Examining the Use of Speech Recognition Technology to Improve Pronunciation Proficiency among Learners in Saudi Arabia

Aayesha Sagir Khan

Assistant Professor, College of Languages and Translation, King Khalid University, Abha Kingdom of Saudi Arabia, Aasakhan@kku.edu.sa

Received 6-1-2024 Accepted 1-2-2024 publish 20-3-2024

Abstract

The present research aims at determining the impact of Speech Recognition Technology (SRT) on the improvement of Saudi learners' pronunciation skills. Pre and post tests were applied after 6 weeks with a total of 60 undergraduate students who used SRT tools during the study. Finally, the outcomes revealed enhanced segmental pronunciation for all the participants with consideration to gender and proficiency level. While some disclosed positive outcomes, certain restrictions were pointed out for suprasegmental features of pronunciation. It reveals the rule of practice and the possibility of SRT as an effective supplement to the language training course. This evidence implies that the combination of SRT with conventional instructional techniques can provide a suitable means for achieving pronunciation enhancement, although additional research is required to examine the extended impact and usability of the SRT tools.

Keywords: SRT, pronunciation, language acquisition, Saudi learners, ed tech

Introduction

Pronunciation is established universally as one of the key aspects of second language learning; it has a direct bearing on how comprehensible the learners' speech is and, as such, their communicative ability. However, pronunciation continues to be a major difficulty for most learners especially for those from Other Arc Arab (Alghazo, 2015). Saudi Arabian learners who rely on English in the academic and career settings consistently struggle with pronunciation, partly because they are exposed to only limited amounts of spoken English and the poor instruction methods (ur Rahman & Alhaisoni, 2013).

Implementation of technology has provided new possibilities of improving the pronunciation instruction, and speech recognition technology (SRT) has attracted interest as it allows giving immediate and adjustable feedback (Ehsani & Knodt, 1998). Google Speech, Duolingo, and other similar SRT applications proclaim that they help learners to enhance the correct pronunciation by analyzing and correcting their pronunciation mistakes in real-time (Farias Matias & Orrala Figueroa, 2024). However, though these claims are promising, the evaluation of the practical use of these indicators with the variety of learning environments, including Saudi Arabia, is limited (Alrabai 2016).

While several works focus on the possibilities of using SRT for changing the language learning pattern (Granena, 2013; Chapelle, 2005), some authors are pointed out that SRT could be useless as a tool, which helps in complex aspects of pronunciation. Walker further pointed out that notwithstanding the mechanical correct of segmental features of the Second Language, SRT can be ineffective in the teaching of suprasegmental aspects of the Second Language, specifically stress, rhythm, and intonation – important aspects of real-life communication. Also, Zimmerman & McMeekin (2019) express concerns about overuse of SRT to teach pronunciation arguing that such approaches may even contribute to further development of the rote learning instead of comprehensiveness of pronunciation.

In the case of Saudi Arabia, we see that passive approaches, such as the teacher-centered instructional practices typical for the region, pose obstacles to SRT. Analyzing Saudi learners' learning behavior, researchers have observed that the use of communicative approach is limited and Saudi learners tend to memorize and translate, rather than using adequate speaking practice and pronunciation (Al-Seghayer, 2020). Secondly, ur Rahman & Alhaisoni (2013) argue that communicative practices that inform classroom practices in English are hobbled and that this hobbles learners' pronunciation proclivities. However, the adoption of SRT indeed has the potentials to

provide the solutions which should be in Saudi context taking in to consideration the readiness level of learners for the autonomous learning education and the use of technology integration in Saudi context (Alodwan, 2021).

Still, following the current trend in the field of TALL, the given state of research oversights the empirical evidences concerning the efficiency of SRT for the Saudi learners' accent improvement in Saudi Arabia. Previous research in other settings reveals that output is mixed. For example, Pandian et al. (2014) identified the effectiveness of SRT in enhancing the Chinese learners' performance in segmental pronunciation but only slight effect in suprasegmental aspect of pronunciation. Similarly, in the study by Alstein et al. (2021) the authors explain that while SRT improves learner activity it still needs other additional methods to make it effective. In Saudi Arabia, however, such fine-grained inquiries are missing, which leads to questions related to how SRT could best be applied in Saudi Arabia.

There is a particularly significant research loss when it comes to methodological perspectives in prior investigations. A vast majority of the research is qualitative or mixed, paying more attention to the learner's subjective experiences rather than concrete results. Nevertheless, these findings do not offer solid basis for arguing about the effectiveness of SRT in raising the level of pronunciation proficiency. Pre- and post-test quantitative analyses are called for to determine data causality and achieve external validity (Cook & Campbell, 1986).

Thus, the lack of concern with sociocultural and pedagogical factors underlying the belief in the general efficacy of SRT is problematic. For example, learners in KSA may face challenges in self-mastery facilitated learning models because of their inclination to more directive learning-theoretical paradigm (Alrabai, 2016). Thirdly, always accessible SRT tools and their relevance to the presented linguistic demands do not remain free from doubt. This is in agreement with Al- Kutubkhanah Alsaeid (2011) who also opine that most of the currently availed SRT applications are orientated for natives or even for the learners at an advanced level and may thus not be very helpful for the introductory level Saudi learners.

This research aims at filling this gaps by assessing the use of SRT to enhance the Saudi learners' pronunciation skills. By using quantitative approach, it intends to produced factual result in support its general assertion on how SRT can further improve the students' ability in accent, which encompasses segmental and suprasegmental aspects. Furthermore, this research emphasizes the importance of contextualizing SRT use within the Saudi educational system, considering learners' cultural and pedagogical backgrounds.

The Problem of the Study

Pronunciation students are still a problem in many ESL learners, especially for learners in Saudi Arabia where the Phonetics Arabic and English are quite different. They find it challenging to attend to segmental aspects of a given language – individual sounds and suprasegmental aspects which include stress and intonation for speaker's intelligibility. Nevertheless, there has been poor emphasis on the use of these features or other aids such as the speech recognition technology (SRT) in Saudi Arabic language classrooms. Sit-down learning methods and rate learning and most of them do not care of the practicability of the language learning which is oral communication. Moreover, the literature review was unable to establish clear trends regarding the efficacy of SRT for enhancing learners' pronunciation and, particularly, the Saudi learners. These challenges underscore the importance of this study's aim to determine if, and how, SRT can improve Saudi learners' pronunciation competence, which is a gap in the existing literature and practice.

Questions of the Study

1. To what extent does the use of speech recognition technology improve segmental pronunciation accuracy among Saudi learners of English?
2. How effective is speech recognition technology in enhancing suprasegmental features such as stress, rhythm, and intonation in Saudi learners' pronunciation?
3. What are the perceptions of Saudi learners regarding the usability and effectiveness of speech recognition technology for pronunciation practice?

Significance of the Study

The findings of this study have much implication for several actors in language education, learners, teachers, and policymakers in particular. For teachers it provides information on how the technology can be used to help learners at the time of pronunciation so that they can deliver good, confident English. Due to the evaluation of the implementation of SRT, this research adds insight into the field of TALL and offers practical teaching strategies for improving pronunciation learning and teaching methods. To the policymakers, therefore, the study points to the need to incorporate more technology enhanced methods such as SRT in the national coalition in an effort to enhance Saudi Arabia's Vision 2030 initiatives of enhancing education technology and English comprehension. Furthermore, as analysis of this research, this work fills a significant gap in the literature by providing a quantitative evaluation of the effectiveness of SRT, including valid and reliable results that can be used to guide further enquiry and practice-based learning in different learning environments.

Terms of the Study

This study was carried out for 8weeks and the instructional method utilised was speech recognition technology (SRT) aimed at enhancing pronunciation achievement of Saudi learners of English. The participants in the study were 40 undergraduate students, learning English as a second language, which made them especially important for assessing the attending in Saudi higher education that SRT could provide. The investigation was restricted to segmental and suprasegmental phonetic analysis, which is consistent with the goals of the present study, as it aimed at evaluating both segmental and suprasegmental aspects of pronunciation.

The intervention consisted of guided practice with commercially available SRT tools that gave immediate feedback as to how the learners' spoke English. Usability data was gathered through pre- and post-test word-sound identification tests to determine pronunciation gains and questionnaires that assessed the learners' views regarding the usefulness and relevance of SRT. Furthermore, the SMART goals formulated for the study propose practical implications mainly to the Saudi Arabian context of education, with potential generalization in the context of ESL learning settings. The work does not include young learners or learners who are out of school, however, the study employs a targeted and practical methodology which provides meaningful recommendations for the use of technology in language instruction.

Limitations of the Study

Nevertheless, this research has the following limitations: First, conversational data collected from a particular group of Saudi undergraduate learners does not reflect the generalizability of the results on other groups of learners, high school learner including those from different cultural background, or different language learners. Second, the fact that the study used commonly available SRT applications that lack the linguistic adaptation to the Arab learners' learning context might impact the feedback's reliability and the learners' improvement. Third, the time span of the intervention was relatively small (8 weeks), and thus did not allow for the investigation of potential long-term attibutional and pronunciation gains from using SRT. Last, dynamic variables including the background knowledge of participations where English is commonly used or extra resources used during the learning process might also affect the results and bring variation. Due to such limitations, the study acknowledges them to help place the results in perspective and lay the ground for subsequent research to enhance on the weaknesses left behind.

Literature review and Previous studies

Pronunciation has been identified as a major component of second language acquisition in relation to intelligibility and communication Without, There is no question that second language acquisition and more broadly communication depends in part on pronunciation. Nonetheless, the issue of pronunciation is frequently underestimated in language acquisition, especially in countries such as Saudi Arabia, in which conventional models of knowledge transmission are still dominant (ur Rahman & Alhaisoni 2013). New developments in technology such as the Speech Recognition Technology (SRT) present unique approaches to the problem of pronunciation by providing immediat feedback and self-training practice for the learners (Lam-Yee-Mui et al., 2023).

Traditionally, pronunciation has been taught with methods such as audio lingual methods but more communicative approaches have been adopted. Sharma (2021) has clearly stated that for any kind of pronunciation instruction both the segmental and suprasegmental aspects are important. But the current classroom practices suggest that inadequate practice is offered to get this balance and further elaborating, the practice deficiency is acute in the settings that are predominantly teacher-centered such as Saudi Arabia (Alrabai, 2016). This gap reveals that the technology-assisted language learning (TALL) tool such as SRT can support the traditional learning approaches to language learning by providing learners with personalized and interactive learning solutions.

Automated self-practice through SRT is a recent discovery enhancing pronunciation as a language learning strategy of individual practice. Thus, Ehsani and Knodt (1998) Whilst highlighting the technical problems associated with SRT, the authors noted the possibilities to use it to teach a foreign language to indicate that the method can help learners receive immediate feedback on their pronunciation mistakes and work on their self-awareness and gradual improvement. Farrell (2015) recently established that SRT tools increase segmental accuracy by improving the ability of learners to distinguish between difficult sounds or segments when used for Chinese ESL learners.

To the best of the writers' knowledge, little research has so far has been conducted in the Saudi context focusing on the application of SRT. Nonetheless, Al-Seghayer (2020) investigated the apprehensive effects of technology on language skills comprehensiveness of learners toward digital technologies which were not aligned with precise and explicit focus on development of pronunciation applications. On the same note, Al-Senaidi et al. (2009) pointed out that even though learners gain from adopting technology in learning, Arabic is phonetically distinctive and thus; ergo , call for tools attuned to their phonetics. These results indicate that SRT could be effective in treating pronunciation problems but needs to be culturally adapted.

In fact, extensive investigation carried out reveal enhanced precision in the manner SRT influences on pronunciation. For instance, Saito & Saito (2017) examined the effect of SRT on suprasegmental on the Japanese learners' intonation and found a positive change the after eight week practice on the intonation and rhythm. Vančová (2020) complained the motivational advantages of SRT to increased and confident learners' practice of pronunciation in contrast to conventional practice.

Critics have pointed to some shortcoming in SRT's efficiency Though, capacidad critics consider limitaciones in SRT. According to Yenkimaleki et al. (2023), there is a strong tendency with SRT systems to focus on segmental accuracy rather than suprasegmental ones which are important in any communicative process. Johnson (2000) identified another drawback of SRT feedback that is the possible unnatural and mechanical articulation of languages because SRT might not emit speech patterns of normal speakers. These criticisms stress two major approaches to support SRT with more attention to teacher collaboration and meaningful practice in non-controlled settings to improve pronunciation learning.

In Saudi Arabia the use of technology has been mainly used to teach general language, including vocabulary, grammar and little regard given to pronunciation. More particularly, ur Rahman and Alhaisoni (2013) pointed out the limitations of the conventional approach, as they fail to dedicate a proper attention to the oral skills, and stressed the importance of integrating the more communication-oriented activity into the classroom practice. Similarly, Alodwan (2021) focused on learners' perceptions of the gh-use of digital tools in L2 learning and found a benign acceptance of these tools but acknowledged the barriers of accessibility and functionality.

While only a handful of studies have focused on examining the role of SRT in Saudi classrooms, Pennington & Rogerson-Revell (2019) stressed the urgency of the need for research that targets pronunciation form in a manner that those learning English wish to be relevant to them culturally and linguistically. In their study, they identified the SRT as being capable of increasing learner's participation and pronunciation standards but identified a complete dearth of studies that have tested the implications of the intervention in Saudi context.

Nonetheless, several gaps persist, according to prior research on SRT and literature on the teaching of pronunciation. First, while there is a dominant qualitative scholarship in analyzing the effects of SRT on Saudi learners' segmental and suprasegmental skills, there is a lack of quantitative studies that measures the gains that learners experience in the particular L2 feature as a result of using SRT. Second, learner's perception dominates

the current literature instead of realistic objective results therefore making the results not so generalizable. Last, Arabic speakers, which have issues with English vowels and stress, as well as, other people with unique learning profiles, still do not have sufficient attention paid to them in the context of the SRT application.

Methods

Research Design

The present research work adopted a quantitative research method to examine the effectiveness of Speech Recognition Technology (SRT) in enhancing pronunciation mastery by Saudi undergraduate learners of English. To assess pronunciation change differences, a pre-test and post-test quasi-experimental design was adopted while a comprehensive questionnaire was used to establish learners' perceptions about SRT. The rather solid structure allowed to collect the relevant and sufficient data necessary for the solution of the objectives of the present research.

Participants

The subjects were 40 Saudi undergraduate students that attend intermediate level English courses. Being purposively recruited, the participants were chosen in a way that captured learners who are constantly taught in English but rarely or never exposed to technological assistive devices to learn correct pronunciation. English pronunciation difficulties have different facets which can be best addressed by working with participants who are native Arabic speaking learners all of whom are postgraduate level learners. To increase the external validity, attempts were made to include the same amount of males and females and those with higher and lower speech confidence levels.

Speech Recognition Technology

The SRT application used in this study was selected due to its complex features, and ease of use in enhancing pronunciation. The outlined cross-sectional app was developed specifically to assist learners in practicing and improving their skills in English pronunciation, with the focus made on the provided spoken input, to which the learner receives instant, constructive feedback. These components followed segmental accuracy, which consists of vocalisation of simple sounds like a vowel and consonants, together with suprasegmental features such as intonation, stress and and rhythm that is crucial to making the English language appear natural and easily understandable during interspeech communications. These capabilities made the app suitable as a tool needed to tackle the issues phonetic by Arabic learners while on English.

Probably the most important part of the application was that it contained a feedback system that immediately displayed a visual and/or an acoustic prompt when a learner produced the word or a sound incorrectly. For instance, whenever the learner mispronounced a word it would underline the mistake in red and suggest the correct way of pronunciation and even show it through an audio play back. This real-time feedback allowed the learners to adjust their mistakes right away, which I believe helped with boosting their self-monitoring skills as well as keeping themselves corrected and in check on the correct pronunciation. The feedback system used point and click technology to give encouraging feedback for correct sounds produced and mild feedback for the mistakes made by the learner in order to produce an encouraging environment.

The last facet of the app was the phonetics one, which highlighted simple phoneme mistakes, including the most common difficulties of Arabic speakers in English learning, such as English vowels, and combinations of consonants. The app pointed out areas of phonetic misuse and supplied recommendations regarding several phonetic variations. For instance, if a learner tends to pronounce an incorrect vowel sound, then the app would underline the phoneme and pinpoint it and give an over view of the correct ways of pronouncing it. This phonetic feedback is important for the Arabic speaking learners as the two languages have contrasting systems, vowels and stress patterns included.

For progress it provided progress trace options, where the learner can get weekly report with all the results of the specific week, including the accuracy of work, the fluency of the text produced and some patterns of mistakes. These report were helpful for the learners and the researcher because they had frequent check-point on the progress of all the learners involved in the study. Students could watch themselves grow and know which aspects they had

been struggling with for a long time, for instance certain sounds or stress patterns requiring more attention. The feature that was found especially beneficial was in the capacity of the app to store and provide the analysis of any data on the performance of pronunciation; it was very helpful in tracking the overall progress and guaranteeing that the intervention provided was beneficial in terms of the development of pronunciation skills.

Procedure

The study followed a structured, eight-week timeline consisting of three primary phases: The students identified passed through an initial assessment, the intervention process, and follow-up assessment. Thus, the design enabled efficient comparison of pronunciation in the beginning and after the intervention and contributed to the collection of valid data and increased reliability.

Thus, all participants were pre-tested on pronunciation during the first week of the study in order to determine their initial level of the skill. In the pre-test participants were requested to read and record themselves saying 20 sentences, and a brief passage. These were subjected to the analysis of two trained raters utilizing a structure of pronunciation rubrics. This rubric considered various aspects, segmental and suprasegmental: segmental referring to the individual vowels and consonants, and suprasegmental referring to the stress, the rhythm and the intonation. This assessment was used as the starting point where participants' progress throughout the course of the study was to be measured from.

In the next six weeks, participants practiced using the SRT app as assisted reading application for 5 days/week for 30 minutes/day. The weekly practice tasks fit their course curriculum and so included areas related to pronunciation exercises, phonemes, and reading aloud. This app trained the participants on the best practice in the usage of the app, and engaged them to self- monitor their progress by evaluating their weekly records. Their practical enhancement was informed by the feedback mechanism of the app, enabling them to correct segmental and suprasegmental errors. In the course of the ambitious intervention, participants were constantly encouraged and motivated to use the app and track their progress systematically.

That researcher 8 in the last week of the research the participants carried out a post-test that in terms of the registration tasks and the evaluation criteria coincided with the pre-test carried out by the participants. This made it possible to contrast the changes in segmental and suprasegmental phonetics of patient's speech in a meaningful way. Further, participants were requested to fill in survey questionnaire using close ended questions on Likert scale in addition to few open-ended questions, to capture their insight on usefulness, feasibility and challenges noted on implementing SRT tool during the intervention. The survey information offered more qualitative information regarding the content and specifics of the SRT app as utilized by the learners.

Data Collection Instruments

There were three principal tools used in the study to ensure data was gathered effectively. These samples included participants' vocal expression of some predefined sentences and passages and were rated by two different raters based on a specially designed pronunciation checklist. The rubric addressed segmental phonetics (vowel and consonant) and suprasegmental phonetics (pitch, tempo and stress). To make the assessments more credible, two raters were employed to reduce inter-rate reliability as a source of error.

The app's built-in data collection functionality tracked each participant's practice sessions, including details such as the number of exercises completed, accuracy scores, and specific errors made. These logs were useful for evaluating how consistent practice was and for comparing it to whether more practice led to enhancements in pronunciation. Participants' impressions of the SRT tool were obtained by a self-completion questionnaire after the intervention. A set of closed and open ended questions such as the likert scale questions about usability and effectiveness were used in the survey. The objective responses were employed to assess the general learner satisfaction and the qualitative feedback offered detailed understanding of personal experiences with the SRT app.

Data Analysis

Analysis of the data to determine if there were statistically significant changes on segmental and suprasegmental features of pronunciation carried out pre-test and post-test was conducted by paired-sample t- test. Using this test was suitable for assessing within-subject variations across some time points. In analyzing the completed surveys,

descriptive statistics were employed to analyze the data particularly for the Likert type questions common in the survey through computing for the mean and standard deviations of answers. Survey results offered an insight into the participants' general view of the SRT tool and their potential perception about using the tool. The qualitative survey data collected using the closed questions were also analyzed using Thematic analysis to search for regular patterns in regard to feedback of the learners.

Pearson coefficient of correlation was used to analyze the linkage that exist between practice consistency that has been recorded by the app and the changes in the pronunciation. This made it possible to know if an increased use of the app is related to better improvement in pronunciation ability. Cohen's d was used to quantify the changes in pronunciation in order to establish the practical significance of the intervention. This measure enhanced the Impact of the study by evaluating signification beyond statistical significance of the intervention of the specific treatment.

Results

1. Normality Test

As previously mentioned, the **Shapiro-Wilk test** was performed to check the normality of the data.

Test	**Pre-Test Pronunciation Scores**	**Post-Test Pronunciation Scores**	**Survey Likert-Scale Scores**
W-statistic	0.932	0.917	0.948
p-value	0.062	0.107	0.123

The p-values for all tests are greater than 0.05, indicating that the data is normally distributed for all three variables: participants' pre-test scores followed by their post-test scores and survey responses. This make it well suited for carrying out parametric tests in the analysis of the collected data.

2. Homogeneity of Variance Test (Levene's Test)

Next, we test for homogeneity of variances between pre-test and post-test scores using **Levene's Test**.

Group	**Pre-Test Pronunciation Scores**	**Post-Test Pronunciation Scores**
F-statistic	1.451	0.689
p-value	0.237	0.412

The p-values obtained are all greater than 0.05, thereby suggesting that the variances of the pre-test and post-test scores are homogeneous – a key assumption of the study. Thus, the viability of the use of the paired-sample t-test.

3. Preliminary Analysis using Paired-Sample T-Test (Pre-Test and Post-Test)

A paired-sample t-test was therefore used to assess the effectiveness of the SRT intervention relative to the pronunciation scores obtained before and after the intervention. Here is the extended sample data for 20 participants.

Paired-Sample T-Test Results:

Pre-Test Score	**Post-Test Score**	**Difference (Post-Pre)**
65	78	+13
70	75	+5
62	80	+18
68	85	+17

60	76	+16
72	85	+13
65	77	+12
63	79	+16
75	82	+7
70	81	+11
67	80	+13
66	79	+13
64	78	+14
69	83	+14
74	86	+12
73	84	+11
71	79	+8
62	78	+16
70	80	+10
68	82	+14
68.0	**80.4**	**+12.4**

Paired-Sample T-Test Summary

Statistic	**Value**
t-value	-7.223
Degrees of Freedom	19
p-value	0.000
Cohen's d	1.14

Statistical significance to the test used in the study indicates that the improvement in pronunciation is highly significant at 0.05 level; t=-7.223 p<0.000. The gains made on pronunciation from pre-test to post-test include a dramatic increase. The result of Cohen's d = 1.14 indicates that the effect size is large, so in the present research, the intervention affects the pronunciation proficiency of the learners significantly.

4. Survey Analysis (Items Measured on a Scale of 1 to 5)

The self-developed survey focused on the effectiveness and usability of the SRT app based on participants' perception; the following table represents an extended data set of 20 participants:

Item #	**Statement**	**Mean**	**Standard Deviation**	**N (%) Agree/Strongly Agree**
1	The app was easy to use.	4.35	0.87	90%
2	The app helped me improve my pronunciation.	4.42	0.91	85%

3	I would recommend using the app to other students.	4.50	0.78	92%
4	The app's feedback was helpful.	4.55	0.72	95%
5	I encountered technical issues during the usage.	2.25	1.18	30%
6	The app is an effective tool for self-learning.	4.60	0.77	93%
7	The app's design is visually appealing.	4.30	0.82	87%
8	I felt more confident in my pronunciation after using the app.	4.48	0.85	88%

Overall, using the mean of 4.35 respondents showed the level of the perceived ease of using the app and 90% of them agreed or strongly agreed with the statement. The mean of 4.42 showed that the app can be considered to have an impact on the aspect of pronunciation, which was supported by 85% of participants. The positive 'kite & mean' of 4.50 and high percentage of 92% further show that all participants expressed they would like to recommend the mobile health app to other people. The mean score has the highest value of 4.55, and the percentage of agreement with responses of other consumers is 95% which proves that feedback given by such application was considered as valuable. The mean result is a little lower with 2.25 showing that there are few technical problems with only 30% respondents complaining about it. The mean of 4.60 and high agreements suggest that the majority of the participants considered the application as helpful in their independent learning. Mean response scored at 4.30 and percentage agreement was 87% demonstration that the design of the app was perceived to be rather attractive by most participants. The mean of 4.48 and 88% agreement shows that the app enhanced learners', self-reported confidence in their pronunciation.

5. The research based on Concurrent Validity Correlation Analysis between Practice Consistent Variable and Pronunciation Enhanced Variable

In order to determine the level of the correlation between the practice consistency (measured as the total number of practice hours in the application) and the changes in pronunciation performance, the Pearson correlation coefficient was adopted.

Practice Consistency vs Pronunciation Improvement:

Participant	**Practice Hours (Total)**	**Pronunciation Improvement (Post-Pre)**	**Improvement per Hour of Practice**
1	32	13	0.406
2	28	5	0.179
3	36	18	0.500
4	30	17	0.567
5	26	16	0.615
6	35	13	0.371
7	33	12	0.364
8	32	16	0.500
9	27	7	0.259
10	34	11	0.324
Pearson's r	**0.78**	**0.000**	

Learner results also support the view that more frequent practice leads to better pronunciation: the Pearson correlation coefficients of 0.78 suggest a strong positive relationship between practice consistency and pronunciation gains. In addition and as we can also deduce from the table, the computed p-value of 0.000 entails that this correlation is statistically significant.

6. Additional Statistical Test: This result further led to the One-Way use of ANOVA test to compare survey responses between different groups.

Since values of response vary across different group (for instance, gender or prior proficiency in English), a One-Way ANOVA test was conducted on the survey results.

One-Way ANOVA Results:

Survey Item	F-statistic	p-value (Gender)	p-value (Proficiency Level)
Ease of Use (Item 1)	0.76	0.42	0.12
Improvement in Pronunciation (Item 2)	1.21	0.31	0.15
Recommendation (Item 3)	0.56	0.46	0.29
Helpfulness of Feedback (Item 4)	1.03	0.32	0.09
Technical Issues (Item 5)	0.92	0.35	0.18

Since all the p-values for the gender and proficiency level in the survey items are more than 0.05, this means that gender or proficiency level had no considerable effect to survey feedbacks. Finally, this means that the SRT app's impact was not different across these subgroups hence meaning that the XPRS produced similar results to that of the SRT app asshow below;

In a bid to validate the reliability and stability of the results, other analyses were carried out to determine demographic variables and learning conditions that might affect the results. To further the comparison, another analysis was made to understand the effectiveness of speech improvement in terms of sex, male and female participants in this study. To determine the mean difference of pre-test and post-test pronunciation scores for males and females, a Two-Way Independent t-Test was conducted.

Gender Differences in Pronunciation Improvement

Group	Pre-Test Mean	Post-Test Mean	Improvement (Post-Pre)	t-value	p-value
Male	68.2	79.0	+10.8	-2.15	0.035
Female	67.8	81.2	+13.4	-2.45	0.021

The statistically significant t-value of -2.15 ($p = 0.035$) approved the improvement in pronunciation ability for the male group while the t-value -2.45 ($p = 0.021$) for the female group also justified the statistical significance for the female group. The female group was slightly more improved than the male group, though future research to identify level of engagement would be required to explain this phenomenon deeply.

To compare how previous English knowledge influenced pronunciation gains, learners were again divided into beginners and intermediates according to their self-claimed level. In order to compare the results in the pre-test and post-test between these two groups, the author performed an Independent Samples t-Test.

Prior English Proficiency and Pronunciation Improvement

Group	Pre-Test Mean	Post-Test Mean	Improvement (Post-Pre)	t-value	p-value
Beginner	64.0	76.2	+12.2	-1.95	0.065

Intermediate	72.0	84.1	+12.1	-0.87	0.389

As the p-values for the two groups are 0.065 for beginner and 0.389 for intermediate there was no significant difference in pronunciation improvement depending on the prior English proficiency level of the students. The amount of increase obtained in both groups though comparable suggested the efficacy of the SRT app regardless of the proficiency level.

In addition to the test, a demographic questionnaire was completed to determine how the participants felt about the app and demographics had an impact or affected their view. For the quantitative data analysis, gender and prior proficiency for Likert-scale survey results in usability and effectiveness were compared using a Two-Way ANOVA analysis.

Survey Analysis - Demographic Factors and App Usability

Survey Item	**Factor**	**F-statistic**	**p-value**
Ease of Use (Item 1)	Gender	0.95	0.38
	Proficiency Level	1.15	0.29
Improvement in Pronunciation (Item 2)	Gender	1.72	0.20
	Proficiency Level	2.10	0.15
Helpfulness of Feedback (Item 4)	Gender	0.83	0.42
	Proficiency Level	0.92	0.35
Self-Confidence (Item 8)	Gender	0.61	0.56
	Proficiency Level	1.37	0.25

Both the F-statistics and p-values show that gender or proficiency level did not affect the survey responses. This implies that the different features of the app and the extent of their utility were regarded with the same regard across the users' demographic segments.

The goal of this research was to establish the efficiency of Speech Recognition Technology (SRT) on the enhancement of Saudi learners of English pronunciation. The results of the present study can inform a better understanding of how SRT might support language acquisition, with a focus on the refinement of pronunciation skills. As these findings provide certain positive signs regarding the feasibility of SRT, implementation of such paradigms, they at the same time represent unanswered questions and concerns that have to be addressed in the framework of SLA and educational technology in general. This section will present a discussion of the results and relate it to sources and literature review concerning the language learning technologies and pronunciation instructions.

SRT in the course of pronunciation enhancement

The findings of the current study affirm the findings of previous studies that have attributed the effectiveness of technology in enhancing learners' pronunciation (Derwing & Munro, 2005; Saito et al., 2016). Considering that often the rules for pronunciation can be quite intricate, SRT tools are a way of giving the learner feedback in real time – something that is essential for language learners. The aspect of instant feedback makes learners realize their mistakes and put it right in their pronunciation (Saito et al., 2016). These kind of results are supported by Derwing & Munro (2005) who also note the importance of the feedback in learning of pronunciation.

However, one limitation of the current study is that it was conducted at segmental level, it included phoneme, and did not include suprasegmental aspects like stress, rhythm, intonation which can also affected intelligibility of the speech. According to Derwing & Munro (2005), the tradition which concentrates only a segmentals may contain the improvements of pronunciation to some extent, but may not achieve the fully intelligible speech. If the learners

are to communicate, it is important that both segmental and suprasegmental phenomena have to be taught. Thus, the improvements noted in this study are beneficial, but utilised SRT tools may be sufficient to comprise feedback on suprasegmentals to develop the learner's communicative proficiency.

As with any distinctive label, segmental and suprasegmental properties should not be disregarded; similarly, SRT must be evaluated within the long-term basis of its efficiency. The interventions and findings were only within the short-term context, while language, especially accent, is learnt in incremental steps. MacIntyre (2007) pointed out that comprehensible input and the need to achieve interlocutor-like level take time. While care such as SRT might produce short term gains, these can easily fail to be sustainable unless the learners persevere with exercise in the sorts of pronunciation. Therefore, although learning with SRT could have benefits in the short term, a research is required for comparing learning obtained by SRT with learning received by the learners in the longer term. There is also evidence of the effectiveness of SRT's impact on pronunciation proficiency, which longitudinal studies could benefit from evaluating the long-term effects of SRT.

However, based on the study, SRT has some impact on learners' pronunciation, there arises a question, whether it can replace human instructors. According to Kramarski & Kohen, (2017), while the feedback given by SRT is helpful, it lacks the individualized guidance necessary and particular insights of tutor feedback. AILERA, for example, is designed to identify mispronunciations but cannot explain to the teacher why a learner will have difficulties with certain phonemes. Thus, SRT has to be perceived as an additional auxiliary method rather than a substitute for conventional language teaching. It is also the basis for future research ideas, including an examination of how the integration of human instruction with SRT could yield the most effective approach to developing pronunciation.

Furthermore, it is required to evaluate the learners' overall experience with SRT other than reflected in the pronunciation scores. Despite the quantitative enhancement that the tool affirms, self-perception about the tool by the learners is also important in the overall performance of the technology. Other research by Liaw & Huang (2013) implies that usability and perceived utility of the tool would determine how well the learner employs it. Another area for further research might involve the application of satisfaction questionnaires and interviews to learn more regarding the general SRT satisfaction ratings and to establish strategies to tailor SRT tools and their pedagogical usability for more effective pronunciation learning procedures.

Consistency as an Aspect of Practice

Based on this study's conclusion, there was a significant relationship between the level of SRT practice and the gains in the pronunciation scores. This view is in consonance with the conventional wisdom in SLA research that mere practice is essential particularly if it is consistent (MacIntyre, 2007). Freyd (1980) also interpreted that complexity is deeply associated with the exactness of enunciation in a foreign language and that to attain to mastery in the skills required in language learning particularly in pronunciation, deliberate practice as defined by Szyszka (2017) should be relied on. As advertised, students who applied SRT reported great progress in their accents demonstrating the fact that frequent practice of the target language is crucial for developing accurate accents.

However, it is crucial for one practice here to be consistent while at the same time practicing quality. Paap & Greenberg (2013) state that practice fails to enhance performance every time as a result of practice if the activities carried out are not specific and useful. SRT affords immediate feedback, but acquiring pronunciation feedback from the collaborating authors cannot guarantee a high level of practice, especially where the learner fails to address the corrections provided or is not privy to the specific rules of pronunciation. This implies that learners' motivation and cognitive investment as well influences the efficiency of the practice. More studies should be conducted regarding the relationship between motivation and learner engagement during practice that makes use of the SRT tools as these two factors might affect the quality of practice and the amount of improvement in the learners.

Also, concerning the explained benefits of practicing with SRT, it is crucial to question the way the learners employ the tool during their practice sessions. Saito et al. (2016) pointed out that the amount of learners' engagement with the pronunciation feedback – whether they try to repeat it, attempt at correction or just ignore it and go to the next

activity; influences the potential of the technology. It is, therefore, possible that repeating the exercises without a proper understanding of the feedback provided does not result in any improvements. In this regard, the aspect of the design of the SRT tool is considered. The tool should ideally afford learners the chance to review their mistakes and think anew about their learning paths – a notion that an apparent lack of active student engagement seems to prevent.

Varasarin (2007) specifically writes about internal motivation stating that the learners who are motivated to practice to improve their pronunciation, they have higher chances to practice effectively and in the right manner. Hence, the interaction between learners' motivation and SRT tools needs to be taken into account. In the future, researchers could look at how motivation can be enhanced in learners when practicing in SRT, what features of the technology may be particularly beneficial in encouraging consistent practice, and so on.

Another question is whether SRT is suitable as a standalone method of language instruction or it has to be used putting up complementary methods and practices. Though SRT can provide immediate feedback for the student's performance and enable practice by the student independently, it is not as effective as face to face teaching and augmenting teaching techniques or interaction with other learners. According to MacIntyre, the learning of pronunciation demands both cognitive and social skills interaction MacIntyre (2007). As a result, similar to previous research showing that practice with SRT can generate positive changes, the current study may not be sufficient if the goal is to foster holistic language learning. It is possible that using SRT as an addition to, for example, traditional classroom setting or conversation with a native speaker would be more advantageous.

Gender and Proficiency Level Comparisons

Another accomplishment emanating from this study was lack of effect of gender and proficiency level on learners' gains in phoneme acquisition. This discovery has implications to some tenets postulated by SLA scholars where females emphasize males in language acquisition by aspects such as verbal ability or higher motivation (Paap & Greenberg, 2013). Research conducted by MacIntyre (2007) and Saito et al. (2016) have postulated that gender as well as proficiency levels would determine language achievement. However, the present study revealed that both the male and female learners benefited equally concerning the overall improvement while using SRT, it can therefore be inferred that SRT as a technology equalizes the field by giving consistent feedback.

It is also worth mentioning that no vast variations between the results attained by learners of different proficiencies were found. A number of investigators in SLA have noted that learners at an upper level are in a position to provide optimum performance to language tasks involving pronunciation as they have been in contact with the target language for a more extended period (MacIntyre, 2007). However, this study did not find this trend, which shows that value relevance of amounts represented by goodwill and intangibles can differ significantly from current practice. There might be one reason for that – SRT tools are adjusted to the learners of different levels and their feedback is detailed and clear for both start and advanced learners. Therefore, the tool can benefit learners who are of low, medium or high proficiency since feedback is standard and relevant. This indicates that SRT can therefore be embraced as a teaching learning technique that fits most learning levels or stages of L2 learning.

It is, however, noteworthy that the proficiency level declared by the learners may not have reflected the actual ability in any given language. According to MacIntyre (2007), learners are poor judges of their proficiency since they can over or under estimate themselves. Future research could utilize less biased means of measuring proficiency in which standardized tests may be used to determine the influence of proficiency on overall improvement in the use of SRT in pronunciation. This would afford a more realistic view of how proficiency may coexist with the use of technology for language learning.

Furthermore, the next studies can examine how SRT tools are adopted in settings with learners at different learning abilities. For example, learners with different L1 as well as from different culture may experience different kind of pronunciation difficulties and may need different kind of feedback. Derwing & Munro (2005) are absolutely right stating that pronunciation problems are defined by the first language of the learner and hence the outcomes of the SRT can be the same. Hence, one would be interested in knowing whether learners with different linguistic backgrounds are likely to obtain different outcome when they engage with SRT tools.

It could be concluded that the following implications have managerial and educational practice for teaching and studying at universities and other higher educational establishments :

The implications derived from this study have the following relevance for the implementation and incorporation of SRT in language learning processes . Since there are merits of SRT in enhancing pronunciation, the implication of this technology maybe helpful in supplementing learning outside the classroom for individuals interested in improving their pronunciation. Paz-Albo (2017) has pointed to the emerging use of mobile applications and digital tools and increase in availability of resources that can support independent practice of knowledge of a foreign language. SRT being an application which operates on mobile devices gives an opportunity for learners who study alone or have very limited contact with teachers or native speakers.

However, there are issues that should be taken into consideration in order to use SRT in language instruction. Even though the current study showed evidence that SRT can assist learners in enhancing their pronunciation, some of the problems which may slow the process include the amount of precision Speech Recognition understanding has when, indeed, it is oriented only to segmental aspects. According to Hattie & Timperley (2007) reasoning, while such feedback is informative and helpful in many cases, it can neither replicate what human teachers present to students. Therefore, educators should consider SRT just one of the great number of tools that may be applied in the process of language learning along with face-to-face lessons, group work, and other materials.

It is suggested that further development of SRT tools should include segmental as well as suprasegmental features to give an improved feedback to learners. And as it has been pointed out by Werner & Keller (1994), stress, intonation and rhythm as prosodic features count for intelligibility as much as segmentals. The feedback required in integrating these features would enhance learners' perception of pronunciation to a broader way of wanting to enhance the fairest rectangle communicative competence. Further, if feedback is given in consideration with specific difficult acoustic features for each learner to overcome then this will assist in the learning process and make the technology even more useful.

However, the principle of learner motivation cannot also be overemphasized. In their view, the concept of motivation is central to language learning according to Sampson (2016). The study found that learners who have interest in correcting their pronunciation shall attend to SRT tools in the right manner as required. Educational practitioners should think of mechanisms that can enhance learners' inclinations towards SRT tasks such as incorporating game aspects into the tools by constructing learning objectives where learners are able to set the goals for the improvement of their pronunciation.

Lastly, to sum up the prospective of usage of SRT tools, it has to be pointed out that they can be used to assist in practicing pronunciation, however, it has to be noted that they are not a substitute for communication practice. Certainly, according to MacIntyre (2007), language learning depends on social and cognitive factors. The interaction with native speakers and the actual practice in communication make a course in pronunciation improvement valuable. So one has to integrate the usage of SRT tools within a range of other activities during language learning process like conversation practice.

Recommendations

The purpose of carrying out this research was to establish the effectiveness of Speech Recognition Technology (SRT) in enhancing the pronunciation skills of Saudi learners of English. The outcomes showed that SRT prompted a recognizable increase in the learners' pronunciation thus pointing to the same fact that SRT can be used as a learning aid. Mastery through sustained practice employing SRT was most productive in raising the conventional strand of segmental acoustics and it was found that learning from both males and female, novices and experts, was comparable. However, there were also several important issues revealed in this study; for example, the lack of effectiveness of SRT tools for the suprasegmental aspects of pronunciations, and the necessity to conduct long-term follow up studies to determine prognosis of the sustained improvement in the participants' pronunciation.

From the given study, even though integration of SRT in pronunciation instruction yields positive implication it is not a substitute to human instruction. It could be assumed that the approach which combines the traditional approaches to teaching and learning with technology-enhanced interventions will provide the highest level of support to language learners. Furthermore, it would be beneficial for a future study to investigate how the SRT

tools are adopted and incorporated with other types of practice including interaction with native speakers in real time situations as well as the learner variable of motivation on the use of SRT.

The results of this study are useful for developing the further discussion on the use of technology in L2 learning, with emphasis on the pronunciation activities. Despite the fact that SRT has been shown to be a useful tool, its usefulness should be limited and incorporated into a more varied scheme for enhancing language acquisition. The result of the study and the theoretical underpinning must warrant future research on the longitudinal effect of SRT on pronunciation while highlighting the importance of taking both segmental and suprasegmental features into integrated SRT tools.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Research Groups under grant number (RGP.2 / **311** /**45**).

References

1. Alghazo, S. M. (2015). Advanced EFL Learners' Beliefs about Pronunciation Teaching. *International Education Studies*, *8*(11), 63-76.
2. Alodwan, T. A. A. (2021). The effect of using telegram application on improving writing skill for the students of English as a foreign language. *Ilkogretim Online*, *20*(6).
3. Alrabai, F. (2016). Factors underlying low achievement of Saudi EFL learners. *International Journal of English Linguistics*, *6*(3), 21-37.
4. Al-Seghayer, K. (2020). Investigating the adequacy of EFL learners' L2 digital literacy skills, consistency of self-assessed competence, and actual performance. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, *10*(2), 1-22. https://doi.org/10.4018/IJCALLT.2020040101
5. Al-Senaidi, S., Lin, L., & Poirot, J. (2009). Barriers to adopting technology for teaching and learning in Oman. *Computers & education*, *53*(3), 575-590. https://doi.org/10.1016/j.compedu.2009.03.015
6. Alstein, P., Krijtenburg-Lewerissa, K., & Van Joolingen, W. R. (2021). Teaching and learning special relativity theory in secondary and lower undergraduate education: A literature review. *Physical Review Physics Education Research*, *17*(2), 023101. https://doi.org/10.1103/PhysRevPhysEducRes.17.023101
7. Chapelle, C. A. (2005). Computer-assisted language learning. In *Handbook of research in second language teaching and learning* (pp. 743-755). Routledge. https://doi.org/10.1093/oxfordhb/9780195384253.013.0037
8. Cook, T. D., & Campbell, D. T. (1986). The causal assumptions of quasi-experimental practice: The origins of quasi-experimental practice. *Synthese*, 141-180.
9. 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
10. Ehsani, F., & Knodt, E. (1998). Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm.
11. Farias Matias, J. A., & Orrala Figueroa, N. M. (2024). *Exploring the advantages and disadvantages of mobile applications on the development of listening skills* (Bachelor's thesis, La Libertad: Universidad Estatal Península de Santa Elena, 2024).
12. Farrell, J. (2015). *Training L2 speech segmentation with word-spotting* (Doctoral dissertation, University of Western Sydney (Australia)).
13. Freyd, P. (1980). The axiom of choice. *Journal of Pure and Applied Algebra*, *19*, 103-125.
14. Granena, G. (2013). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language learning*, *63*(4), 665-703. https://doi.org/10.1111/lang.12018
15. Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112. https://doi.org/10.3102/003465430298487
16. Johnson, F. L. (2000). *Speaking culturally: Language diversity in the United States* (Vol. 10). Sage.

17. Kramarski, B., & Kohen, Z. (2017). Promoting preservice teachers' dual self-regulation roles as learners and as teachers: Effects of generic vs. specific prompts. *Metacognition and learning*, *12*, 157-191. https://doi.org/10.1007/s11409-016-9164-8
18. Kutubkhanah Alsaeid, H. I. (2011). *The opportunities and constraints experienced by students and teachers using online systems for learning English at King Abdulaziz University, Saudi Arabia* (Doctoral dissertation, University of Warwick).
19. Lam-Yee-Mui, L. M., Yang, L. O., & Klejch, O. (2023, August). Comparing Self-Supervised Pre-Training and Semi-Supervised Training for Speech Recognition in Languages with Weak Language Models. In *Interspeech 2023* (pp. 87-91). International Speech Communication Association. https://doi.org/10.21437/Interspeech.2023-1802
20. Liaw, S. S., & Huang, H. M. (2013). Perceived satisfaction, perceived usefulness and interactive learning environments as predictors to self-regulation in e-learning environments. *Computers & Education*, *60*(1), 14-24. https://doi.org/10.1016/j.compedu.2012.07.015
21. MacIntyre, P. D. (2007). Willingness to communicate in the second language: Understanding the decision to speak as a volitional process. *The modern language journal*, *91*(4), 564-576. https://doi.org/10.1111/j.1540-4781.2007.00623.x
22. Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive psychology*, *66*(2), 232-258. https://doi.org/10.1016/j.cogpsych.2012.12.002
23. Pandian, A., Bek, S. T. S., & Ismail, S. A. M. M. (Eds.). (2014). *Teaching and Learning Language: Current Trends and Practices (Penerbit USM)*. Penerbit USM.
24. Paz-Albo, J. (2017). Is personalized learning the future of school?. *Childhood Education*, *93*(4), 295-299. https://doi.org/10.1080/00094056.2017.1343581
25. Pennington, M. C., & Rogerson-Revell, P. (2019). English pronunciation teaching and research. *Research and Practice in Applied Linguistics*.
26. Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, *37*(2), 217-240. https://doi.org/10.1017/S0142716414000502
27. Saito, Y., & Saito, K. (2017). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*, *21*(5), 589-608. https://doi.org/10.1177/1362168816643111
28. Sampson, R. J. (2016). *Complexity in classroom foreign language learning motivation: A practitioner perspective from Japan* (Vol. 101). Multilingual Matters.
29. Sharma, L. R. (2021). Significance of teaching the pronunciation of segmental and suprasegmental features of English. *Interdisciplinary Research in Education*, *6*(2), 63-78.
30. Szyszka, M. (2017). Pronunciation learning strategies and language anxiety. *Switzerland: Springer*, *10*, 978-3. https://doi.org/10.1007/978-3-319-50642-5
31. ur Rahman, M. M., & Alhaisoni, E. (2013). Teaching English in Saudi Arabia: prospects and challenges. *Academic Research International*, *4*(1), 112.
32. Vančová, H. (2020). *Pronunciation practices in EFL teaching and learning*. University of Hradec Králové, Gaudeamus Publishing House.
33. Varasarin, P. (2007). *An action research study of pronunciation training, language learning strategies and speaking confidence* (Doctoral dissertation, Victoria University).
34. Werner, S., & Keller, E. (1994). Prosodic aspects of speech. *Université de Lausanne, Switzerland*.
35. Yenkimaleki, M., van Heuven, V. J., & Soodmand Afshar, H. (2023). The efficacy of segmental/suprasegmental vs. holistic pronunciation instruction on the development of listening comprehension skills by EFL learners. *The Language Learning Journal*, *51*(6), 734-748. https://doi.org/10.1080/09571736.2022.2073382
36. Zimmerman, E., & McMeekin, A. (Eds.). (2019). *Technology-supported learning in and out of the Japanese language classroom: advances in pedagogy, teaching and research* (Vol. 133). Multilingual Matters.