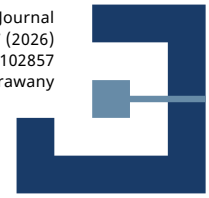
The JALT CALL Journal
ISSN 1832-4215, 22, 102857 (2026)
https://doi.org/10.29140/jaltcall.2026.102857



Castledown

The effects of automatic speech recognition (ASR)-based peer feedback on EFL diphthong pronunciation improvement and retention

Marwa Said Mustafa El-Garawany
Faculty of Archaeology and Languages, Matrouh University, Egypt
marwaal-garwani@mau.edu.eg

The present study examined the effects of automatic speech recognition (ASR)-based peer feedback on Egyptian English-majored students' EFL diphthong pronunciation improvement and retention. The participants comprised 28 first-year students in the English Language Department, Faculty of Archaeology and Languages, Matrouh University, throughout the Spring semester of the academic year 2023/2024. Participants were randomly assigned into a control group (n = 14) which received traditional pronunciation instruction and an experimental group (n = 14) which used ASR-based peer feedback. An EFL diphthong pronunciation test was devised and administered as a pre-, post-, and delayed posttest. Semi-structured interviews were also held to ascertain the students' perceptions toward the use of ASR-based peer feedback in their diphthong pronunciation training. For three weeks, students completed the assigned in-class and out-of-class activities where they practiced producing the English diphthongs using the ASR application Speechnotes with peer feedback. Results indicated that the experimental group significantly surpassed the control group on the posttest and the delayed posttest. Thus, ASR-based peer feedback had significant effects on students' diphthong pronunciation improvement and retention. The students' responses also showed positive perceptions toward its use. Such results suggest that Egyptian EFL students could improve their diphthong pronunciation through using ASR-based peer feedback.

Keywords: automatic speech recognition, ASR-based peer feedback, Speechnotes, diphthong pronunciation improvement, diphthong pronunciation retention

Introduction

With English elevated to the status of international lingua franca, correct English pronunciation is unquestionably important for EFL learners' effective and successful oral communication as it impacts intelligibility and comprehensibility. Pronunciation errors may cause serious communication breakdowns, more detrimental to the outcome than lexical or grammatical problems (Jarosz, 2019; Pennington & Rogerson-Revell, 2019). Despite such importance, pronunciation teaching is largely neglected and overshadowed by the emphasis on developing other language components such as vocabulary and grammar, resulting in frequent pronunciation inaccuracies in students' oral production (Elimat & AbuSeileek, 2014; Isaacs, 2009). Because of this neglect, EFL teachers often feel unprepared for teaching pronunciation and tend to rely on such traditional techniques like covert rehearsal, listening and imitating, minimal pair drilling of the most troublesome sounds (e.g. long vowels, diphthongs), and students' self-recording of their own performance to listen for errors. Though these techniques can assist EFL students with their pronunciation practices, they do not offer them opportunities to get feedback on their oral performance. Moreover, some students may struggle to develop their EFL pronunciation due to negative influences from their L1, and they often cannot discern their own pronunciation mistakes (McCrocklin, 2016). Therefore, providing accurate and immediate feedback may be essential for students to foster their pronunciation skills (Satio & Lyster, 2012). With advancements in artificial intelligence (AI), automatic speech recognition (ASR) has been developed to provide useful and instant feedback that informs students of their pronunciation weaknesses (Edmett et al., 2023), since it "applies various technologies such as natural language processing (NLP), natural language understanding (NLU), and speech-to-text (STT) to respectively identify, understand, and process spoken and written human language" (Nekoobahr et al., 2021, p. 165). The potential of ASR to deliver personalized feedback is also an advantage since teachers are unlikely to offer as much individualized feedback to each student because of the excessive workload and time constraints (Offerman & Olson, 2016). For these reasons, ASR has been the subject of several researchers, but its impact on pronunciation training is still unclear (Ngo et al., 2024; Yaniafari et al., 2022). The utilization of such promising technology in the Egyptian EFL context for such an objective has not yet been implemented. Obtaining empirical data on the effects of ASR in countries like Egypt, where the educational system prioritizes exam-based assessment centered on reading, grammar, and writing, is necessary, particularly given the restricted resources and access to authentic input (El-Garawany, 2021, 2024). Besides, the retention effect of using ASR on pronunciation has been barely examined in prior literature. Using a delayed posttest design to measure the long-term impacts of ASR can develop a more nuanced understanding of its effectiveness (Ngo et al., 2024). In addition, since ASR applications lack phonetic representation capabilities and visual models of articulation, students need more assistance with their pronunciation perception (Liakin et al., 2014). Peer feedback can provide such a scaffold, as few studies indicate that utilizing a combination of feedback methods, such as peer

plus ASR feedback, may lead to more pronunciation gains compared to employing a single feedback mode, like autonomous ASR feedback (Dai & Wu, 2021; Evers & Chen, 2020; Tsai, 2019; Wu et al., 2021). Furthermore, while limited investigations have explored EFL students' perceptions of using ASR-based peer feedback in their pronunciation training (Dai & Wu, 2021; Evers & Chen, 2020; Sun, 2023; Tsai, 2019; Wu et al., 2021), further studies are needed to delve deeper into students' attitudes toward it, especially regarding its impact on their diphthong pronunciation development (the focus of the present study). To plug these gaps, the present study explored the effects of ASR-based peer feedback on Egyptian English-majored students' EFL diphthong pronunciation improvement and retention along with their perceptions toward it.

Context of the problem

Egyptian English-majored students at the Faculty of Archaeology and Languages, Matrouh University, are required to study the compulsory Phonetics and Phonology course in the Spring semester of their first year. It has been noticed that they encounter difficulties in accurately producing diphthongs in their oral practices, assignments, and final exams, bringing about misrepresentation or miscomprehension of the spoken message. Diphthong pronunciation is widely documented as problematic among EFL learners (Abker, 2020; Anwar & Kalisa, 2020; Aziez & Sofiana, 2023; Pratiwi & Indrayani, 2021; Sari et al., 2024) because of the minimal and inadequate teaching and assessment practices, the limited exposure to the FL, the negative L1 interference, and the different sound and orthography systems of both languages (Fouly et al., 2020; Galal-Eldin, 2023; Khalifa, 2020).

The Arabic and English phonological systems are quite different with respect to the number of the sounds used, particularly vowels. Arabic has a smaller and simpler vowel system than English. While English has twelve monophthongs and eight diphthongs, Arabic has only six monophthongs (i.e., three short, equivalent to /æ/, /ɪ/, /ʊ/, and three long, equivalent to /ɑ:/, /i:/, /u:/) and two diphthongs equivalent to /aɪ/ and /aʊ/. Moreover, the three short vowels are of secondary importance as they are almost allophonic and rarely written in Arabic scripts – except in religious or educational texts. They are indicated by small diacritical marks placed above or below letters to change pronunciation. In contrast, the long vowels and diphthongs, which are represented by letters, are more crucial for conveying meaning, guiding pronunciation, and minimizing ambiguity (Bani Salameh & Abu-Melhim, 2014; Connor-Linton, 2006; Naji & Almakrob, 2023; Swan & Smith, 2001). Owing to this "relatively small number of vowels in Arabic, learners tend to use just one (Arabic or English) vowel to represent more than one English sound" (Kharma & Hajjaj, 1997, p. 14).

In producing diphthongs, Egyptian students tend to simplify their pronunciation by approximating them to the closest monophthongs whether short or long. Some of the most common confusions include (a) /eɪ/ and /əʊ/ are usually pronounced rather short and are confused with /e/ and /ɒ/ respectively: *red* /red/ for raid /reɪd/; *hop* /hɒp/ for *hope* /həʊp/, (b) /ɪə/ and /ʊə/ are often altered to

/i:/ and /u:/ respectively: *clear* /klɪə/ becomes /kli:/; *ensure* /ɪnʃʊə/ becomes /ɪnʃu:/, and (c) /əʊ/ is sometimes altered to /ɔ:/: *social* /səʊʃəl/ becomes /sɔ:ʃəl/. Another feature relates to vowel insertion. Egyptian students tend to insert the short vowel /ɪ/ between a diphthong and a consonant (e.g. between /aɪ/ and /d/ in *classified*, between /aʊ/ and /t/ in *about*) (Abdelreheem, 2023; Swan & Smith, 2001).

For documenting the problem, the researcher interviewed 12 first-year students at the end of the Spring semester of the academic year 2022/2023 to explore their perceptions relating to diphthongs which were thought to be the most difficult sounds for them due to their articulatory complexity. The results indicated that most respondents rated their ability to produce diphthongs as either fair or poor. They also expressed dissatisfaction with the traditional pronunciation instruction and the limited feedback they get from their lecturers. They needed new instructional and feedback practices to facilitate their pronunciation learning. Hence, this study was executed to help students to develop their EFL diphthong pronunciation through utilizing ASR-based peer feedback. It posed the following research questions:

RQ1: To what extent can ASR-based peer feedback affect Egyptian English-majored students' EFL diphthong pronunciation improvement and retention?

RQ2: What are the Egyptian English-majored students' perceptions of the usefulness of ASR-based peer feedback in their diphthong pronunciation training?

Literature review

Diphthong pronunciation

Originated from the Greek *díphthongos*, meaning 'two sounds', the term diphthong refers to a vowel in which there is a noticeable change/glide in vowel quality within a single syllable, as in *day* /deɪ/, *home* /həʊm/, and *noise* /nɔɪz/ (Brown & Miller, 2013; Crystal, 2008; Richards & Schmidt, 2010). Diphthongs are usually regarded as single phonemes but represented with two phonetic symbols (Brown & Miller, 2013). To illustrate, English diphthongs are created by combining two vowel sounds where the vocal organs (i.e., the tongue and the lips) move from one vocalic position to another. For example, when producing the diphthong /aɪ/, the tongue moves from the low position of /æ/ towards the high front position of /ɪ/ as in *eye* /aɪ/, *pie* /paɪ/, and *sigh* /saɪ/. Alternatively, for producing the diphthong /aʊ/, the movement is from low towards high back, combining /æ/ and /ʊ/ as in *how* /haʊ/, *now* /naʊ/, and *cow* /kaʊ/. In certain descriptions, the movement is characterized by involving a glide such as /j/ or /w/, leading to the diphthongs represented as /aɪ/ and /aʊ/ occasionally being seen as /aj/ or /aw/ (Yule, 2014). Based on such direction of tongue movement, diphthongs are classified into closing diphthongs (i.e., the tongue rises, thereby closing the space between it and the roof of the mouth) and centring diphthongs (i.e., the tongue moves towards the central vowel sound /ə/ for producing /ɪə/, /eə/, and /ʊə/). The closing diphthongs are further subdivided into fronting (i.e., the tongue moves towards the close front vowel /ɪ/ for producing

/eɪ/, /aɪ/, and /ɔɪ/) and backing (i.e., the tongue moves towards the close back vowel sound /ʊ/ for producing /əʊ/ and /aʊ/). Regarding change in lip shape, two indicators of roundedness are employed: (a) from unrounded to rounded, e.g. /əʊ/ as in *load* and /aʊ/ as in *loud*, and (b) from rounded to unrounded, e.g. /ʊə/ as in *sure* and /ɔɪ/ as in *join*. However, the diphthongs /eɪ/, /aɪ/, /ɪə/, and /eə/ are lip spread throughout their articulation (Collins et al., 2019; O'Connor, 1980).

Diphthong acquisition and articulation can therefore be challenging for EFL students, particularly for those whose L1 (i.e., Arabic) does not involve similar sounds (Naji & Almakrob, 2023). This necessitates the adoption of effective instructional methods and feedback practices to tackle the recurring problems that students may encounter when developing their EFL diphthong pronunciation (Hewings, 2004). To the researcher's knowledge, limited empirical efforts have been made to improve it, but some instructional practices leveraging technology have shown promise. For instance, utilizing online dictionaries that include audio pronunciations positively impacted students' diphthong pronunciation (Chairunnisa, 2023; Krismonika et al., 2021). Employing the flipped classroom model significantly fostered it through using PowerPoint files and YouTube videos (Alzahrani & Alqurashi, 2023) and listening to audio-recordings of assigned texts containing the target sounds (Özdemir, 2019). Moreover, computer-assisted pronunciation training (CAPT) systems have been used to facilitate students' diphthong pronunciation proficiency as they offer personalized feedback and substantial amounts of multifaceted digitized pronunciation practice. These systems allow for recording and displaying students' speech signals as visual representations (i.e., pronunciation models, articulatory movements, waveforms, and spectrograms) (Evers & Chen, 2020). For example, using programs like Accent Master Software (Saleh & Gilakjani, 2021), icSpeech Pronunciation Coach (Ghounane & Rabahi, 2021), and Praat (Behr, 2022; Nurjannah et al., 2023) significantly improved students' diphthong pronunciation skills and raised their awareness of their speech errors.

ASR-based peer feedback in pronunciation training

Despite their established efficacy, certain CAPT applications may be technically difficult for students to operate or interpret their visual graphs (El-Garawany, 2021; Levis, 2007; Li, 2019; Neri et al., 2008; Wong, 2016) and might restrict their education to pre-determined activities or pre-recorded utterances from a native speaker for comparison (McCrocklin, 2016; McCrocklin et al., 2019; Neri et al., 2008). One type of CAPT technology that enables students to practice pronunciation easily and freely is ASR (Evers & Chen, 2020), which refers to "the computational transformation of spoken language to machine readable text" (Brown & Miller, 2013, p. 43) as it "takes the acoustic speech signal and converts it into a text transcription" (Cooper, 2019, p. 1788). Likewise, Wilson (2024, p. 35) defines it as a technology that can recognize and transcribe oral language into text as it "analyzes audio signals, such as speech recordings or real-time audio streams, to identify spoken words and phrases and convert them into written texts". ASR systems, therefore, consist of two basic aspects: a front-end and a

back-end. The front-end aspect processes the audio input and captures fundamental attributes of speech while excluding irrelevant information through filtering, noise reduction, and feature extraction using mel-frequency cepstral coefficients (MFCCs). The back-end aspect decodes the extracted features and generates a sequence of words that most accurately matches the spoken input. This entails employing various algorithmic approaches such as hidden Markov models (HMMs), deep neural networks (DNNs), and end-to-end models to map the acoustic features to meaningful sequences of sounds, words, and sentences (Lin, 2025; Kumar & Yadav, 2024). ASR systems process diverse applications, including virtual assistants (e.g. Apple's Siri, Amazon's Alexa, Google Assistant) that respond to voice instructions, real-time video captioning services that enhance content accessibility for the hearing impaired, and dictation software products (e.g. Apple Dictation, Google Speech Recognition, Windows Speech Recognition) that transform voice recordings, interviews, and lectures into written texts, eliminating the need for manual notetaking and facilitating the documentation of work meetings (Lin, 2025; Kumar & Yadav, 2024; Wilson, 2024).

Notwithstanding such initial purpose, utilizing ASR dictation programs for language learning has garnered increasing interest among scholars in recent years (Liakin et al., 2017). These programs often do not need purchase, enabling the accumulation of large-scale speech databases that substantially enhance recognition accuracy. Moreover, the accessibility of these programs facilitates their integration into classroom settings or self-study practices (Evers & Chen, 2020). Currently, there exist three categories of ASR systems: speaker-dependent; -independent; and -adaptive. Speaker-dependent ASR requires preliminary training on data from the user, enabling the program to accurately recognize individual speech characteristics. In contrast, speaker-independent ASR operates without any prior training, relying instead upon a comprehensive database of speech samples. Speaker-adaptive ASR employs an extensive speech database alongside specialized adaptation algorithms tailored for each user (Junqua & van Noord, 2001; Moore, 2021). For pronunciation training, ASR dictation platforms need to be speaker-independent to remain unaffected by the students' accents or mispronunciations, ensuring unbiased feedback and improvement opportunities. If a student has an accent and mispronounces certain phonemes, speaker-dependent or -adaptive ASR will learn to disregard these errors. For instance, if a student consistently articulates the phoneme /v/ more like /f/, the software will analyze the context and adapt to his/her pronunciation, providing transcriptions that include /v/, even when /f/ is pronounced (e.g. when receiving /fɑ:st/, the ASR system will transcribe it as *vast*). With time, the program will yield transcriptions increasingly aligned with the student intended speech rather than his/her actual pronunciation. Therefore, feedback from speaker-dependent or - adaptive ASR cannot be deemed objective (Evers & Chen, 2020). Accordingly, the speaker-independent ASR software was selected for employment in the present study.

These ASR programs also require high output and error detection accuracy since inaccurate transcription can lead to frustration and demotivation (Liakin

et al., 2017; McCrocklin et al., 2019; Mroz, 2018). Several researchers have compared the recognition rates of native and non-native speakers. Derwing et al. (2000) utilized Dragon NaturallySpeaking as an ASR tool and concluded that it accurately recognized and transcribed about 72% of words articulated by non-native speakers, showing that the program's recognition rate was not as accurate as human listeners. Besides, the system's accuracy in identifying native speakers was about 18% higher compared to non-native speakers at an advanced proficiency level. Likewise, Ashwell and Elam (2017) found that the transcription success rate of Google Web Speech API for Japanese learners of English was 66%, whereas that of native speakers was almost 23% higher. McCrocklin et al. (2019) found that the accuracy of Google's engine ranges from 88% for sentence dictation to 93% for free speech in the case of non-native speakers, and from 92% for sentence dictation to 98% for free speech in the case of native speakers, which is comparable to human-based accuracy reported to be between 89% and 98% (Edalatishams, 2017). However, Windows Speech Recognition correctly decoded 74% of sentence transcription and 54% of free speech of non-native speakers and 87% of sentence transcription and 60% of free speech of native speakers (McCrocklin et al., 2019). The average recognition accuracy of Siri was 67% for spoken sentences by Japanese EFL students (Daniels & Iwago, 2017). These studies indicate that the recognition rate of ASR systems for non-native speech remains relatively low, possibly due to the speakers' accents or pronunciation errors. Alternatively, it could be posited that the error detection accuracy for non-native speakers is higher, which may be useful for authentic pronunciation training (Xiao et al., 2023). In addition, Google Speech Recognition provides higher accuracy rates compared to other widely used free ASR dictation programs, making it, along with other applications built on it, the preferred choice for pronunciation practice (Daniels & Iwago, 2017, McCrocklin et al., 2019). Thus, despite not matching the error detection rate of humans, ASR with a sufficient accuracy rate can provide meaningful feedback to language learners under controlled teaching conditions, enhancing their awareness of refining their English pronunciation (Xiao et al., 2023). Current research has documented positive impacts of individual practice with ASR on students' EFL pronunciation improvement (Aljabr, 2025; McCrocklin, 2016, 2019; Xiao et al., 2023; Yuan & Liu, 2020; Ziane & Dou, 2024).

To avoid encountering ASR limitations, pronunciation instruction should facilitate opportunities for students to exchange their perspectives on pronunciation. Such exchanges may provide insights that are more accurate than the generated ASR feedback (Evers & Chen, 2020). Moreover, ASR dictation applications lack functions that offer detailed corrections and contextual explanations. Thus, students need more support to understand their pronunciation (Liakin et al., 2014). Incorporating peer feedback alongside ASR can significantly enhance pronunciation instruction (Dai & Wu, 2021; Evers & Chen, 2020; Sun, 2023; Tsai, 2019; Wu et al., 2021). This suggestion shares theoretical linkages with the Vygotskian sociocultural theory in that cognitive development is a mediated process shaped by social interactions and cultural tools, including digital educational resources (Vygotsky, 1978). Incorporating ASR with peer feedback

provides students with valuable feedback and abundant practice opportunities to speak and evaluate their technology-mediated output, enabling them to develop their pronunciation knowledge in an interactive and stimulating learning environment. The integration of peer feedback further amplifies the social aspect of the learning process, allowing students to interact with each other and collaborate to detect their pronunciation flaws and refine their pronunciation skills (Dai & Wu, 2021; Sun, 2023). Moreover, ASR-based peer feedback is situated within two other theoretical frameworks, namely the noticing hypothesis which emphasizes the importance of consciously recognizing language features (Schmidt, 1990) and the interaction hypothesis which highlights the role of meaningful communication in language development (Long, 1991). These theories relate to ASR-based peer feedback as they elucidate how students process and notice gaps in their EFL pronunciation through the automated and peer feedback they receive (Sun, 2023).

Empirical studies on the effectiveness of ASR-based peer feedback

To date, a few studies have investigated the pedagogical effects of ASR-based peer feedback on EFL pronunciation skills, particularly among Chinese-speaking students. In the context of Taiwan, Tsai (2019) carried out a qualitative study to examine the influence of collaborative practice using the ASR-based system MyET versus self-directed practice on 60 English majors' pronunciation acquisition of segmentals and suprasegmentals. After 10 weeks, the analysis of students' reflective reports revealed that the collaborative group recorded largely positive remarks regarding the advantages of the ASR software as they enjoyed using its functions and became more aware of their common pronunciation errors. Evers and Chen (2020) explored the effects of ASR-based peer feedback on 63 Taiwanese working adults' pronunciation performance for 12 weeks. During the weekly two-hour sessions, the participants dictated a given text to the ASR software Speechnotes and then practiced mispronounced words by themselves (the control group, $n = 31$) or with feedback from group members (the experimental group, $n = 33$). The texts targeted seven English problematic phonemes, five without analogues in their L1 (/v/, /θ/, /ð/, /æ/, and /e/) and two with analogues in their L1 (/r/ and /ʃ/). Results of the reading-aloud and spontaneous conversation tasks indicated that peer feedback was significantly more effective than independent practice in fostering students' pronunciation comprehensibility. Besides, the experimental group was more satisfied with the ASR software than the control group. Though, no significant differences were detected in accentedness or perceived ease of use.

In the context of China, Dai and Wu (2021) investigated the effect of feedback provided by peers and/or ASR in mobile-assisted pronunciation learning, targeting 10 troublesome English phonemes; five vowels (/i:/, /ʌ/, /eɪ/, /ɑ:/, and /ə/), three consonants (/v/, /θ/, and /ð/), and two consonant clusters (/tr/ and /str/). A total of 84 university students were allotted to three conditions using WeChat: peer feedback, ASR feedback, and peer plus ASR feedback. A three-week procedure was applied: the pretest in Week 1, the treatment plus the

immediate posttest in Week 2, and the delayed posttest in Week 3. The results of the eight-sentence-reading task showed that the three groups improved their pronunciation, but the peer feedback and the peer plus ASR feedback groups significantly outperformed the ASR feedback group on the posttest and the delayed posttest. The interview data showed that the students who utilized ASR appreciated its immediate feedback but encountered some access problems. Wu et al. (2021) explored the effectiveness of ASR-based peer feedback in reducing 60 Chinese first-year secondary school students' sound deletion during an eight-week experiment. Each weekend, the participants utilized DingTalk to practice reading their textbook passages. They assessed the errors in the ASR by themselves (the control group, n = 30) or with feedback from partners (the experimental group, n = 30). The results of the post-test revealed that the peer feedback group performed better than the self-corrected group in overall pronunciation performance and sound deletion. Moreover, both groups held positive attitudes toward the ASR-based pronunciation practice. Sun (2023) studied the impact of a 14-week ASR-based peer feedback treatment on improving 61 Chinese intermediate-level students' pronunciation and speaking skills. The participants were randomly divided into a control group (n = 32) which received conventional teacher-led instruction and feedback and an experimental group (n = 29) which used Speechnotes with peer correction. The results of read-aloud tasks, spontaneous conversations, and IELTS speaking tests indicated that the experimental group significantly outclassed the control group in pronunciation comprehensibility and accentedness and global speaking abilities. Responding to the interviews, the students recognized Speechnotes as a beneficial ASR platform for developing their pronunciation and speaking skills. The present study aimed to extend that previous research to the Egyptian context, assessing both the immediate improvements and the enduring effects of ASR-based peer feedback on EFL diphthong pronunciation, while also examining students' perceptions toward its utility, mainly in their diphthong pronunciation training.

Method

Design and participants

In this study, the mixed-methods approach was adopted using the pretest-posttest control-group experimental research design. The mixed-methods approach was chosen to enable a thorough understanding of the influence of ASR-based peer feedback on EFL diphthong pronunciation, combining quantitative data to validate the impact of the study intervention and qualitative insights to explore students' attitudes, challenges, and perceived benefits. The pretest-posttest control-group design was employed to measure the efficacy of ASR-based peer feedback in enhancing diphthong pronunciation, facilitating direct comparisons between the experimental and control groups while accounting for initial pronunciation differences, as the pretest establishes a baseline and the posttest assesses changes post-intervention. The participants

(*N* = 28, aged 18–19; six males and 22 females) were first-year students in the English Language Department, Faculty of Archaeology and Languages, Matrouh University, throughout the Spring semester of the academic year 2023/2024. Based on the participants' records, they were identified as intermediate learners of English, as indicated by their scores on the English language entrance exam. They were native speakers of Arabic (i.e., Egypt's official language) who had studied EFL for roughly 12 years and received the same pre-tertiary education before entering university. They were also digitally literate but had no prior experience with ASR dictation programs before the experiment. They were randomly assigned into a control group (*n* = 14) which received traditional pronunciation instruction and an experimental group (*n* = 14) which used ASR-based peer feedback. Before the study commenced, the participants were informed about the purpose, instruments, and procedures of the study and signed the consent. The experimental group students had the freedom to withdraw at any time without penalty. However, they were required to submit regular assignments upon withdrawal, as the completion of these assignments is integral to the course grade according to the faculty regulations. The confidentiality and anonymity of the participants' responses were assured. All students willingly participated and allowed data collection and analysis. After the conclusion of the study, the control group employed ASR-based peer feedback in their diphthong pronunciation training to guarantee equal educational opportunities.

EFL diphthong pronunciation test (DPT)

For collecting quantitative data, the researcher devised an EFL diphthong pronunciation test to be used as a pre-, post-, and delayed posttest (see Appendix A). It contained three questions to measure the students' ability to correctly produce diphthongs in words, sentences, and paragraphs. The students were individually tested and audio recorded within a time limit of 10 minutes. During recording, they were permitted to repeat the words and/or phrases if they believed their pronunciation was incorrect, with just the final pronunciation included in the analysis. This approach mitigated the impact of accidental errors or anxiety. Before responding to the questions, each student was given five minutes to familiarize himself/herself with the test items. In Question 1, the students were asked to read aloud eight lists of words, covering the eight English diphthongs. In each list, there were five different target diphthongs with five examples of each target diphthong in total. The words were divided into sets of five to make it easier for students to process them, breaking up what would otherwise be a long list of 40 words. In Question 2, the students were asked to read aloud 10 sentences, containing three words for each diphthong. In Question 3, the students were required to read aloud a short paragraph, including sufficient examples of the eight English diphthongs, different from those in the lists or sentences. Using such read-aloud tasks, rather than spontaneous speaking tasks, was for two reasons. First, the reading aloud task, a type of controlled pronunciation production, enhances comparability and increases

score reliability. Second, employing an assessment of spontaneous speech does not guarantee that the specific pronunciation features (i.e., words containing diphthongs) to be tested will manifest in students' responses, or at least in sufficient quantity to render the evaluation valid (Hewings, 2004; Richards, 2015). Materials for constructing the test were adapted from O'Connor (1980), Hewings (2004), and Baker (2006).

As proposed to ensure objectivity, two scorers (the researcher and another lecturer of TEFL of the same teaching experience and qualifications) listened to the audio recordings, assessed the students' EFL diphthong pronunciation, and calculated the mean. Question 1 was scored by awarding one mark for each correct response/word, resulting in a total possible score of 40. Questions 2 and 3 were assessed by using the developed rubric which contained two skills suitable to the purpose of the present study: accuracy and comprehensibility. Each skill was measured on a five-point rating scale (1 = *very poor*, 2 = *poor*, 3 = *fair*, 4 = *good*, and 5 = *very good*), leading to a total score ranging from two to 10 for each question. Thus, the DPT was scored out of 60. As proposed to achieve its content validity, an expert jury of seven Egyptian university professors of linguistics and applied linguistics were requested to evaluate it according to three criteria: appropriateness to the students' academic level, clarity of its instructions and items, and pertinence to the objective it intended to measure. They indicated that the test could be an accurate measure of EFL diphthong pronunciation. To compute its reliability, the DPT was piloted on a group of 12 first-year English-majored students- out of the study sample- at the end of the Spring semester of the academic year 2022/2023. Since two scorers marked it, the inter-scorer reliability was performed where the Pearson's correlation coefficient calculated was 0.907, thereby signifying high test reliability.

Semi-structured interview

For obtaining qualitative data, the semi-structured interview was conducted to gather more detailed information regarding the perceived usefulness of ASR-based peer feedback. Therefore, the researcher developed an interview guide (see Appendix B), including four open-ended questions to ascertain the experimental group students' perceptions about the use of ASR-based peer feedback in improving their EFL diphthong pronunciation and the specific benefits they gained from using it. To confirm its content validity, it was submitted to the same jury who validated the DPT. As the researcher requested voluntary participation, 10 students consented to be interviewed. They were afforded a comfortable environment to candidly articulate their views and experiences, and they were allotted sufficient time to respond to each question. The researcher attentively listened to the ideas expressed by the students and attempted to refrain from steering the interviews in any particular direction. The interviews lasted from 15–20 minutes and were audio-recorded for transcription. Following the steps outlined by Johnson and Christensen (2014), the obtained data was carefully read, systematically segmented, and repeatedly scrutinized to assign broad categories (codes), which were further studied to

reveal subthemes. This iterative process ensured an in-depth analysis of the qualitative data, resulting in accurate identification and interpretation of principal themes. As proposed to increase the reliability of the coding procedure, the inter-coder reliability was used, which involved two coders (the same scorers of the DPT) coding the data, defining themes, comparing their findings, and resolving any differences to reach consensus, thereby maintaining objectivity and trustworthiness.

ASR dictation application speechnotes

Speechnotes (https://speechnotes.co/) is a speaker-independent ASR dictation application that uses connected word recognition, allowing students to articulate their utterances without necessitating pauses between words for better recognition. Students should momentarily pause only after a sequence of interconnected words that the program then transcribes together. This method enhances accurate identification of homonyms (i.e., words that are phonetically same yet possess distinct meanings such as *bear* /beə/ and *bare* /beə/ and no /nəʊ/ and *know* /nəʊ/) and does not disrupt fluency (Evers & Chen, 2020). To deliver accurate transcription, Speechnotes employs Google's high-end speech-recognition engines. Students' speech is transmitted to Google where it is processed by advanced parallel servers and algorithms and subsequently returned to Speechnotes as a stream of potential transcription outcomes. With the right handling of these outcomes and a set of commands to the STT engines, students attain results that maintain high accuracy, rivalling even the most sophisticated and costly software available on the market. Transcription accuracy is anticipated to exceed 90% for native speakers (Štuikys & Burbaitė, 2024) and 88% for non-native speakers, as previously mentioned (McCrocklin et al., 2019). On the application webpage (also available on phones/tablets), there is a clear explanation of its functionalities, including how to dictate, proofread, and punctuate (Yaniafari et al., 2022).

Experimental procedures

The experiment of this study was carried out as a part of the Phonetics and Phonology course offered to first-year English-majored students who met with the researcher for four hours of face-to-face pronunciation instruction each week. Before experimentation, under the same circumstances, both groups were given the pretest on 16th March 2024. Owing to the small sample size, the students' scores on the pre-administration of the DPT were analyzed using the nonparametric Mann-Whitney U Test to determine whether there were any statistically significant differences between the control and experimental groups' mean ranks before the experiment. Results revealed that the two groups had similar mean values with no statistically significant differences between the two groups' mean ranks in each pronunciation level and in overall EFL diphthong pronunciation ($U = 92.5$, $p > 0.01$), suggesting that both groups were comparable in their initial diphthong pronunciation as they entered the

study (see Table 1). The performance of both groups in word-, sentence-, and paragraph-level pronunciation was considered average. Concerning word-level pronunciation, the students managed to accurately articulate 61% of the words. Regarding sentence- and paragraph-level pronunciation, the students demonstrated a basic understanding of diphthong pronunciation as they pronounced some of the words containing diphthongs properly, with several conspicuous faults that rendered their pronunciation occasionally difficult to understand. It was observed and concluded (by the two scorers) that the most prevalent errors pertained to /əʊ/, /ɪə/, /eə/, and /ʊə/, followed by /eɪ/ and /ɔɪ/, representing the six diphthongs without analogues in the students' L1. The least frequent errors were associated with /aɪ/ and /aʊ/, representing the two diphthongs with analogues in their L1. This also might be due to misusing the place of articulation and misunderstanding the rules of vowel pronunciation. Based on these results, an intervention was deemed necessary to improve the students' EFL diphthong pronunciation.

Table 1. U-scores for the differences between the control and experimental groups' mean ranks of the EFL diphthong pronunciation on the pretest

Level/question	Group	N	M	SD	Mean rank	Sum of ranks	U	Sig.
1. Word-level pronunciation	Control	14	24.36	1.610	14.5	406	92.5	No sig. > 0.01
	Exp.	14	24.54	1.834				
2. Sentence-level pronunciation	Control	14	5.07	0.675	14.5	406	86.5	No sig. > 0.01
	Exp.	14	5.21	0.671				
3. Paragraph-level pronunciation	Control	14	5.50	0.620	14.5	406	86.5	No sig. > 0.01
	Exp.	14	5.36	0.633				
EFL diphthong pronunciation (total)	Control	14	34.93	2.702	14.5	406	92.5	No sig. > 0.01
	Exp.	14	35.04	2.919				

Experimentation lasted for three weeks where the students studied /eɪ/, /aɪ/, and /ɔɪ/ in the first week, /aʊ/ and /əʊ/ in the second, and /ɪə/, /eə/, and /ʊə/ in the third. Each week, the class period was divided into two two-hour sessions. During the first session, the researcher taught the target sounds using regular lecturing and phonetic training followed by such techniques like listening and imitating- with the help of the researcher and audio recordings- and minimal pair drilling, starting from word-level to sentence-level pronunciation. During the second session, the students practiced the target sounds in paragraph-level pronunciation. For this part, they were required to read aloud eight selected short scripts (i.e., paragraphs/dialogues, usually containing 10–15 sentences/lines); each targeted one diphthong in about 15 words. Thus, they dictated three scripts in Week 1, two in Week 2, and three in Week 3 to the ASR software. During Week 1, an extra 30-minute orientation session was given to familiarize the students with the Speechnotes application and its capabilities and to train them on how to interpret the ASR feedback. Then, students were

randomly divided into pairs. In each pair, Student A dictated the assigned script to Speechnotes and got real-time transcriptions of his/her utterances. Student B compared the generated transcriptions with the original ones, wrote down/marked the mispronounced words or phrases, and helped the practicing student to correct his/her mistakes who kept practicing till the transcripts matched the model texts. After that, the feedback giver took on the role of the practicing student where he/she received feedback from his/her classmate. This process was repeated until students practiced all the assigned texts. During these in-class activities, students also engaged in pronunciation-based discussions to rate each other's pronunciation performance using the developed scoring rubric (see Appendix A) and suggest strategies to improve it. The researcher circulated among the students to make sure the pairs were staying on task and performed the assigned activities properly. The students with their partners were also required to perform out-of-class activities by reading aloud eight more short scripts to continue practicing diphthongs following the same procedures. On their WhatsApp group, the students were required to include their recordings, some captured screenshots of their Speechnotes results (i.e., the first and final trails), and examples of the offered peer feedback, providing tangible evidence of their progress in diphthong pronunciation as their final ASR-generated transcripts matched the provided model scripts. This was to make sure that they incorporated the ASR application and peer correction in their pronunciation training each week. Figure 1 presents some samples of the students' work on Speechnotes. The assigned materials were taken from O'Connor (1980), O'Connor and Fletcher (1989), Baker (2006), and Baker and Goldstein (2008). Regarding the control group students, they received the same face-to-face pronunciation instruction. Though, in the second in-class session, each student read aloud the same given scripts before the class, focusing on the target sounds. The researcher directly provided feedback by modeling correct pronunciation for their mistakes and giving explicit explanations and suggestions for improvement. For the out-of-class activities, they were also asked to record themselves reading aloud the same extra given materials and upload their recordings to their WhatsApp group. The researcher listened carefully to these recordings, analyzed the students' pronunciation performance, and provided delayed corrective feedback (i.e., written comments and voice notes) on their mistakes.

At the end of the three-week experiment, both groups took the posttest on 7th April 2024. After two weeks, they took the delayed posttest on 21st April 2024. On the following day, the researcher held semi-structured interviews with the 10 volunteer students. Utilizing the nonparametric Mann-Whitney U Test, the students' scores on the two post-administrations of the DPT were analyzed. It was employed to compare the differences between the control and experimental groups' mean ranks on the two post-administrations to examine the effects of ASR-based peer feedback on EFL diphthong pronunciation improvement and retention. Moreover, textual data from the interviews were subjected to thematic analysis to define themes related to the use of ASR-based peer feedback in improving the students' diphthong pronunciation.

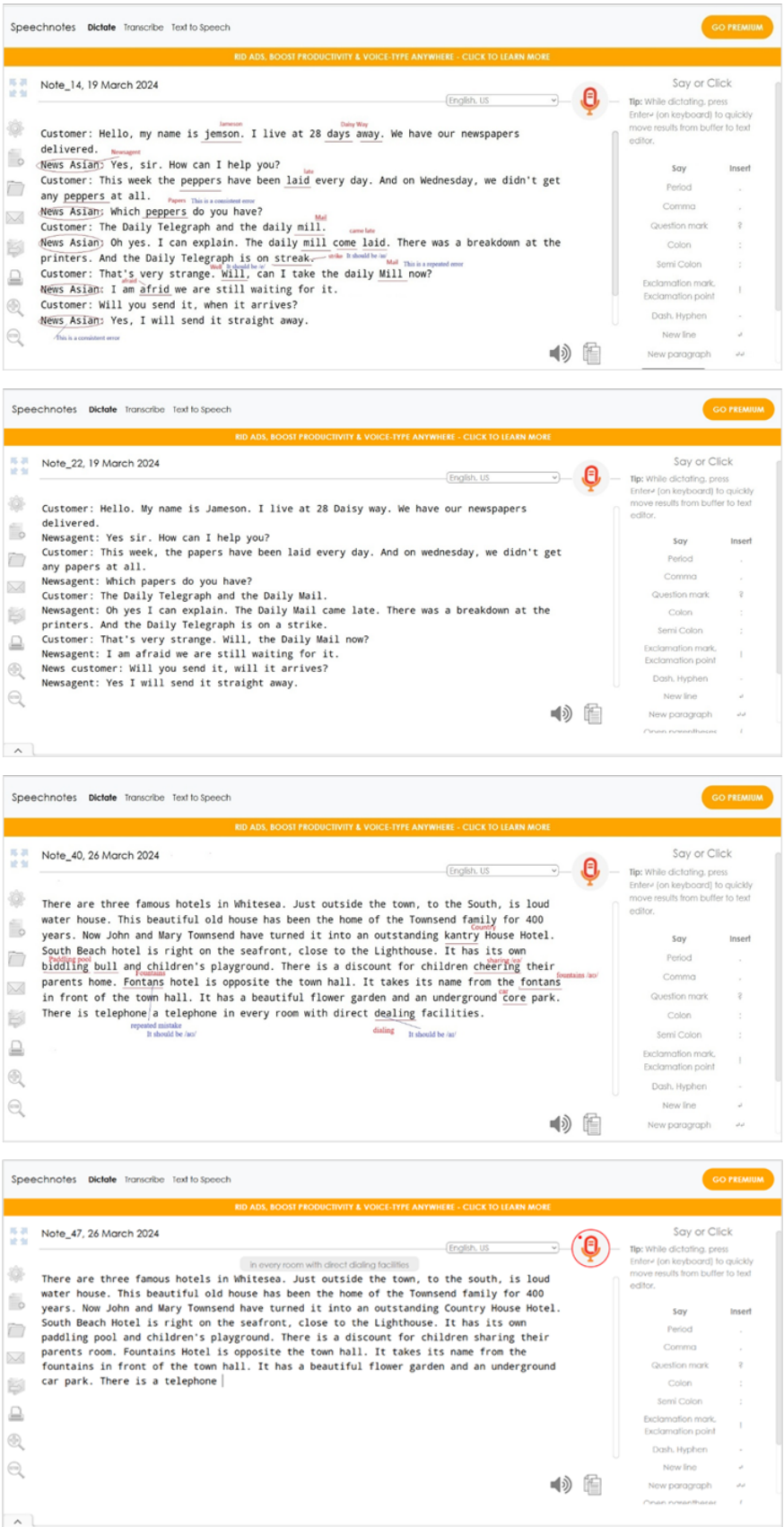


Figure 1. Samples of the students' first and final trails on Speechnotes

Results

RQ1: To what extent can ASR-based peer feedback affect Egyptian English-majored students' EFL diphthong pronunciation improvement and retention?

Table 2 provides the mean values and U-scores for the differences between the control and experimental groups' mean ranks of the EFL diphthong pronunciation on the posttest and the delayed posttest. The mean values of the experimental group were greater than those of the control group, thus revealing the positive effects of ASR-based peer feedback. Table 2 also displays that there existed statistically significant differences at the 0.01 level between the control and experimental groups' mean ranks in each pronunciation level and in overall EFL diphthong pronunciation on the posttest ($U = 33.5$, $p < 0.01$) and the delayed posttest ($U = 33.5$, $p < 0.01$) in favor of the experimental group. This indicates that the experimental group students significantly surpassed their control group counterparts in EFL diphthong pronunciation improvement and retention. This progress might be attributed to the integration of ASR-based peer feedback into the students' diphthong pronunciation training. Furthermore, the effect size was estimated using Cohen's d to determine the magnitude of the mean differences between the control and experimental groups in each pronunciation level and in overall EFL diphthong pronunciation. The adopted cut-off points were 0.2, 0.5, and 0.8, denoting small, medium, and large effects, respectively. Table 2 indicates that the mean values of both groups in each level and in overall diphthong pronunciation on the posttest ($d = 1.391$) and the delayed posttest ($d = 1.425$) were very different, as shown by the large effect sizes. This might be ascribed to the employment of ASR-based peer feedback.

Table 2. U-scores for the differences between the control and experimental groups' mean ranks of the EFL diphthong pronunciation on the posttest and the delayed posttest

	Level/question	Group	N	M	SD	Mean rank	Sum of ranks	U	Sig.	Effect size
The Posttest	1. Word-level pronunciation	Control	14	27.96	1.704	14.5	406	34	0.01	1.371 Large
		Exp.	14	30.46	1.936					
	2. Sentence-level pronunciation	Control	14	5.89	0.712	14.5	406	36.5	0.01	1.214 Large
		Exp.	14	6.75	0.700					
	3. Paragraph-level pronunciation	Control	14	6.18	0.608	14.5	406	35	0.01	1.347 Large
		Exp.	14	7.04	0.664					
	EFL diphthong pronunciation (total)	Control	14	40.04	2.899	14.5	406	33.5	0.01	1.391 Large
		Exp.	14	44.25	3.155					
The Delayed Posttest	1. Word-level pronunciation	Control	14	27.43	1.959	14.5	406	31.5	0.01	1.349 Large
	2. Sentence-level pronunciation	Control	14	5.68	0.749	14.5	406	37.5	0.01	1.252 Large
		Exp.	14	6.57	0.675					
	3. Paragraph-level pronunciation	Control	14	6.00	0.620	14.5	406	31.5	0.01	1.433 Large
		Exp.	14	6.89	0.626					
	EFL diphthong pronunciation (total)	Control	14	39.11	3.151	14.5	406	33.5	0.01	1.425 Large
		Exp.	14	43.50	3.013					

RQ2: What are the Egyptian English-majored students' perceptions of the usefulness of ASR-based peer feedback in their diphthong pronunciation training?

Five positive themes emerged from the thematic analysis of the data obtained from the semi-structured interviews. To tangibly demonstrate the defined themes, some representative quotations from the participants were selected. For clarity and reference, these quotations were marked with participant identifiers as P1, P2 etc.

Theme 1: Perceived usefulness of ASR-based peer feedback. The ten participants believed that using ASR-based peer feedback was effective in improving their EFL diphthong pronunciation. Notably, they highlighted the recognition accuracy of the ASR dictation tool Speechnotes in detecting their pronunciation mistakes and the help of their peers in interpreting the ASR feedback and correcting those mistakes. For instance, P1 noted that "I found Speechnotes beneficial in identifying my pronunciation faults. With my classmate, I kept practicing and dictating my texts to the ASR dictation tool until my errors decreased and my transcripts improved as they matched the original scripts". P2 stated that "The ASR generated transcripts plus the peer correction helped me to improve my pronunciation through getting more accurate feedback than depending on my own judgement or the teacher's feedback". Besides, some participants were

initially skeptical, but experienced positive outcomes after employing the tool with their peers, as observed by P5 who reported that "Using ASR-based peer feedback was useful in detecting and modifying my pronunciation mistakes. At first, I had some doubts about the efficacy of the ASR tool in developing my diphthong pronunciation, but with the assistance of my partner and after engaging with it several times, I observed an obvious reduction in the number of mistakes and a marked improvement in the pronunciation of diphthongs".

Theme 2: Enjoyment and motivation. Nine students indicated that using ASR-based peer feedback provided them with a more enjoyable and motivating educational experience to practice diphthong pronunciation than traditional instruction. They valued the peer correction aspect of the proposed intervention, as expressed by P3 who stated, "It made practicing diphthong pronunciation more enjoyable than just performing the regular classroom activities. When dictating the given scripts to the tool, my partner and I took turns comparing them with the generated transcripts to spot the differences, rate the overall performance in terms of accuracy and comprehensibility, and correct the mistakes. So, we kept practicing till our transcripts improved, and we were very happy to see that the final trials matched the original texts". P6 also said that "I enjoyed using the features of Speechnotes and getting involved in the process especially when we evaluated each other's performance to improve the accuracy and comprehensibility of our diphthong pronunciation. We were very satisfied with our final trials and the progress we achieved each week". Similarly, P8 stated that "The progress achieved with each trial pleased and encouraged us to continue practicing".

Theme 3: Metalinguistic awareness. Seven students revealed that the utilization of the ASR dictation tool in a cooperative manner enhanced their awareness of their mistakes, subsequently motivating them to get improved. P2 mentioned that "I was unaware of how often I mispronounced diphthongs until I began using Speechnotes. Seeing the Speechnotes feedback helped me to pay attention to the common mistakes I make in my pronunciation and concentrate on refining it". Likewise, P4 noted that "Using the ASR application made us aware of our weaknesses and helped us to compare our generated transcripts with the original ones". Moreover, P9 said that "When Speechnotes spotted our pronunciation mistakes, we started to think about and discuss the inaccuracies and their corrections to get improved on the next trails".

Theme 4: Timely and tailored feedback. Six students valued the ASR dictation software for its capacity to deliver instant and customized feedback on their diphthong pronunciation mistakes, facilitating targeted improvement. Besides, students' peers offered feedback attuned to their ability. For example, P1 stated that "The Speechnotes website assisted in identifying the specific diphthongs that I was having trouble with. It offered me more personalized feedback than my lecturer, enabling me to concentrate more on such issues and make greater progress. Moreover, my classmate modeled the pronunciation of

mispronounced words at a much slower speed by splitting the words into phonemes or syllables". Likewise, P7 reported that "The ASR tool was successful in detecting my own diphthong pronunciation errors instantly without having to wait for the lecturer feedback. With the help of my partner, it facilitated the improvement of my pronunciation in ways I could not have achieved otherwise". P10 also noted that "The software automatically identified our diphthong pronunciation errors, allowing us to observe the ASR-generated feedback in real time while dictating our scripts to the program".

Theme 5: Increased practice opportunities. Four students appreciated the use of ASR-based peer feedback for providing them with increased opportunities to practice their diphthong pronunciation inside and outside the classroom in a pressure-free environment, as illustrated by P2 who stated that "I enjoyed practicing my pronunciation with the ASR tool and my partner as I felt less nervous compared to practicing before my teacher. This process gave me more opportunities to practice inside and outside the classroom without feeling self-conscious". Similarly, P4 mentioned that "Using Speechnotes and getting feedback from my classmate created a comfortable learning environment where I could practice more without being embarrassed from reading aloud my scripts in front of the whole class and the teacher". P5 indicated that "Using the ASR-based peer feedback activities provided us with sufficient practice to enhance our diphthong pronunciation skills".

Discussion

The effects of ASR-based peer feedback

Concerning the first question, the significant progress in the experimental group students' overall EFL diphthong pronunciation on the posttest and the delayed posttest indicates that ASR-based peer feedback might have been more effective than regular instruction received by the control group students. This might be attributed to two primary factors: the ASR various affordances and the effective scaffolding provided by the peer feedback. Regarding the first factor, the implemented software provided students with a range of empowering capabilities. The students might have profited from the Speechnotes speaker-independent system which uses its massive speech database to recognize the pronounced words, making it less sensitive to accents. This, in turn, prevented mispronunciations related to the students' non-native accent (Evers & Chen, 2020). It also utilizes Google's advanced speech recognition technology, enhancing its capacity to accurately identify words with which students experienced difficulties, thereby facilitating targeted correction endeavours (McCrocklin et al., 2019). Furthermore, its intuitive design fostered seamless dictation and interaction with the tool. No technical frustrations were observed or reported throughout the experiment, which are often linked with CAPT systems (El-Garawany, 2021; Levis, 2007; Li, 2019; Neri et al., 2008; Wong, 2016). It also provided students with immediate and individualized feedback which

enabled them to see instantaneous transcriptions of their utterances tailored to their pronunciation needs. This personalized approach encouraged students to contemplate their specific weaknesses in articulating diphthongs while reviewing the generated feedback, enabling them to direct their efforts more efficiently. Guided by cognitive load theory (Sweller et al., 2011), the Speechnotes easy-to-use interface and real-time feedback, might have reduced the distraction factor associated with operating complicated programs (Ng, 2015) and the risk of cognitive overload connected with receiving delayed feedback from teachers, allowing students to focus entirely on their diphthong production (Sun, 2023). Informed by the noticing hypothesis (Schmidt,1990), the efficacy of Speechnotes could also be explained as its automated feedback facilitated noticing. The utilization of ASR prompted students to evaluate their pronunciation by analyzing the generated transcriptions of their dictation trials and comparing them to the original texts to identify discrepancies for correction, thus fostering the development of their diphthong pronunciation. Underpinned by skill acquisition theory (DeKeyser, 2015), focused practice is necessary to transform declarative knowledge of pronunciation-related rules into actionable behaviours of producing specific pronunciation features (i.e., procedural knowledge). By utilizing ASR, students engage in repeated practice where the proceduralization process becomes increasingly effortless and reaches automatization manifested in faster articulation times and fewer errors (Dai & Wu, 2021). In addition to practice, ASR feedback plays a crucial role in facilitating the proceduralization and automatization of pronunciation (Sato & Lyster, 2012) as students receive immediate and diagnostic feedback that can significantly assist in their skill acquisition (Neri et al., 2008). Moreover, pronunciation aspects associated with declarative knowledge are typically the first to be forgotten. In contrast, those aspects that have become automatized or internalized through practice are more likely to be retained (O'Malley & Chamot, 1990). Thus, the deliberate and thorough in-class and out-of-class training afforded by Speechnotes might have contributed to the students' diphthong pronunciation improvement and retention.

Concerning the second factor, drawing on the Vygotskian sociocultural theory (Vygotsky, 1978) and the interaction hypothesis (Long, 1996), the peer feedback component of the treatment might have enhanced the students' overall diphthong pronunciation improvement and retention. Since ASR applications cannot offer effective guidance in situations where users cannot differentiate between the target sound and their own pronunciation (Evers & Chen, 2020), the combination of peer instruction with ASR feedback might have yielded better outcomes in the students' diphthong pronunciation by highlighting the mistakes in their dictation that they cannot rectify on their own. With the peer assistance, the students compared their transcriptions with the model scripts and corrected their mistakes till their transcripts matched those model texts. This peer collaboration enabled the students to develop their pronunciation skills while providing several opportunities to observe and learn from their classmates throughout the whole process, thus benefiting from each other's tactics and insights. Moreover, the students engaged in pronunciation-focused

discussions to rate each other's performance using the developed scoring rubric. This facilitated their evaluation of the weaknesses in their peers' pronunciation accuracy and comprehensibility, thus increasing their motivation to refine them. Furthermore, according to the skill acquisition theory (DeKeyser, 2015), practice with peer feedback significantly enhances the processes of proceduralization and automatization compared to independent practice (Sato & Lyster, 2012). This suggests that collaborative learning environments, where students provide and receive feedback from their peers, facilitate the development of the targeted skills. This kind of interaction helps students to understand their mistakes, refine their strategies, and reinforce correct articulation, leading to a deeper internalization of such skills and a more automatic execution of tasks over time (Dai & Wu, 2021; Sun, 2023).

Unlike the average performance of the control group, the performance of the experimental group in word-, sentence-, and paragraph-level pronunciation on both the posttest and the delayed posttest was above average. Regarding word-level pronunciation, the experimental group students managed to accurately pronounce 75% of the words, whereas the control group articulated only 68% of them. Concerning accuracy in sentence- and paragraph-level pronunciation, the experimental group students pronounced most of the words containing diphthongs accurately, with few errors, indicating a good understanding of diphthong pronunciation. As for comprehensibility, their diphthong pronunciation was mostly easy to understand. However, the control group students pronounced some of the words containing diphthongs accurately, with several noticeable errors, demonstrating a basic understanding of diphthong pronunciation. Their pronunciation was sometimes difficult to understand. Thus, the quantitative results of this study showed that the experimental group significantly outperformed the control group in their overall EFL diphthong pronunciation improvement and retention, indicating the positive effects of ASR-based peer feedback. This is congruent with the findings from studies by Tsai (2019), Evers and Chen (2020), Dai and Wu (2021), Wu et al. (2021), and Sun (2023), who concluded that utilizing ASR-based peer feedback supports students' EFL pronunciation learning and offers various affordances for promoting it during practice sessions.

Experimental group students' perceptions of ASR-based peer feedback

Relating to the second question, the qualitative analysis indicated that the students' responses reflected positive perceptions regarding the use of ASR-based peer feedback. All students stated that using ASR with peer feedback served as an effective means to improve their EFL diphthong pronunciation. Most of them declared that it presented a more enjoyable and stimulating pronunciation practice alternative compared to traditional methods. More than half of them specified that the ASR dictation tool offered them individualized feedback and raised their self-awareness of their common diphthong pronunciation mistakes, allowing them to focus more on rectifying those mistakes with the help of its functionalities and their peers. Some students reported that the in-class

and out-of-class ASR-based peer feedback activities provided them with various opportunities to practice in a supportive learning environment. This indicates that the interactive features of ASR, along with peer correction, increased the students' engagement in their diphthong pronunciation acquisition. Their ability to monitor each other's progress constituted a significant motivational component in their pronunciation learning, and their capacity to visualize such progress cultivated a sense of accomplishment, encouraging the students to persist in their practice. Consequently, the amalgamation of ASR with peer feedback seemed to convert pronunciation instruction into a more interesting and rewarding educational experience, enabling the students to work at their own pace in a stress-free environment. These qualitative findings complement the quantitative ones and might clarify why using ASR-based peer feedback positively influenced the experimental group students' overall EFL diphthong pronunciation development. Though, some students were initially doubtful about its utility in enhancing their diphthong pronunciation. Thus, when introducing ASR to students, it might be advantageous to highlight its beneficial effects on their pronunciation skills, citing relevant literature. This approach could increase students' acceptance of new technologies in their EFL pronunciation learning. These findings align with the results of studies by Tsai (2019), Evers and Chen (2020), Dai and Wu (2021), Wu et al. (2021), and Sun (2023), who documented students' positive perceptions of ASR-based peer feedback as a valuable technique for promoting their EFL pronunciation performance.

Conclusion

This study aimed to achieve two main objectives: (a) to investigate the effects of ASR-based peer feedback on Egyptian English-majored students' EFL diphthong pronunciation improvement and retention, and (b) to determine the students' perceptions regarding its utilization. Two instruments were devised to collect data and answer the established research questions. Results revealed that the experimental group students who used ASR-based peer feedback in their diphthong pronunciation training significantly outclassed their control group counterparts who received regular instruction in both variables. Besides, the experimental group students' attitudes toward the utility of ASR-based peer feedback were highly positive. Thus, ASR-based peer feedback proved to have positive effects on the students' EFL diphthong pronunciation improvement and retention. These findings are significant as they demonstrate that ASR-based peer feedback can effectively enhance students' diphthong pronunciation improvement and retention (i.e., an aspect that has been underexplored in the ASR literature). Thus, this study contributes to the advancement of pronunciation research by providing EFL university educators, practitioners, and curriculum designers with a novel techno-pedagogical intervention based on synergizing ASR and peer feedback, that could be integrated into the English pronunciation instruction to address the shortcomings of the current EFL classroom environment. However, this study faced specific methodological constraints, including the short treatment duration and the small sample size. Accordingly, it explores

more possibilities for future studies about using ASR-based peer feedback in long-term investigations with larger samples, thereby augmenting the external validity of the results. Besides, the significant effects might have been influenced by assigning adult learners as participants, since they might have better interpreted the ASR feedback and provided constructive suggestions. The peer effect among young learners should be further explored in future research. Other suggestions include replicating this study through using other types of ASR software or different scaffolding strategies.

Acknowledgements

The author wishes to express her sincere thanks to the first-year English-majored students for participating in this study as well as the Editor-in-Chief, editorial staff, and anonymous reviewers of The JALT CALL Journal for reading this manuscript and providing valuable feedback to improve it.

References

Abdelreheem, H. H. S. (2023). The key English pronunciation difficulties for Egyptian EFL learners. *Anglica: An International Journal of English Studies, 32*(2), 115–136. https://doi.org/10.7311/0860-5734.32.2.06

Abker, I. A. A. (2020). Pronunciation problems of English diphthongs sounds encountered by Saudi students at Albaha University, Saudi Arabia: A case study in Almandag. *International Journal of Applied Linguistics & English Literature, 9*(4), 37–48. http://dx.doi.org/10.7575/aiac.ijalel.v.9n.4p.37

Aljabr, F. (2025). ASR using Speechnotes for EFL learners: A study of the effects on English pronunciation and prosody skills. *Journal of Ecohumanism, 4*(2), 979–987. https://doi.org/10.62754/joe.v4i2.6384

Alzahrani, S. A., & Alqurashi, H. S. (2023). Using the flipped classroom model to improve Saudi EFL learners' English pronunciation. *Linguistics and Culture Review, 7*(S1), 51–71. https://doi.org/10.21744/lingcure.v7nS1.2260

Anwar, Y., & Kalisa, P. (2020). Students' problems in pronouncing non-existing English diphthongs in Indonesian language. *Language Circle: Journal of Language and Literature, 15*(1), 17–25. https://doi.org/10.15294/lc.v15i1.26261

Ashwell, T., & Elam, J. R. (2017). How accurately can the Google web speech API recognize and transcribe Japanese L2 English learners' oral production? *The JALT Call Journal, 13*(1), 59–76. https://doi.org/10.29140/jaltcall.v13n1.j212

Aziez, F., & Sofiana, I. (2023). Deep in the diphthong problem: A study on Indonesian EFL learners' pronunciation development and the influence of individual differences. *Komposisi: Jurnal Pendidikan Bahasa, Sastra, dan Seni, 24*(2), 91–101. https://doi.org/10.24036/komposisi.v24i2.122553

Baker, A. (2006). Ship or sheep? An intermediate pronunciation course (3rd ed.). Cambridge University Press.

Baker, A., & Goldstein, S. (2008). Pronunciation pairs: An introduction to the sounds of English (2nd ed.). Cambridge University Press.

Bani Salameh, M. Y., & Abu-Melhim, A. (2014). The phonetic nature of vowels in modern standard Arabic. Advances in Language and Literary Studies, 5(4), 60–67. https://journals.aiac.org.au/index.php/alls/article/view/425/365

Behr, N. S. (2022). English diphthong characteristics produced by Thai EFL learners: Individual practice using PRAAT. Computer Assisted Language Learning Electronic Journal (CALL-EJ), 23(1), 401–424. https://callej.org/index.php/journal/article/view/386/314

Brown, K., & Miller, J. (2013). The Cambridge dictionary of linguistics. Cambridge University Press.

Chairunnisa, A. (2023). Student's perception on learning diphthong pronunciation using homophone word in Dictionary.com. https://repository.uinjkt.ac.id/dspace/bitstream/123456789/74669/1/11190140000025_SKRIPSI_ALYA%20CHAIRUNNISA_025.pdf

Collins, B., Mees, I. M., & Carley, P. (2019). Practical English phonetics and phonology: A resource book for students (4th ed.). Routledge.

Connor-Linton, J. (2006). Writing. In R. Fasold, & J. Connor-Linton (Eds.), An introduction to language and linguistics (pp. 401–432). Cambridge University Press.

Cooper, S. (2019). Speech recognition. In J. S. Damico, & M. J. Ball (Eds.), The SAGE encyclopedia of human communication sciences and disorders (pp. 1787–1790). Sage Publications, Inc. https://doi.org/10.4135/9781483380810.n580

Crystal, D. (2008). A dictionary of linguistics and phonetics (6th ed.). Blackwell Publishing.

Dai, Y., & Wu, Z. (2021). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. *Computer Assisted Language Learning, 36*(5–6), 861–884. https://doi.org/10.1080/09588221.2021.1952272

Daniels, P., & Iwago, K. (2017). The suitability of cloud-based speech recognition engines for language learning. *The JALT CALL Journal, 13*(3), 229–239. https://doi.org/10.29140/jaltcall.v13n3.j220

DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed.) (pp. 94–112). Routledge.

Derwing, T. M., Munro, M. J., & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly, 34*(3), 592–603. https://doi.org/10.2307/3587748

Edalatishams, I. (2017). *Non-native speech and recognition accuracy of two ASR applications: Dragon and Dictation.* Paper presented at the meeting of Pronunciation in Second Language Learning and Teaching, Salt Lake City, UT, United States.

Edmett, A., Ichaporia, N., Crompton, H., & Crichton, R. (2023). *Artificial intelligence and English language teaching: Preparing for the future.* British Council. https://doi.org/10.57884/78EA-3C69

El-Garawany, M. S. M. (2021). Using Praat to develop English majors' EFL intonation production. *Sohag University Journal of Education, 92*(3), 91–125. https://doi.org/10.21608/edusohag.2021.208708

El-Garawany, M. S. M. (2024). The effects of a QuillBot-based intervention on English language majors' EFL writing performance, apprehension, and self-efficacy. *Language Teaching Research Quarterly, 43,* 167–189. https://doi.org/10.32038/ltrq.2024.43.10

Elimat, A. K., & AbuSeileek, A. F. (2014). Automatic speech recognition technology as an effective means for teaching pronunciation. *The JALT CALL Journal, 10*(1), 21–47. https://doi.org/10.29140/jaltcall.v10n1.j166

Evers, K., & Chen, S. (2020). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning, 35*(8), 1869–1889. https://doi.org/10.1080/09588221.2020.1839504

Fouly, S. A., Methias, N. W., & Ahmed, H. A. (2020). Pronunciation errors and repair strategies made by Northern Upper Egypt EFL learners: A case study of the production of English sound segments by first year English Department students in Beni-Suef University. *Journal of the Faculty of Arts-Beni-Suef University, 55*(April-June), 5–22. https://doi.org/10.21608/jfabsu.2020.154604

Galal-Eldin, N. (2023). Types of English pronunciation errors among Egyptian English major post-graduate students. *Journal of the Faculty of Arts-Fayoum University, 15*(1), 1976–2030. https://doi.org/10.21608/jfafu.2023.239534.1964

Ghounane, N., & Rabahi, H. (2021). The use of computer assisted pronunciation training in teaching English pronunciation for first-year EFL students at Saida University. *International Journal of Applied Linguistics & English Literature, 10*(6), 76–81. http://dx.doi.org/10.7575/aiac.ijalel.v.10n.6p.76

Hewings, M. (2004). *Pronunciation practice activities: A resource book for teaching English pronunciation.* Cambridge University Press.

Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Canada Journal, 27*(1), 1–12. https://doi.org/10.18806/tesl.v27i1.1034

Jarosz, A. (2019). *English pronunciation in L2 instruction: The case of secondary school learners.* Springer.

Johnson, R. B., & Christensen, L. (2014). *Educational research: Quantitative, qualitative, and mixed approaches* (5th ed.). Sage Publications, Inc.

Junqua, J., & van Noord, G. (2001). *Robustness in language and speech technology.* Springer Science+Business Media Dordrecht.

Khalifa, M. F. (2020). *Errors in English pronunciation among Arabic speakers: Analysis and remedies.* Cambridge Scholars Publishing.

Kharma, N., & Hajjaj, A. (1997). *Errors in English among Arabic speakers: Analysis and remedy.* York Press.

Krismonika, Norahmi, M., & Asi, N. (2021). The effect of U-Dictionary on vowel pronunciation ability of the tenth grade students. *Journal of English Language Teaching, Linguistics, and Literature, 1*(1), 22–33. https://doi.org/10.37304/ebony.v1i1.3099

Kumar, A., & Yadav, J. (2024). Multiview learning-based speech recognition for low-resource languages. In L. A. Kumar, D. K. Renuka, B. R. Chakravarthi, & T. Mandl (Eds.), *Automatic speech recognition and translation for low resource languages* (pp. 375-404). Scrivener Publishing LLC.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27,* 184–202. https://doi.org/10.1017/S0267190508070098

Li, K. (2019). The application of Praat in English pronunciation teaching. *Advances in Social Science, Education, and Humanities Research, 378,* 374–376. https://dx.doi.org/10.2991/assehr.k.191217.150

Liakin, D., Cardoso, W., & Liakina, N. (2014). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal, 32*(1), 1–25. https://doi.org/10.1558/cj.v32i1.25962

Liakin, D., Cardoso, W., & Liakina, N. (2017). Mobilizing instruction in a second-language context: Learners' perceptions of two speech technologies. *Languages, 2*(3), 11–31. https://doi.org/10.3390/languages2030011

Lin, B. (2025). *Reinforcement learning methods in speech and language technology.* Springer.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie, & T. Bhatia (Eds.), *Handbook of research on second language acquisition* (pp. 413-468). Academic Press.

McCrocklin, S. (2019). ASR-based dictation practice for second language pronunciation improvement. *Journal of Second Language Pronunciation, 5*(1), 98–118. https://doi.org/10.1075/jslp.16034.mcc

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System, 57*(2016), 25–42. http://dx.doi.org/10.1016/j.system.2015.12.013

McCrocklin, S., Humaidan, A., & Edalatishams, E. (2019). ASR dictation program accuracy: Have current programs improved? In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference, Ames, IA, September 2018* (pp. 191–200). Iowa State University. https://iastate.box.com/shared/static/wtnv3yg890ze2ibtkihwdpts7bfojt8h.pdf

Moore, M. (2021). Speech recognition for individuals with voice disorders. In T. McDaniel, & X. Liu (Eds.), *Multimedia for accessible human computer interfaces* (115–144). Springer. https://doi.org/10.1007/978-3-030-70716-3_5

Mroz, A. P. (2018). *Noticing gaps in intelligibility through automatic speech recognition (ASR): Impact on accuracy and proficiency.* Paper presented at 2018 Computer-Assisted Language Instruction Consortium (CALICO) Conference, Urbana, IL, United States.

Naji, M. H. S., & Almakrob, A. Y. (2023). Arabic-speaking EFL learners' pronunciation of British English vowels: A production-based study. *Journal of Language Teaching and Research, 14*(3), 730–739. https://doi.org/10.17507/jltr.1403.21

Nekoobahr, F., Hawkins, J., Santi, K. L., Antonelli, J. R. S., & Thorpe, J. L. (2021). Identifying and evaluating language-learning technology tools. In M. Khosrow-Pour (Ed.), *Handbook of research on modern educational technologies, applications, and management* (pp. 160–177). IGI Global.

Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning, 21*(5), 393–408. https://doi.org/10.1080/09588220802447651

Ng, W. (2015). *New digital technology in education: Conceptualizing professional learning for educators.* Springer.

Ngo, T. T., Chen, H. H., & Lai, K. K. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL, 36*(1), 4–21. https://doi.org/10.1017/S0958344023000113

Nurjannah, N., Nofiana, N., Jannah, M., & Farah, F. (2023). Enhancing EFL learners diphthong pronunciation ability's using PRAAT. *SCOPE: Journal of English Language Teaching, 8*(1), 139–146. https://journal.lppmunindra.ac.id/index.php/SCOPE/article/viewFile/19567/6071

O'Connor, J. D. (1980). *Better English pronunciation* (2nd ed.). Cambridge University Press.

O'Connor, J. D., & Fletcher, C. (1989). *Sounds English: A pronunciation practice book.* Longman.

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition.* Cambridge University Press.

Offerman, H. M., & Olson, D. J. (2016). Visual feedback and second language segmental production: The generalizability of pronunciation gains. *System, 59*(2016), 45–60. https://doi.org/10.1016/j.system.2016.03.003

Özdemir, S. (2019). *The role of listening skills in the pronunciation of diphthongs through flipped classrooms.* [Master's thesis, Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü, Turkey]. https://hdl.handle.net/20.500.12462/9978

Pennington, M. C., & Rogerson-Revell, P. (2019). *English pronunciation teaching and research: Contemporary perspectives.* Palgrave Macmillan.

Pratiwi, D. R., & Indrayani, L. M. (2021). Pronunciation error on English diphthongs made by EFL students. *Teknosastik, 19*(1), 24–30. https://doi.org/10.33365/ts.v19i1.486

Richards, J. C. (2015). *Key issues in language teaching*. Cambridge University Press.

Richards, J., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics* (4th ed.). Pearson Education Limited.

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning, 62*(2), 595–633. https://doi.org/10.1111/j.1467-9922.2011.00639.x

Saleh, A. J., & Gilakjani, A. P. (2021). Investigating the impact of computer-assisted pronunciation teaching (CAPT) on improving intermediate EFL learners' pronunciation ability. *Education and Information Technologies, 26,* 489–515. https://doi.org/10.1007/s10639-020-10275-4

Sari, P. K., Indrawati, N. L. K. M., & Aryawibawa, I. N. (2024). English diphthongs pronunciation errors found in North Bali bilingual school's 5th grade students. *Sintaksis: Publikasi Para ahli Bahasa dan Sastra Inggris, 2*(5), 138–147. https://doi.org/10.61132/sintaksis.v2i5.1004

Sato, M., & Lyster, R. (2012). Peer interaction and corrective feedback for accuracy and fluency development: Monitoring, practice, and proceduralization. *Studies in Second Language Acquisition, 34*(4), 591–626. https://doi.org/10.1017/S0272263112000356

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11,* 206-226. https://doi.org/10.1093/applin/11.2.129

Štuikys, V., & Burbaitė, R. (2024). *Evolution of STEM-driven computer science education.* Springer.

Sun, W (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. *Frontiers in Psychology, 14*(2023), 1–14. https://doi.org/10.3389/fpsyg.2023.1210187

Swan, M., & Smith, B. (2001). *Learner English: A teacher's guide to interference and other problems* (2nd ed.). Cambridge University Press.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Tsai, P. (2019). Beyond self-directed computer-assisted pronunciation learning: A qualitative investigation of a collaborative approach. *Computer Assisted Language Learning, 32*(7), 713–744. https://doi.org/10.1080/09588221.2019.1614069

Vygotsky, L. (1978). *Mind in society: Development of higher psychological processes.* Harvard University Press.

Wilson, K. (2024). *Computer jargon: The illustrated glossary of basic computer terminology.* Elluminet Press.

Wong, C. (2016). Teaching pronunciation to learners of English as a lingua franca. In W. Renandya, & H. Widodo (Eds.), *English language teaching today: Linking theory and practice* (pp. 241–255). Springer. https://doi.org/10.1007/978-3-319-38834-2_17

Wu, X., Liu, X., & Chen, L. (2021). Reducing EFL learners' error of sound deletion with ASR-based peer feedback. In W. Jia, Y. Tang, R. S. T. Lee, M. Herzog, H. Zhang, T. Hao, & T. Wang (Eds.), *Emerging technologies for education. SETE 2021. Lecture Notes in Computer Science, 13089* (pp.178–189). Springer. https://doi.org/10.1007/978-3-030-92836-0_16

Xiao, H., Ou, K., Wang, H., & van de Weijer, J. (2023). The effects of ASR apps on monophthong pronunciation improvement and generalization to new words in English. In K. Arai (Ed.), *Intelligent Computing. SAI 2023. Lecture Notes in Networks and Systems, 711* (pp. 1410–1433). Springer. https://doi.org/10.1007/978-3-031-37717-4_93

Yaniafari, R. P., Olivia, V., & Suharyadi (2022). The potential of ASR for improving English pronunciation: A review. *Proceedings of the International English Language Teachers and Lecturers (iNELTAL), KnE Social Sciences 2022*, 281–289. https://doi.org/10.18502/kss.v7i7.10670

Yuan, Y., & Liu, X. (2020). An empirical study of the effect of ASR-supported English reading aloud practices on pronunciation accuracy. In L. Lee, L. H. U, F. L. Wang, S. K. S. Cheung, O. Au, & K. C. Li (Eds.), *Technology in education: Innovations for online teaching and learning. ICTE 2020. Communications in Computer and Information Science, 1302* (pp. 75–87). Springer. https://doi.org/10.1007/978-981-33-4594-2_7

Yule, G. (2014). *The study of language* (5th ed.). Cambridge University Press.

Ziane, S., & Dou, A. (2024). Using ASR technology to assess EFL learners' pronunciation. *Proceedings of IAC in Prague 2014,* 95–103. https://www.conferences-scientific.cz/file/9788088203384

Appendix A

EFL Diphthong Pronunciation Test (DPT)

1. Read the following lists of words aloud: (40 marks)

1.	tones	pure	avoid	shape	doubt
2.	bear	tourist	white	grouse	real
3.	coat	loiter	kind	mouth	rarely
4.	sail	voice	cruel	stone	cure
5.	furious	boils	trade	rise	fierce
6.	down	airs	lied	steered	home
7.	wary	same	poor	joint	beard
8.	proud	phoned	hair	night	paper

2. Read the following sentences aloud: (10 marks)

1. I can see annoying black bugs in this hay.
2. Mustafa got three toys for his birthday.
3. He found that it's the best cow in the dairy farm.
4. Mr. Lear looked for Snowball in the closet.
5. Joe tried not to tear the jacket.
6. The insurers were delighted with the business.

7. It is clear that it is a very unusual pier.
8. Roy wasn't curious about Claire.
9. Farida had a wonderful private boat tour.
10. Her husband left her facing a mountain of debt.

3. Read the following paragraph aloud: (10 marks)

In here, the dearest things were the electrical appliances- things like a dish-washer, a refrigerator, a toaster, and a stove. There was no tableware, so I bought some knives and forks. I'd been given some bowls, but no cups and saucers, so I bought six of each. I needed furniture, too. I had to make some difficult choices. I bought brown curtains for the dining room. That wasn't my idea, but they should look really great. They'll help keep out the noise. And I obtained a nice old wooden table and some chairs. They cost about a thousand dollars. I've painted the walls pale yellow and hung a large poster near the window. I've put an oil painting on the opposite wall with a strange pair of lamps. Yesterday, I bought a ceiling light, a shower, and a beige carpet for the stairs. Another job is to paint the outside of the house and the garage. Surely, I'll enjoy doing that.

Scoring rubric for the EFL Diphthong Pronunciation Test (DPT)
(Questions 2 and 3)

Level / Skill	1 Very poor	2 Poor	3 Fair	4 Good	5 Very good
1. Accuracy	Students pronounce words containing diphthongs inaccurately and unclearly, with significant errors, demonstrating little or no understanding of diphthong pronunciation.	Students pronounce few words containing diphthongs accurately and clearly, with many errors, demonstrating a limited understanding of diphthong pronunciation.	Students pronounce some words containing diphthongs accurately and clearly, with several noticeable errors, demonstrating a basic understanding of diphthong pronunciation.	Students pronounce most words containing diphthongs accurately and clearly, with minimal errors, demonstrating a good understanding of diphthong pronunciation.	Students pronounce all words containing diphthongs accurately with perfect clarity and consistency, demonstrating a strong understanding of diphthong pronunciation.
2. Comprehensibility	Students' diphthong pronunciation is impossible to understand.	Students' diphthong pronunciation is often difficult to understand.	Students' diphthong pronunciation is sometimes difficult to understand.	Students' diphthong pronunciation is mostly easy to understand.	Students' diphthong pronunciation is always easy to understand.

Appendix B

Semi-structured interview guide

1. Can you express your experience of using ASR-based peer feedback in your diphthong pronunciation training? What aspects did you find most beneficial and what challenges did you encounter during the process?
2. How do you think ASR-based peer feedback impacted your EFL diphthong pronunciation? Please, include specific instances to illustrate this impact.
3. In what ways did ASR-based peer feedback differ from traditional instruction? Which method do you perceive to be more useful, and why?
4. Did you enjoy using the ASR dictation tool with peer feedback? How satisfied were you with using it in developing your EFL diphthong pronunciation? Please, give reasons.