research addresses the following questions:

1. Does the use of Speech Texter lead to measurable improvements in students' pronunciation accuracy?
2. How does the intervention affect learners' ability to use appropriate stress and rhythm?
3. Does the tool enhance overall intelligibility and clarity of speech?

Literature Review

Pronunciation cannot be defined merely as the ability to produce single sounds properly but entails suprasegmental characteristics, including stress, rhythm, intonation, which are important in terms of intelligibility (Gilakjani and Sabouri, 2016). Intelligibility, which is sometimes regarded as the extent to which the speech can be understood by the listener has been regarded as a more realistic pedagogical objective than being able to speak the language in a native-like manner (Munro and Derwing, 1995). Such learners who cannot be accurate in the production of individual sounds or who use stress and rhythm inappropriately tend to have a problem with making themselves intelligible, no matter how good their vocabulary and grammatical competence are (Gilakjani and Sabouri, 2016). Historically, the teaching of pronunciation was based on teacher feedback, imitation and drills. Although they worked in certain cases, these methods have been criticised as being monotonous and teacher centred (Celce-Murcia et al., 2010). In addition, the learners hardly got a chance to practise beyond the classroom with enough corrective feedback. The development of computer-assisted language learning (CALL) and the mobile-assisted language learning (MALL) have revolutionised the teaching of pronunciation. PRAATT, Rosetta Stone and speech recognition applications allow learners to do some personal practise, obtain feedback and become more motivated (Neri et al., 2008). It has been argued that learners do well when timely feedback on their pronunciation errors is given, as this assists them to become aware of the difference between their intended and actual pronunciation (Saito, 2012). Speech-to-text (STT) applications convert oral words into written text and their speech which gives the learner immediate feedback on pronunciation performance. The software does not recognise the words when they are mispronounced, or replaces them with other words. This live transcription enables the learners to keep a track on the articulation, stress and rhythm. Nevertheless, there are comparatively limited empirical studies that have explored the use of STT tools in South Asia, especially in Pakistan, where the problem of pronunciation is further complicated by the effect of the first language (Rahman, 2020). The study fills this gap by assessing the effects that Speech Texter has on Pakistani undergraduate students' pronunciation.

Methodology

Research Design

The study followed a quasi-experimental design with pre-test and post-test to measure the efficacy of the Speech Texter application to improve the pronunciation ability of the students. The study was conducted within a four-week long intervention with the undergraduate students studying at the University of Lahore. The learners interacted with Speech Texter through guided sessions that would help them improve articulation, stress, rhythm and intelligibility of the spoken English. The quasi-experimental design was considered appropriate in this study because it enabled to systematically study the effects of the intervention and take into

consideration the practical limitations of operating in an educational environment. This design has been extensively utilised in studies of the second language acquisition to evaluate the effectiveness of the digital tool in the development of oral proficiency (Creswell and Creswell, 2018).

Participants

The sample size used in the study was 20 students who were purposely chosen among the 3rd semester students of the BS English programme at the University of Lahore. Purposive sampling was regarded to be suitable since it enabled the researchers to select the participants strategically to help realise the aims of the research and the subjects had the relevant nature. The students in this case were typical ESL learners with intermediate level of fluency on the English language, and thus they were good candidates to be included in the study of the effectiveness of a pronunciation-oriented intervention. Moreover, all the participants were not trained in pronunciation before in any formal way, except the one in the classroom instruction, which guaranteed a relatively homogenous group with regards to previous exposure. This was a controlled selection procedure which made the findings more valid since it reduced the possibility of confounding factors associated with prior specialised training. According to Patton (2015), the purposive sampling method is especially useful in research in the educational field as a researcher intends to obtain detailed insights into the particular learner groups.

Instrumentation

Intervention Tool

Speech Texter application was utilised in this study as the major intervention tool. This computerised application is mobile based, i.e. it is used to translate verbal representation to written text, in this way, the learner can keep track of their pronunciation precision on the spot. It was a suitable tool in pronunciation practise among the undergraduate students because of its open-ended nature and easy interface.

Pronunciation Test

The researchers were keen to determine how students had progressed and thus came up with a pronunciation test which revolved around four important dimensions of spoken English, namely, articulation, stress, rhythm, and intelligibility. The test was created in such a way that it was able to capture the segmental and suprasegmental aspects of speech and a comprehensive evaluation of the pronunciation performance of learners was done.

Rating Rubric

The rating rubric based on Derwing and Munro (2005) was used to test the learner performance. The rubric used Likert scale of five points between 1 (poor) and 5 (excellent) with four points in four categories of accuracy, stress, rhythm and intelligibility. This scheme provided the opportunity to assess particular features in details and also assess the quality of pronunciation on the whole.

Research Procedure

To achieve the reliability and validity of the findings, the study procedure was done in three stages. During the pre-test, the participants were asked to read a set passage out loud. Their oral readings were taped in a consistent setting and graded with the help of a developed scoring rubric. This move offered a valid baseline of pronunciation and fluency and created consistency between the participants. The intervention process took place in the period of four weeks where students were subjected to use of Speech Texter application thrice a week in about 30 minutes sessions at a time. The activities involved reading passages prepared, word lists practising, and free speech exercise with the transcription accuracy of the software being observed. In order to

increase the validity, the identical kind of tasks were used to all participants, and the learners were provided with systematic instructions on how to analyse the differences between what they say and what is written in the text. This process brought about self-monitoring and reflection coupled with making sure that the treatment was applied uniformly. During the post-test stage the respondents were required to engage in the same standard reading task that was deployed in the pre-test. Their performances were once again tape-recorded and assessed with a similar rubric and same conditions like the pre-test. This consistency of the methodology enhanced the validity of the comparison of the pre- and post-test outcomes that made it possible to draw appropriate conclusions about the effect of the intervention.

Quantitative Data Analysis

Paired-sample t-tests were used in SPSS to analyse quantitative data in terms of pre-test and post-test to understand whether the post-test and pre-test differences in the pronunciation performance of the learners were statistically significant. This statistical test in the study was suitable because it could directly compare the identical sample of population before and after the intervention, hence individual variability was suppressed. Besides quantitative, qualitative observations were also taken systematically during the intervention period. These were the field notes of the instructor on the intensity of the engagement of the learners, their responsiveness to the feedback and the application used by the learners through the Speech Texter application during the practise sessions. Additionally, self-reported experiences were also considered; they provided the learners with a perspective on their understanding of the progress, and difficulties and the overall utility of the tool in the pronunciation practise. Combining both quantitative and qualitative evidence has given a more detailed idea of the effect of the intervention so that the statistical results could be justified by the experiences of the learners and the observation of their behaviours. This combination of methods is in line with the focus on the necessity of Creswell and Creswell (2018) to mix quantitative data with qualitative information to enhance the validity and the richness of research results.

Qualitative Data Collection

Student self-reports were used as a qualitative source of data collection in order to get the perceptions and experiences of learners. The use of written self-reports instead of other qualitative means of observations, like interviews or classroom observations, was not accidental. Self-reports gave the learners some privacy and time of reflection in which they could express themselves freely without feeling pressured by being in the presence of the researcher. This method reduced possible social desirability and the students were able to talk about their problems and achievements freely. Moreover, the written reflections were more feasible in the time limits of the four weeks intervention since it did not demand much scheduling and could be done in less time at the convenience of the learners. Self-reports provided a direct access to the internal perspective of the learners of pronunciation challenges, their feelings, strategies, and self-awareness in comparison with classroom observations, which capture mainly outward behaviours. This approach was quite consistent with the hyper-independent character of the study and the focus on autonomy and personal consideration without the loss of any meaningful qualitative data to supplement the quantitative results.

Towards the conclusion of every week of intervention, the participants were to give brief written comments on how they used Speech Texter. Such considerations prompted learners to make remarks about various things, such as the words or sounds that they had the hardest time pronouncing, how they interpreted what the application had to say to them and how they went about correcting the inaccuracies. Students were also asked to explain how they felt after the immediate feedback, i.e., did it make them want to practise more or were they frustrated, and to report whether they practised more independently when they were not in the classroom. These written reflections were incredibly informative about the knowledge of learners about their own problems in pronunciation and how they interacted with the speech-to-text

technology. They also emphasised the metacognitive strategies used by the learners including stress patterns surveillance, detection of repeated mistakes, and the ability to correct oneself in accordance with the accuracy of transcription. Notably, the self-reports provided a testimony of augmented learner autonomy, since the students frequently commented on how they did their practise by choice outside the classroom environment, and utilised the flexibility that the mobile tool provided. Therefore, the qualitative data acquired by use of self-reports were supplementary to the quantitative test outcomes, which gave an opportunity to understand how Speech Texter influenced the performance of learners, as well as their attitudes to the pronunciation learning.

Results and Discussion

Quantitative Findings

The results of the quantitative analysis showed that all four dimensions of pronunciation, namely, accuracy, stress, rhythm and intelligibility, improved significantly, which substantiates the effectiveness of the Speech Texter intervention. Table 1 shows the dimensions along with the values of pre-test, post-test, and p-value against each dimension.

Table 1 Quantitative Findings of Pre-test and Post-test Scores on Pronunciation Dimensions

Dimension	Pre-test Mean	Post-test Mean	*p*-value
Pronunciation Accuracy	2.15	3.90	< 0.01
Stress	2.05	3.65	< 0.01
Rhythm	2.10	3.75	< 0.01
Intelligibility	2.20	4.00	< 0.01

Note. Scores were measured on a 5-point scale (1 = poor, 5 = excellent). All post-test improvements were statistically significant at $p < 0.01$.

The analysis of pre-test and post-test scores showed that there were significant improvements in all the four dimensions of pronunciation. Concerning the accuracy of pronunciation, the mean score rose slightly in the pre-test (2.15) and in the post-test (3.90) ($p < 0.01$), which also means significant improvement in the pronunciation skills of the learners and a decrease in phonological errors. Equally, the learners registered significant improvement in the stress segment with mean score increasing by 2.05 to 3.65 ($p < 0.01$). This enhancement is an indication that the intervention was effective in making learners lay more emphasis on syllables and words in a more appropriate manner and hence, generate speech patterns that were more similar to those of individuals who speak English well. The rhythm scores also had a significant improvement in the mean score, as the mean score rose to 3.75 in the after-test, compared to 2.10 in the pre-test ($p < 0.01$). One of the most important features of fluency is the ability to retain a natural flow and timing of speech and the findings show that learners became more proficient in the given aspect once the Speech Texter application was used. Figure 1 given below presents a graphical representation of comparison between all four dimensions.

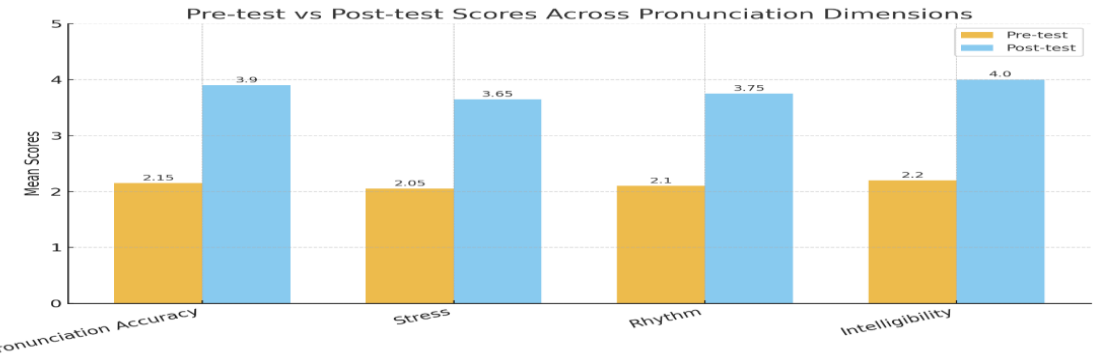

Figure 1

The most significant improvements were however noticed in the intelligibility where the scores improved to 2.20 in the pre-test and 4.00 in the post-test ($p < 0.01$). This implies that the general speech of the learners was much more understandable to the listeners and that is the ultimate aim of the teaching of pronunciation. Intelligibility increased most of the four categories and emphasises the feasible effect of the intervention in increasing the communication skills in the real life. Although a combination of accuracy, stress and rhythm led to these improvements, the spike in intelligibility indicates that learners could combine these aspects in an enhanced way leading to a better and more lucid verbal output. All these enhancements show that the application of Speech Texter offered both focused practise and valuable feedback that allowed the learners to monitor themselves and improve on their pronunciation. The statistically significant findings prove that the positive changes were not accidental but the direct output of the specified intervention, and it is an excellent argument to include mobile-assisted means in the process of learning L2 pronunciation.

Qualitative Observations

Along with the statistical changes, the qualitative observations supported and self-reported experiences of learners were helpful to understand how successful the use of Speech Texter as a pronunciation tool could be. Student feedback produced four themes that were revealed by common themes, showing the impact the application had on their learning. Table 2 shows the various themes extracted from the students' reports:

Table 2 Themes from Qualitative Findings with Supporting Evidence and Sample Quotes

Theme	Supporting Evidence	Example Quote from Students
Immediate Feedback	Real-time transcription motivated learners to notice and correct errors instantly.	*"When the app didn't recognize my word, I knew I had said it incorrectly, so I tried again until it was right."*
Identification of Mispronounced Words	Learners became aware of recurring pronunciation problems and targeted them for practice.	*"I realized I always mispronounced the word 'technology,' and the app highlighted it every time."*
Awareness of Stress Patterns	Incorrect stress often led to distorted transcriptions, making learners conscious of rhythm.	*"When I put the stress on the wrong syllable, the word appeared differently in the text, so I paid more attention to stress."*
Encouragement of Independent Practice	The tool provided flexibility for autonomous practice outside classroom hours.	*"I used the app at home in the evening. It felt like extra practice without waiting for the teacher."*

Note. Themes were generated from learners' self-reported experiences with Speech Texter during the four-week intervention.

To begin with, students always underlined the importance of instant feedback, stating that transcription in real time encouraged them to learn more and more often and systematically. One of the respondents said, *"when the app failed to know my word, then I knew that I was saying it wrong, and I would reword it until it was identified as correct."* This display of mistakes promoted proactive correction and students felt that it was more effective compared to late feedback by instructors. Second, the respondents stated that the application made them discover the regularly mispronounced words, which made them aware of the regularly occurring weaknesses. As an example, a student mentioned that she always learned to pronounce the word technology incorrectly, and the app reminded her every time. This recognition provided a chance to practise specifically and helped students to track their progress in the long term. The third theme was an increased awareness of patterns of stress. Students elaborated that stress when misplaced tended to lead to incorrect or inaccurate transcriptions and therefore they had to be aware of syllable emphasis and word rhythm. One of the learners mentioned, *"I concentrated more on stress because when I stressed the other syllable the word came out different in the text."* This heightened awareness helped in natural and intelligible speech. Lastly, the participants valued that Speech Texter promoted practising outside of the classroom. The user-friendliness of the application enabled them to practise passages, word lists, and free speech in relaxed environments, which enabled independence and more confidence in their pronunciation skills. As one of the respondents observed, *"I was using the app at home in the evening. It was such a workout without having to wait until the teacher."* Collectively, these qualitative data supplement the quantitative data because the results prove that Speech Texter helped not only to enhance measurable areas of pronunciation but also to promote the motivation, self-awareness and autonomy of the learners-the factors that should help them to develop language in a long-term perspective.

Discussion

The findings confirm that speech-to-text technology, in particular, Speech Texter, can be used to a considerable extent in enhancing pronunciation in learners. The present research results coincide with the argument of the literature on the problems of multidimensionality of pronunciation. Gilakjani and Sabouri (2016) suggest that pronunciation does not just involve the possibility of producing single sounds but mastering suprasegmental factors, such as stress, rhythm, and intonation, which are pertinent to the intelligibility. The results showed that the learners indicated significant gains in all these domains where post-test scores recorded a significant rise compared to pre-test. As an example, the pronunciation accuracy improved to 3.90, stress to 3.65, rhythm to 3.75 and intelligibility to 4.00 (all $p < 0.01$). These gains indicate that suprasegmental practise results in general intelligibility which supports the argument by Munro and Derwing (1995) regarding intelligibility as the main pedagogical objective as opposed to native-like pronunciation. Intelligibility was improved the most, which confirms previous claims according to which students with false stress and rhythm are usually unable to be heard, irrespective of their grammatical and lexical proficiency (Gilakjani and Sabouri, 2016). On the other hand, the research proves that learners are more capable of combining both segmental and suprasegmental features in case of systematic practise and instant feedback, which results in more understandable and intelligible speech. This observation also answers the historical criticism of the classic methods of teaching pronunciation that were mostly teacher-centred, founded on imitation and drill, and provided the student with minimal possibilities to practise independently (Celce-Murcia et al., 2010). Comparatively, using Speech Texter enabled the learner to practise independently and simultaneously to give corrective feedback which made the process more interesting and effective. These findings also coincide with Neri et al. (2008) and Saito (2012), who claim that pronunciation training using technology improves the motivation of learners and their phonological awareness after providing them with feedback in time. Specifically, the Speech-to-Text feature of Speech Texter served as a diagnostic device, with real-time feedback, showing errors as an inaccurate

or missing transcription, thus, requiring the learner to correct him/herself. This echoes the view of Saito (2012) that the instant feedback is of high importance in enhancing the phonological awareness of learners. Moreover, the notable gains documented in this work are also providing empirical data to the lack of studies on speech-to-text (STT) applications in South Asia, which is also noted by Rahman (2020). The study confirms the theoretical claims regarding the positive effect of Speech Texter on Pakistani undergraduates; therefore, it will not only widen the empirical foundation of technology-based training of pronunciation in the environment of L1 interference, but also prove the truth of earlier theoretical proposals on the effect of the latter. In conjunction with these statistical gains, qualitative observations also gave us a better understanding of the way learners perceived the intervention. The four major themes exhibited by student reflections represent the pedagogical value of Speech Texter. Firstly, students always pointed out the significance of instant feedback as they said that practising in real-time made them be more likely to do it more often and systematically. As one of the respondents described, in case the app did not recognise my word, I knew that I was pronouncing it incorrectly and as such I would rephrase until the correct word is recognised. This instant feedback of errors encouraged active correction and was seen to be more productive than later instructor feedback. Second, the learners described the way the application assisted them in recognising commonly misspelt words, which created more awareness of the recurring weaknesses. Indicatively, one of the students said, I was aware I was always pronouncing technology wrongly, and the app kept on reminding me. This focused attention created an opportunity to practise and track the progress. The third theme was the increased sensitivity to stress patterns. Being the cause of incorrect transcription in most cases, misplaced stress led students to be more attentive to the syllable emphasis and word rhythm. One learner shared the view that she focused more on stress since the word differed in text whenever I stressed the other syllable. This increased awareness led to the use of natural and understandable speech. Lastly, the students appreciated the encouragement of independent practise out of the classroom by Speech Texter. Its availability and ease of operation enabled them to train passages, word lists, and free speech in non-challenging environments and help them develop independence and self-confidence. According to one of the participants, he was using it at home in the evening. It was an exercise without necessarily waiting until the teacher. Taken together, these qualitative results complement the quantitative information in that Speech Texter did not only positively impact the quantifiable aspects of pronunciation, it also helped learners to become more motivated, more self-aware, and independent all of which are key to the development of language on a long-term basis. Overall, this paper confirms previous theoretical statements about the importance of the centrality of intelligibility and timely feedback and also adds new empirical details to this subject on a South Asian setting. The findings indicate that mobile-assisted devices like Speech Texter could be effectively used to complement the traditional classroom teaching approach to provide the learners with higher degrees of independence, motivation and accuracy in their pronunciation growth.

Pedagogical Implications

Speech Texter should be viewed as a supplementary tool rather than a replacement for teacher feedback, as the guidance of instructors remains essential for nuanced correction and deeper understanding of pronunciation features. However, its accessibility and cost-free nature make it a valuable low-cost solution, particularly in resource-limited contexts such as Pakistan, where educational institutions often face budgetary and technological constraints. Moreover, by encouraging students to engage in independent practice, Speech Texter and other speech-to-text (STT) tools play a vital role in fostering learner autonomy and promoting lifelong learning habits. This combination of teacher support and technological assistance creates a balanced learning environment where students benefit from both structured classroom instruction and flexible, self-directed practice.

Conclusion

This study aimed at establishing the effectiveness of Speech Texter, a speech-to-text application, in enhancing pronunciation abilities of undergraduate students in Pakistan. The results prove that the tool contributed greatly to the accuracy of pronunciation, stress, rhythm, and the general intelligibility in learners in a relatively very limited period of four weeks. The quantitative outcomes, which were proven by the help of paired-sample t-tests, and the qualitative considerations confirm that Speech Texter was able to give instant and relevant feedback, help the students recognise the common pronunciation difficulties, become more aware of the stress and rhythm patterns, and start practising independently without classroom supervision. Collectively, these results demonstrate the pedagogical importance of using digital tools in learning the English language. The study also demonstrates that the most significant progress was revealed in intelligibility the ultimate objective of pronunciation instruction. This result supports the previous ones that the primary goal of pronunciation teaching should be intelligibility and not native-like perfection. Speech Texter overcame the traditional drawbacks of pronunciation teaching that have long been experienced by traditional methods by using real-time feedback and learner agency, which frequently uses tardy teacher feedbacks and limited practise of pronunciation skills. Notably, the research is relevant to ESL/EFL settings of resource-limited setting such as Pakistan. Being a convenient and affordable solution, Speech Texter may be used in the classroom to complement the work done by teachers and help students become more independent and self-motivated in language learning. Although teacher feedback is still highly important in offering more detailed feedback, speech-to-text technology is a viable and scalable method of expanding practise opportunities out of the classroom. Simultaneously, the study admits its limitations such as limited sample size, limited time, and the use of one application. These reasons limit the external validity of the findings and indicate that more studies are required using larger, more varied samples, more extended interventions, and comparative studies of different speech-to-text programmes. It is also possible that future research will examine the effects of long-term usage of such applications on the long-term effect of learners to improve their pronunciation and confidence in communicating in real life. To sum it up, this study offers empirical data that speech-to-text programmes such as Speech Texter can be used as effective adjunctive devices in the pronunciation training process. This study demonstrates that technology-enhanced teaching of pronunciation can help to change the traditional ways of working in a classroom, motivate the learners to achieve independence and, finally, make the English speakers in ESL/EFL settings more intelligible and confident speakers.

References

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge University Press.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approach* (5th ed.). Sage Publications.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*(3), 379–397. https://doi.org/10.2307/3588486

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins.

Gilakjani, A. P., & Sabouri, N. B. (2016). Learners' listening comprehension difficulties in English language learning: A literature review. *English Language Teaching, 9*(6), 123–133. https://doi.org/10.5539/elt.v9n6p123

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Sage Publications.