Bouchhioua, N. (2024). FOSTERING THE INTERPLAY BETWEEN ACOUSTIC PHONETICS AND AI-POWERED PRONUNCIATION LEARNING: A TEACHER-ACTION RESEARCH. *Teaching English with Technology*, *24*(3), 41–67.
https://doi.org/10.56297/vaca6841//BFFO7057/YGDQ5334

FOSTERING the INTERPLAY BETWEEN ACOUSTIC PHONETICS and AI-POWERED PRONUNCIATION LEARNING: A TEACHER-ACTION RESEARCH

Nadia Bouchhioua[1]

[1] English Department, Manouba University, Tunisia

Keywords: AI-powered pronunciation apps, CAPT systems, acoustic phonetics, prosody

https://doi.org/10.56297/vaca6841//BFFO7057/YGDQ5334

Teaching English with Technology

Vol. 24, Issue 3, 2024

Most regular practices of teacher-action research in ELT consist of identifying performance problems and providing specific treatments for the sake of improvement. However, teacher-action research does not dictate a specific methodology. The present teacher-action research is concerned with the implementation of an Advanced Prosody course in the Tunisian higher education context. The course aims to raise students' awareness of the interplay between experimental phonetics and speech technology in the development of computer-assisted pronunciation training (CAPT) systems and AI-powered pronunciation apps. Students' engagement with the course as well as their beliefs and attitudes toward the utilization of AI-powered apps to learn English pronunciation were measured through an online survey. A quantitative and qualitative analysis of their responses revealed a high engagement with the course content. Although the participants expressed positive attitudes toward the use of these apps, their responses revealed a dual awareness. They acknowledged the distinct value of the teacher-student connection. They also acknowledged the necessity for teachers to embrace technology and become proficient in its efficient utilization. The findings are discussed within the ongoing debate about the integration of AI in the educational sphere.

1. Introduction

Within the educational sphere, there is a growing belief that as artificial intelligence (AI) systems evolve, they can assume tasks previously carried out by humans, such as grading exams, providing feedback, and other activities traditionally associated with teaching. This shift has raised concerns among educators, parents, and stakeholders worldwide who fear that excessive reliance on AI could hinder students' development of critical thinking and problem-solving skills, leading to decreased self-reliance.

The emergence of ChatGPT heightened these concerns, particularly regarding its potential risks when utilized by learners. However, the integration of AI in education, including language teaching and learning, has roots that extend back several decades with the development of computer-assisted language learning (CALL) systems. CALL has evolved into a distinct discipline that leverages new technologies and media to enhance language

education. Yet, it is essential to question why there is apprehension toward a field like AI, which has historically thrived on interdisciplinary collaboration and innovation. Indeed, advancements in various sciences have continually contributed to computer science, including AI, driving the creation and implementation of diverse applications, software, and systems.

In this context, this paper highlights how advancements in phonetic and phonological research have provided a solid theoretical foundation for computer-assisted pronunciation training (CAPT) systems and pronunciation learning apps, aligning with new objectives in the field of L2 teaching. Furthermore, it presents findings from a survey conducted among Tunisian EFL (English as a Foreign Language) learners who participated in an Advanced Prosody course blending acoustic phonetics and AI-powered pronunciation training systems and speech analyzer software. This study, conducted as teacher-action research in the Tunisian EFL context, delves into students' engagement with and responses to the Advanced Prosody course while exploring their beliefs and attitudes toward technology and the role of AI in language learning. The following section examines the literature concerning CAPT systems and AI-powered pronunciation apps. It endeavors to elucidate the key role that experimental phonetics has played in shaping the development of CAPT systems, pronunciation apps, and speech analysis software.

2. Literature review

2.1. CAPT systems and AI-powered pronunciation apps

Lim and Toh (2024) provided a systematic review of research studies published from 2010 to 2021 on language apps used to teach English as a second language. Most of the research surveyed was conducted on apps that supported teaching subsystems and skills such as grammar, vocabulary, reading, writing, speaking, and listening. However, the review did not focus on apps or software devoted to supporting pronunciation teaching.

Automatic speech recognition (ASR) systems and CAPT programs are among the earliest and most widely used programs that integrate technology in the domain of education by utilizing ASR in pronunciation teaching and learning. Technological advancements have seamlessly integrated ASR systems into foreign language pronunciation training, revolutionizing computer-assisted language learning. These speech-enabled systems offer a myriad of advantages for language learners. They cleverly address individual learners' challenges, allowing for practice at a personalized pace. Thus, they liberate learners from the constraints of teacher availability.

CAPT systems provide automatic, instant feedback on pronunciation, which greatly benefits individual learners (Pennington, 1999; Rogerson-Revell, 2021). Moreover, advancements in ASR technology enable feedback to be tailored to the specific needs of each learner. The tireless and non-judgmental

nature of these applications engages students in autonomous study, providing limitless opportunities to review materials and access additional assistance. These systems facilitate the storage of student profiles in specialized records, allowing both students and teachers to monitor progress and address any challenges encountered.

CAPT systems play a crucial role in assisting students who struggle with public speaking, offering them valuable support in improving their language skills (Levis, 1997; Levy, 1997). They offer the learner an interactive environment in a range of modes such as whole class, small group, pair, and teacher-to-student (Pennington, 1999). Similarly, teachers benefit from employing CAPT systems in their pronunciation classes as they offer students drilling practice, which teachers find tedious and time-consuming.

Rogerson-Revell (2021) suggests that CAPT has the potential to enhance conventional learning approaches by providing more opportunities for exposure to a wide range of spoken language variations and accents, encompassing diverse L1 and L2 accents as well as various speech genres and styles. Additionally, CAPT systems offer the possibility of incorporating diverse audiovisual content through various platforms, spanning from educational websites to mobile applications and social media platforms. This incorporation of visual elements underscores the potential benefits of integrating visual animations into language learning tools, which can assist learners in visualizing and understanding the production of sounds (Lord, 2021).

Some CAPT systems are designed to address the pronunciation needs of learners with specific first-language backgrounds. An exemplary instance of such CAPT models is the system developed by Kawai and Hirose (2000), which enables the assessment of intelligibility and non-nativeness of phone quality in language pronunciation training, specifically designed for Japanese learners of English. The system detects errors in the choice of phones, reports the degree of non-nativeness of the learner's pronunciation, and suggests ways to improve spoken language abilities. A similar system is SLIM—an Italian acronym for Multimedia Interactive Linguistic Software—developed by Delmonte et al. (2004) to help Italian learners of English communicate intelligibly and as close as possible to natives. The authors succeeded in developing SLIM, based on the information provided by experimental phonetics, to detect significant deviation from a native speaker's word, phrase, or utterance production. It also offers a visual aid and a written diagnosis of the problem, as well as an indication of how to overcome and correct the mistake. CAPT systems rely significantly on dense phonetic information and input in their design, necessitating collaboration between phoneticians and speech technologists.

2.2. The role of experimental phonetics in the development of CAPT systems

Early attempts to create systems for automatic speech recognition (ASR) were mostly directed by the theory of acoustic phonetics. Acoustic phonetics measures speech segments (consonants and vowels) and provides explanations and descriptions of how they are physically realized in a spoken utterance. It provides measurements of the duration of speech segments, their fundamental frequency (F0), and their intensity. It also provides values of vowel formants. In the early 1950s, Davis et al. (1952) built a system that recognized isolated digits spoken by a single speaker with the formant frequencies measured during vowel regions of each digit.

Similar acoustic-phonetic knowledge has also nourished the design of other recognition systems based on syllable and vowel duration of single speakers (Forgie & Forgie, 1959; Olson & Belar, 1956). The reciprocal exchange of knowledge and the intricate interplay between phonetic science and computer science have steadily evolved, being manifested in a broad range of systems that now encompass sophisticated modern computer programs. Indeed, due to its inherently scientific nature, which involves the study of the tangible phenomenon of human speech, which can be quantified and measured, phonetics establishes a robust connection with computer science and technology, particularly in the realms of education and academia.

Experimental phonetics employs empirical methods and speech analysis software for the acoustic analysis of speech sounds. The measurements and insights it generates regarding speech sounds and their attributes serve as foundational elements for the creation of automatic speech recognition systems, speech analysis software, and innovative CAPT programs. Collaboration between the phonetician and the speech technologist is best manifested in the development of applications such as *Praat*, a free computer software package that analyzes speech phonetics, which was created by Boersma and Weenink (2018). This software records speech utterances, sets the recorded sound at the required volume, saves the recorded material in a .wav file format, segments and labels the speech signal, creates text grids for the segmented data, and saves them.

The functionalities of this software help the user determine the acoustic characteristics of speech sounds and study prosodic features such as stress, intonation, and speech rhythm. Figure 1 below shows some of the functionalities offered by *Praat* as measurements of a speech signal. The waveform and the spectrogram of the English phrase “He asked”, produced by a Tunisian female EFL learner (a participant in the present study), are displayed in the screenshot in Figure 1. The blue line indicates the pitch contour. Different options and settings are available to measure the pitch range, trace the intonational contour, as well as to measure the duration and the intensity of the speaker’s voice.

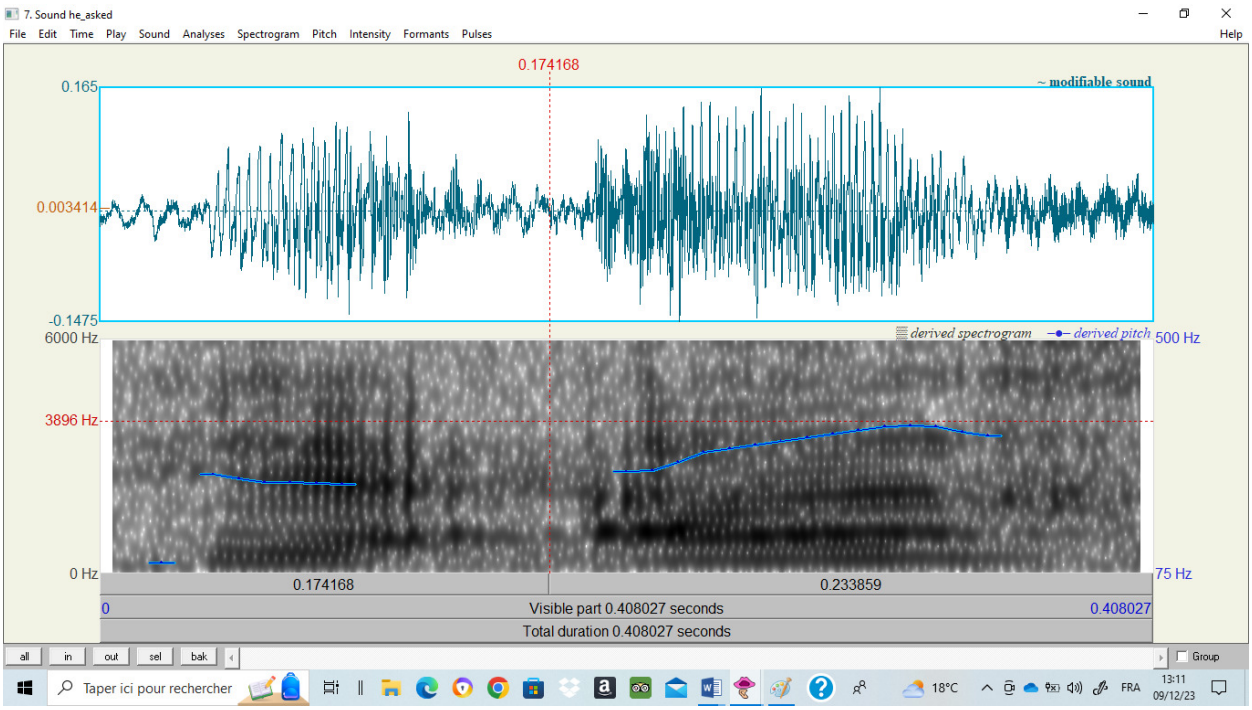

Figure 1. Voice trace of a Tunisian female speaker producing the English utterance "He asked"

The synergistic partnership between linguists, who furnish crucial phonetic insights, and speech technologists, who have technical prowess in system development, culminates in the creation of indispensable tools benefitting learners and educators alike. The phonetic insights offered regarding sound duration variations across different languages seem to be crucial for teaching languages such as English, where duration is a correlate of lexical stress, to students whose native language employs phonemic duration, such as Arabic. For instance, Bouchhioua (2008a) reported that Tunisian EFL learners consistently produced English segments and words longer than native English speakers, which may reveal their non-nativeness. Language apps, which provide individualized feedback to learners depending on their specific needs—segment and word duration in this case—and L1 background, could be very useful.

In another study, Bouchhioua (2008b) showed that while both vowel formant values (F1 and F2) are affected by lexical stress in English, only gradient F1 lowering could be used to predict lexical stress in Tunisian Arabic. Tunisian EFL learners seem to have transferred this tendency from their mother tongue. They did not reduce unstressed vowels in their production of English utterances. Incorporating comparative phonetic research of this nature, which uses acoustic measurements facilitated by advanced speech analysis software, constitutes a rich source of technical phonetic insights. These insights can be harnessed in constructing pronunciation apps tailored to individual learners, accounting for the prosodic nuances inherent in their native language.

The production of pitch is characterized by significant variability across individuals, languages, and even dialects. This diversity underscores the importance of CAPT programs, such as *My English Tutor (My ET)*, that analyze pitch patterns generated by non-native speakers and juxtapose them with native speech. These programs are widely embraced and adaptable across diverse linguistic and contextual settings. The strength of those systems is that they obtain intelligibility scores and instruct the non-native learner on how to correct his/her pronunciation (Benchaaben, 2023). Those systems use speech recognition algorithms to accurately measure pitch and align it with the location of each phone in the learner's speech. This technology could be useful in teaching pitch accents and intonation contours to non-native learners.

Critical voices have scrutinized pronunciation apps and CAPT systems for their emphasis on isolated articulatory mechanics, which some argue mirror audiolingual methods involving repetitive drills and mimicry. Critics contend that such approaches may fail to foster communicative or phonological competence within a language. In response to this criticism and with the continuous growth of AI and phonetic research, more sophisticated, pedagogically informed systems were developed through technological and academic collaborations, such as Cauldwell's (2012) *Cool Speech* application (http://www.speechinaction.org/cool-speech-2), which focuses on the features of fluent natural speech and is based on extensive academic research. There is also the *Sounds of Speech* application (https://soundsofspeech.uiowa.edu) that uses AI to emphasize communication. Peng et al. (2018) evaluate a 3-D talking head on an application developed for learning Mandarin Chinese. Such animated heads converse with the user, aiming to augment personalized pronunciation training for non-native Mandarin language learners. As another application exemplifying the combination of conversational AI and language learning, *ELYSAI* (https://www.elysai.com/) is designed with a specific focus on aiding learners in refining their pronunciation abilities. Through interactive dialogues facilitated by AI-driven interfaces, *ELYSAI* empowers users to engage with virtual companions on a platform for immersive conversational experiences. Using advanced AI technologies, this application provides learners with real-time feedback on pronunciation accuracy and tailored guidance for improvement. Figure 2 illustrates the use of this app by a Tunisian EFL learner who is a participant in the action research reported in this paper.

Previous research examining the use of these applications in pronunciation learning often overlooks the role of phonetic science in their development. Additionally, few studies offer a comprehensive exploration of learners' engagement in such courses or their perceptions of the benefits and limitations of employing AI- powered apps and CAPT systems in L2 pronunciation learning. This paper seeks to address these gaps.

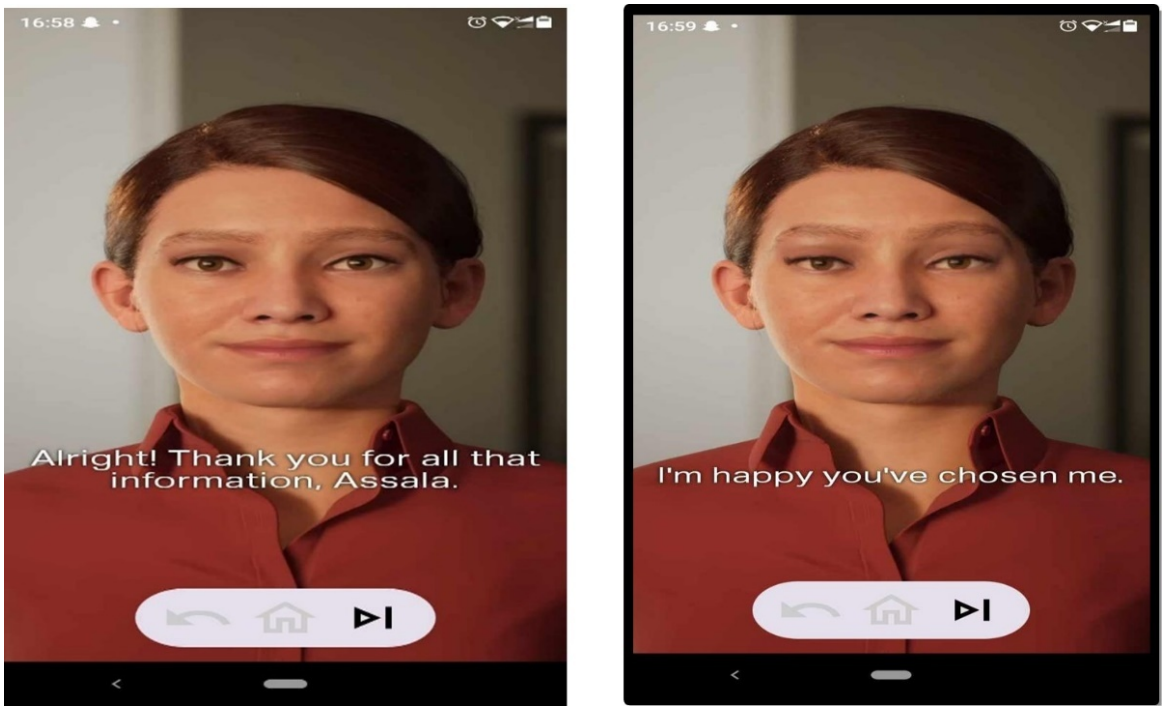

Figure 2. Screenshots of the *ELYSAI* app used by a Tunisian EFL learner

3. Methodology

The present investigation constitutes a teacher-action research attempt, focusing on students' engagement with and reaction to an Advanced Prosody course, alongside an exploration of their beliefs and attitudes concerning technology and AI. It is an optional course taken as one of the requirements of the Master's degree in linguistics for English major students. It integrates elements of acoustic phonetics, CAPT systems, and speech analyzer software. Its primary objective is to foster students' comprehension of the symbiotic relationship between experimental phonetics and speech technology. The study's specific context is the Tunisian EFL setting, characterized by diglossia and multilingualism. Within this milieu, Tunisian Arabic, representing a "low" form of Arabic, functions as the vernacular (mother tongue), whereas Modern Standard Arabic, designated as the "high" variety and taught in schools, serves as the official language. Additionally, French is acquired as a second language, while English and other European languages constitute foreign language instruction within the educational framework. The action research reported in this paper aims to answer the following questions:

1. How did students engage with and respond to the Advanced Prosody course under the guidance of the teacher-researcher?
2. What are the prevailing beliefs and attitudes among Tunisian EFL learners regarding the utilization of AI-powered pronunciation apps and CAPT systems?

Table 1. The Advanced Prosody course details

Duration	No. of students	Age	Gender & nationality	Language background
24 hours (two hours over 12 weeks)	20	22-23	All females All Tunisians in nationality (three had mothers who are not Tunisian)	• Native language: Tunisian Arabic • French and English as foreign languages • English is the major language of study in the BA and MA program.

3.1. Course details

The teacher-researcher delivered this course to a cohort of 20 MA students from a Tunisian public university, spanning twelve weeks with sessions held for two hours per week, between October and December 2024. Table 1 provides details of the course and the participants. The level of proficiency in English of the participants at the beginning of the course was B2 according to the CEFR.

3.2. Course objectives

This course was designed to acquaint MA students in linguistics with the analysis of prosodic elements within the English sound system, including stress, intonation, and rhythm. These aspects pose challenges for many non-native English learners of English as a second language (ESL) or foreign language (EFL). Named "Advanced Prosody", this course delved into intricate facets of phonetic science, including physical cues of prosodic elements, areas typically untouched in the undergraduate-level phonetic coursework. Students were guided in understanding how these features are often either influenced by their L1 characteristics or produced with an approximation of the English prosodic system, resulting in non-native-sounding speech patterns (Celce-Murcia et al., 1996).

Under the teacher-researcher's guidance, students were trained to approach these features through the lens of second language acquisition research. The course delved into the nature, components, and functions of these prosodic features within both native English and L2 English contexts. Moreover, discussions explored the pivotal roles of these features in facilitating successful communication among users of English as an international language.

As a preparatory endeavor, this course equipped students with the foundational knowledge necessary for researching these aspects of the English language. It fostered their awareness of the interconnectedness among acoustic phonetics, computer science, and AI in developing speech analysis tools like *Praat* and in crafting and refining CAPT systems and pronunciation apps.

3.3. Course methodology

The course was goal-oriented and project-based. Students were asked to form groups of two to four and choose a topic from the course outline. They were tasked with conducting research about the topic and delivering it orally through PowerPoint. Students were mandated to record spoken utterances produced by native and non-native speakers of English and to analyze them acoustically. Students were helped, as the course progressed, to develop knowledge about the prosodic system of English and to compare it to prosodic systems of other languages.

The teacher-researcher introduced students to the use of the speech analyzer software *Praat*. Training included recording sounds, making acoustic measurements, and analyzing different sound patterns using this software. In addition, students were also introduced to various CAPT systems, such as *Tell Me More* and *My English Tutor (My ET*), and to various language apps, such as *Duolingo, ELSA speak,* and *ELYSAI*, all of which focus on teaching English pronunciation. As the course progressed, students gained insights into the roles of technology, AI, and acoustic phonetics in developing these systems. They also developed an appreciation for the collaborative efforts between phoneticians, speech technologists, and computer engineers in advancing technology, research, and teaching methods.

Throughout the course, tasks assigned by the teacher-researcher gradually familiarized students with *Praat* and its diverse functionalities. This included recording speech, adjusting sound levels, saving recordings in appropriate formats, segmenting and labelling speech signals, and creating text grids for segmented data, among other tasks.

Assignments in the course were designed to assess the various skills developed throughout the duration of the program. One assignment involved students recording themselves reading a short English text aloud. This text included specific target items for measuring word stress, weak forms, and intonation acoustically. Students were instructed to note the duration (in milliseconds), F0 (in Hertz), and intensity (in decibels) values of the target syllables in their speech and compare these values to those of native English speakers (Appendix A).

Previous research (Ghazali & Bouchhioua, 2003) reported a severe intelligibility loss among a comparable group of Tunisian EFL learners because of a lack of vowel reduction in unstressed syllables and the overstressing of function words, due to L1 interference. Therefore, the objective of this activity was to raise students' awareness of the importance of correct stress placement and vowel reduction in function words for intelligibility and successful communication.

Toward the end of the course, students were tasked with writing an individual research paper on a related topic. This paper was expected to highlight their understanding of the topics covered in class, their ability to synthesize and critically reflect on these topics, as well as demonstrate their proficiency in academic writing and referencing skills. Assessments were conducted based on the quality of the oral presentations, the completion of assignments, and the quality of the written research paper.

3.4. The survey

After the course concluded, students were invited to participate in an online survey, designed by the teacher, anonymously. Its primary objective was to assess students' reactions to the course and their level of engagement with its content. Additionally, the survey aimed to measure students' attitudes regarding the utilization of technology and AI-powered systems in learning English pronunciation. It consisted of 15 questions, comprising 12 closed questions and three open-ended questions. The inclusion of open-ended questions allowed participants to express their concerns and ideas freely, providing valuable qualitative insights. The types of questions in the survey were factual, behavioral, and attitudinal, covering a range of perspectives and experiences related to the course content and technological tools used.

Given that the questionnaire included both closed and open-ended items, the data analysis process would incorporate both quantitative and qualitative approaches. Quantitative analysis would involve quantifying responses to closed questions, while qualitative analysis would involve interpreting responses to open-ended questions to gain deeper insights into students' perceptions and experiences (Appendix B).

3.5. Results

Twenty female students participated in the course and completed the questionnaire. Among them, 85% fell within the age range of 22 to 25, with the remaining 15% being older than 25. All 20 participants hailed from Arabic-speaking backgrounds, with 17 having Tunisian lineage on both sides. Notably, three participants had non-Tunisian mothers, which accounted for the discrepancy where only 17 out of 20 indicated proficiency in French—a language ingrained in Tunisia's education system from as early as 7-8 years old, serving as a widely spoken second language in the country. All the participants were proficient in English (B2 level), aligning with their academic focus, with 100% reporting fluency in the language. A mere 3% of respondents noted additional linguistic competencies beyond Arabic, French, and English as displayed in Figure 3.

In the second segment of the questionnaire (Questions 4-8), students' attitudes regarding their course experience were examined. Question 4 delved into their overall impressions of the Advanced Prosody course. The feedback

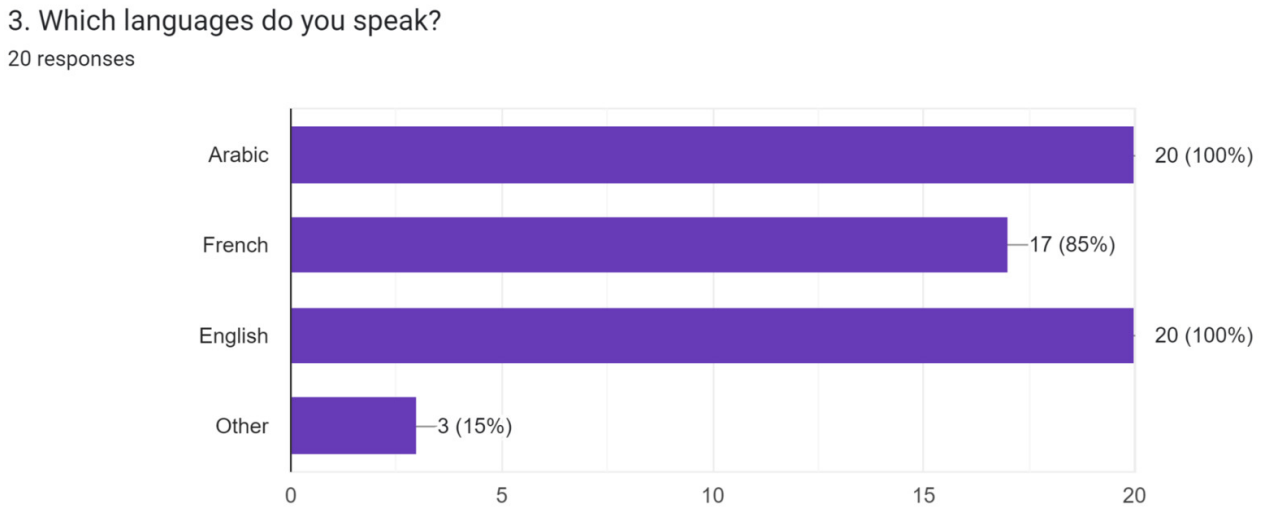

Figure 3. Languages spoken by the participants

revealed that the vast majority of participants (95%) found the course "Interesting". Notably, none of the participants expressed finding the course either too challenging or overly simplistic for their level of proficiency.

Question 5 aimed to disclose the aspects that contributed to the respondents' perception that the course was of interest. Multiple options were provided, and participants were instructed to select all that applied to them (Appendix B). Among the options, the "prosody component (stress, intonation, rhythm, juncture, etc.)" received the highest response rate at 92.3%, indicating a strong interest in this area. Following closely was the "pronunciation learning and teaching component" at 76.9%, demonstrating a significant engagement with this aspect of the course. The "acoustic phonetic component" received 38.5% of responses, suggesting a moderate level of interest. Regarding the use of *Praat* and its functionalities, 61.5% of respondents expressed interest in this tool. Surprisingly, the combination of all components received a lower response rate of 30.8%, indicating that students had a stronger preference for individual components rather than their amalgamation (Figure 4).

Afterward, students were queried regarding their prior knowledge about the role of acoustic phonetics in the creation of computer systems like *Praat* and other pronunciation learning software (Question 7). Among the 20 respondents, 55% answered affirmatively ("Yes"), while the remaining 45% replied negatively ("No"). Subsequently, Question 8 sought to delve into how the course altered participants' awareness regarding the interaction between phonetic science and computer science in developing such systems. The focus was on the 45% who initially responded "No" to Question 7, constituting eight participants. This question was open-ended, allowing participants to express their viewpoints freely. Consequently, the eight responses underwent qualitative analysis through thematic analysis for deeper insights.

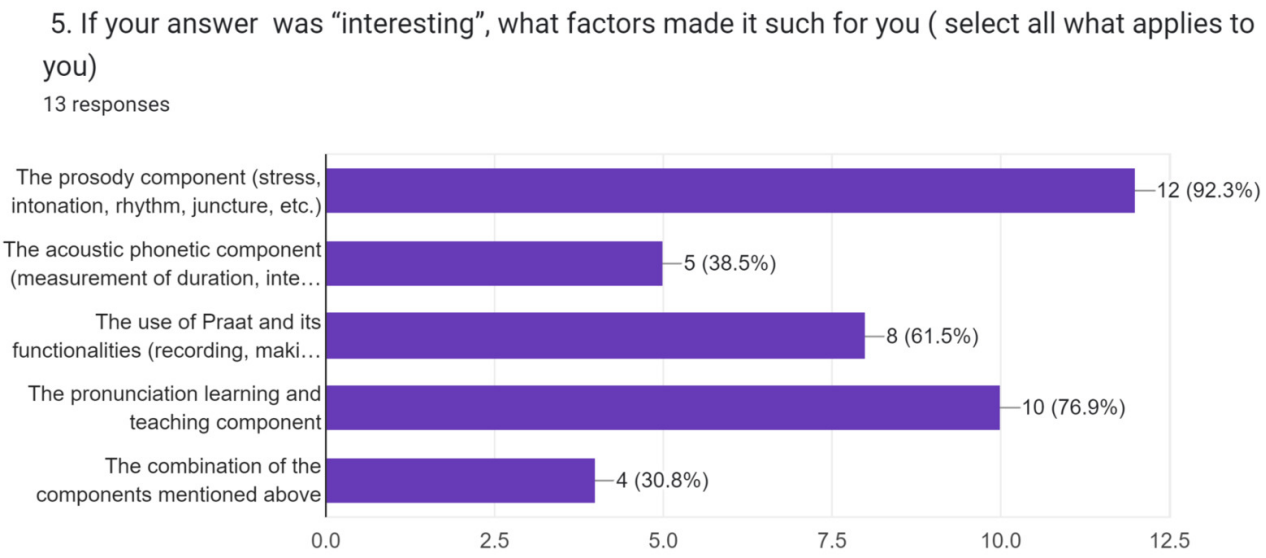

Figure 4. Students' responses to Question 5

The qualitative thematic analysis of the students' reflections on Question 8, "How did the course change your awareness about the interaction between phonetic science and computer science for the development of such systems?", unveiled a prevalent theme centered on the *Praat* software. Students viewed *Praat* as a key example of the fusion between acoustic phonetics and computer science, facilitating the analysis of phonetic attributes such as pitch and duration of speech sounds. Their insights underscored the essential role of acoustic phonetics in quantifying the physical aspects of human speech, synergizing with computer science to advance and refine such systems. This collaboration was eloquently captured in their recurrent use of terms like *"complementarity"*, *"combination"*, and *"interaction"*. The course successfully achieved its objective of heightening students' awareness of this symbiotic relationship, as evidenced by expressions in their answers to Question 8, such as: "*Through the course I learnt"*; *"The course showed me"*; *"The course raised my awareness*"; and *"I became aware"*. The responses from one of the participants adeptly integrated and synthesized all these ideas.

> The Advanced Prosody course showed me how linguists and computer scientists work together on systems like Praat and pronunciation apps. Linguists understand speech sounds, while computer scientists use their skills to analyze speech using things like spectrograms, pitch, and intensity. This teamwork makes sure these systems are both accurate and efficient. Without it, the computers wouldn't be able to do the job properly.

Another student wove together these ideas in her response:

> I learned that acoustic phonetics plays an important role in the development of computer systems such as praat. this is partially because the theoretical framework and methodologies needed to analyze various aspects of speech, such as pitch, intensity and duration are provided by acoustic phonetics. indeed, throughout it we focus on the movements of articulatory organs (lips, tongue, vocal cords..) and how these movements create specific acoustic patterns.

Though not articulated with complete precision, the students' responses demonstrate an effort to comprehend the course content and its objectives.

The following segment of the questionnaire (Questions 9-15) delved into the utilization of pronunciation apps and CAPT systems, aiming to understand students' behaviors and attitudes towards these tools. Question 9 specifically explored their pre-course engagement with speech analyzer software or any pronunciation apps. The findings revealed that a substantial majority (85%) had already acquainted themselves with such systems before enrolling in the course (Figure 5).

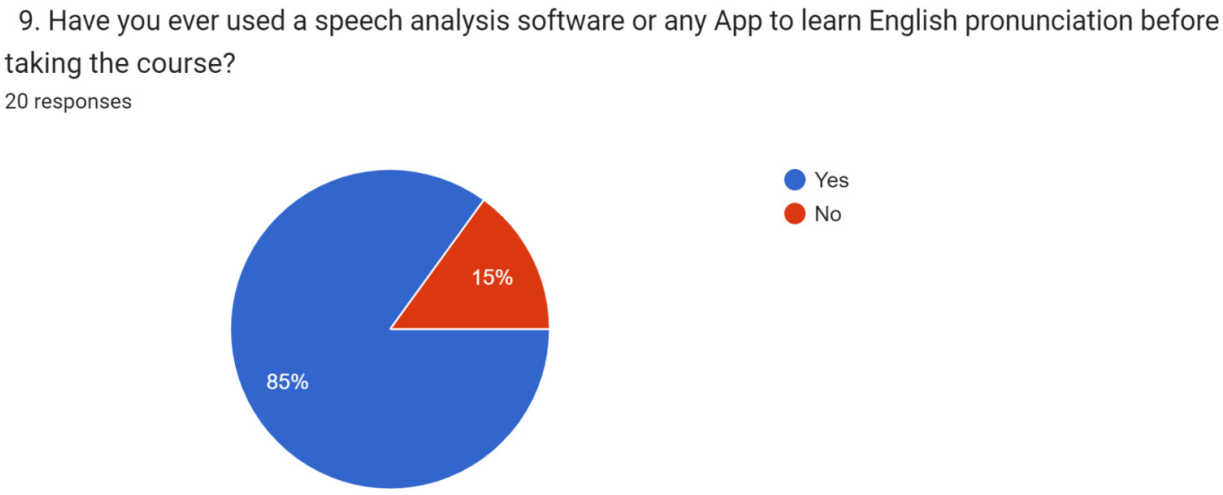

Figure 5. Students' responses to Question 9

Respondents who indicated using language apps for pronunciation learning were presented with options to select in Question 10. These options were curated considering the free availability of apps, given that many participants in the study lacked the means to purchase apps and typically relied on free alternatives, despite the potential limitations of those apps compared to ones they pay for. The data revealed that the free language learning app *Duolingo*, despite its limited pronunciation content, was the most widely recognized among students (82.4%). Following was *ELSA Speak*, with a usage rate of 23.5% (Figure 6). About 15% of the students reported not using any apps or software to learn pronunciation before the course. This could be attributed

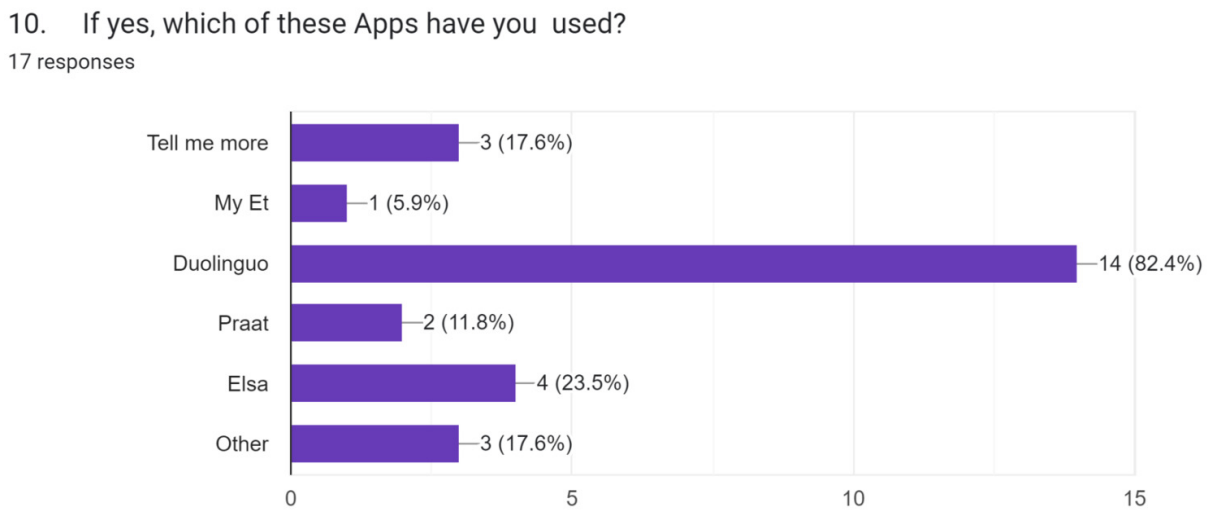

Figure 6. Students' responses to Question 10

to the lack of institutional support for the use of such technology, as students and parents must provide their own mobile phones or PCs. It could also be because some students prefer to learn pronunciation through authentic exposure to English through films and songs.

In Question 11, students were asked about "the advantages of using pronunciation software or language apps". They were provided with several options. The option that received the highest rate of responses was that those systems offer "*unlimited opportunities to review my speech*" (80%), followed by the option *"more time for practice"* (65%), *"autonomous learning"* (25%), *"technical feedback"* (15%), and *"neutrality"* (1%). As observed, the selected options revolve around the opportunities these systems offer for independent learning, allowing learners to review their speech without external pressure from teachers or peers in classroom settings, and enabling them to practice at their own pace and comfort level. Figure 7 displays students' responses to Question 11.

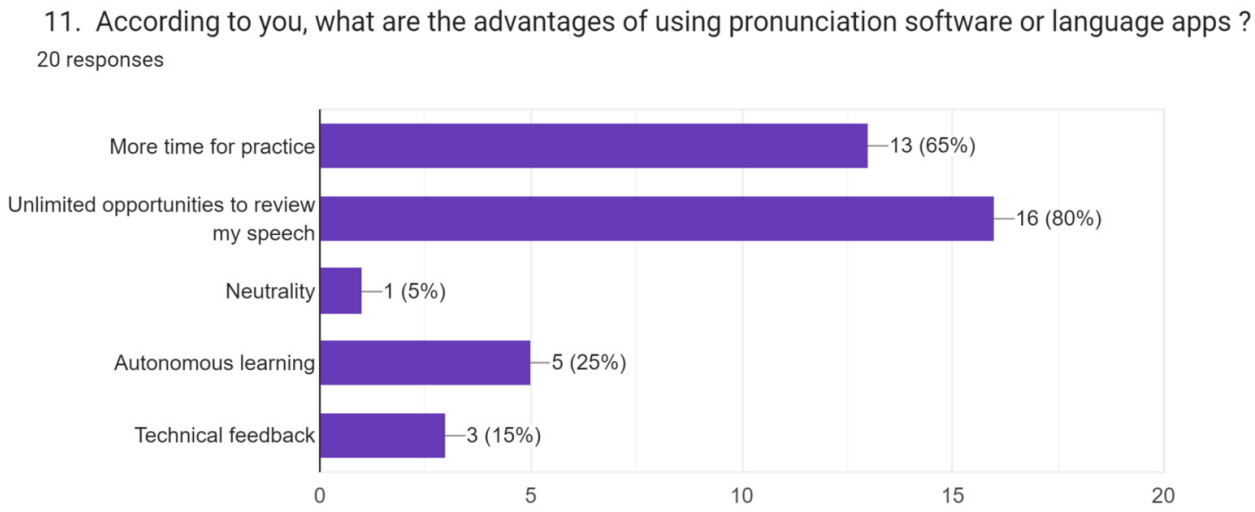

Figure 7. Students' responses to Question 11

While Questions 10 and 11 explored students' familiarity and usage of pronunciation apps and CAPT systems, Question 12 delved into their preferences regarding pronunciation learning. Participants were presented with several options. They were asked to select all that applied to them. The results revealed that 60% of the participants in the survey preferred learning pronunciation "with an app or software" (12 out of 20), while 50% indicated a preference for learning it "with the teacher in the classroom" (10 out of 20). Interestingly, among the 12 students who preferred learning pronunciation "with an app or software", some also expressed a preference for learning it "with the teacher in the classroom". This suggests that students have varied preferences for how they learn pronunciation, with 45% of respondents also choosing the option "independently through my exposure to English" (Figure 8).

After investigating students' preferences regarding the different methods of learning pronunciation, Question 13 focused on the option of preferring to learn pronunciation "with software or a language app" and inquired about the reasons behind this preference. Participants were presented with multiple options (Figure 9). The option "I can practice at my own pace" received the highest rate (about 79%), and the option "I can practice at any time I choose" was selected by 63.2% of the participants. The selection of these two options by the participants corroborates the results of Question 11, where students were asked to select the advantages of using pronunciation software and language apps. The options "unlimited opportunities to review my speech" and "more time for practice" received the highest percentages (80% and 65%). This indicates that the students' responses are consistent and emphasize independence, lack of pressure, and increased practice time as the primary criteria influencing their choice.

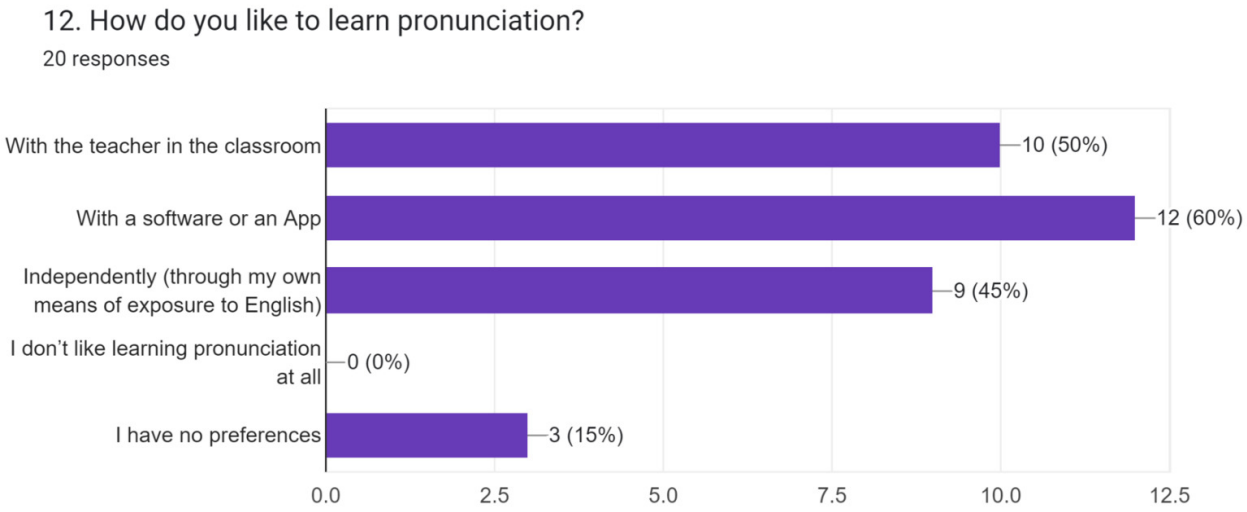

Figure 8. Students' responses to Question 12

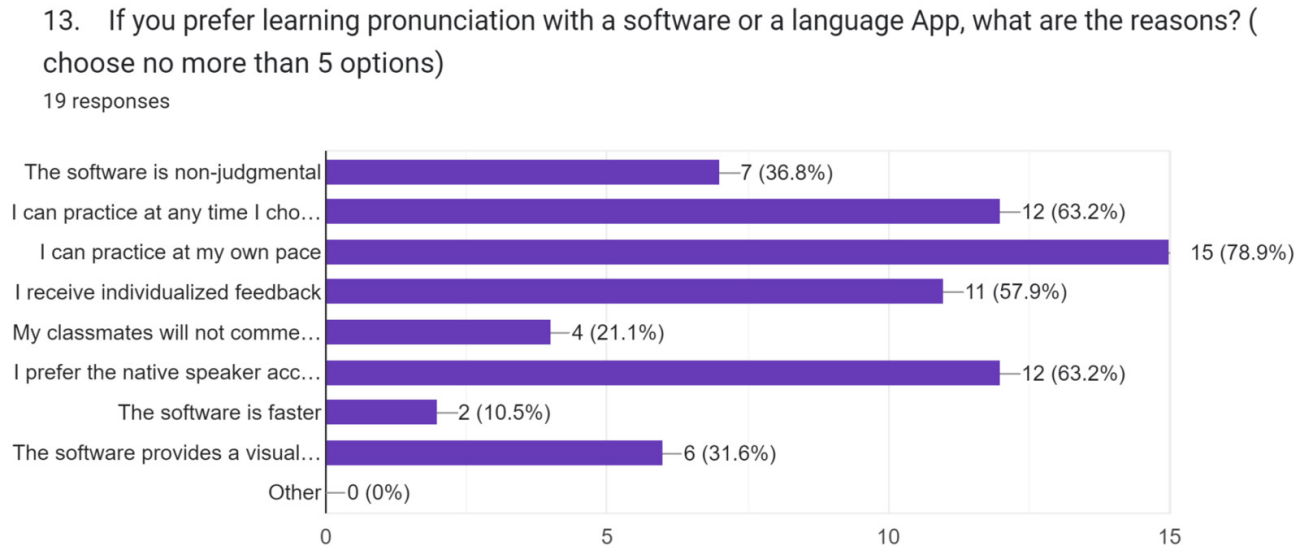

Figure 9. Students' responses to Question 13

The results also showed that the respondents preferred "the native speaker accent of the software" (63.2%) and the "individualized feedback" they received (57.9%). Approximately 37% found that the software's non-judgmental nature was advantageous, allowing them to avoid potential criticism about their pronunciation from classmates (21%). Additionally, about 31% believed that the visual feedback provided by the software or app is beneficial, while 10% perceived that learning pronunciation with such systems is faster, compared to traditional classroom methods (Figure 9).

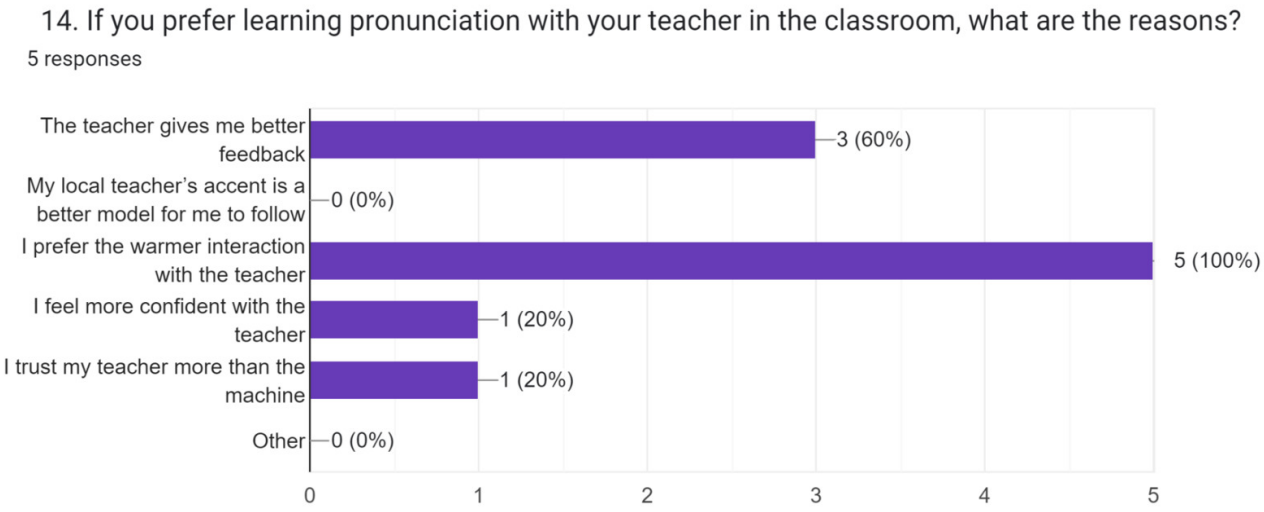

Figure 10. Students' responses to Question 14

Question 12 included the option "with the teacher in a classroom" for learning pronunciation. Therefore, respondents who selected this option were subsequently asked in Question 14 about the reasons behind their preference (Figure 10).

From Figure 10, it is evident that only five out of the 10 students who selected the option "with a teacher in the classroom" in Question 12 responded to Question 14. The primary reason they cited for this preference was that they "prefer the warm interaction with the teacher" (100%). Additionally, 60% of respondents mentioned that they believe the teacher provides more constructive feedback. One student expressed feeling more confident with the teacher, while another student mentioned trusting the teacher more than a machine. Interestingly, none of the participants indicated that their "local teacher accent" is a better model for them to follow. These findings suggest that while students value their teachers' warm interaction and feedback, they still prefer the native speaker accents offered by apps to their teachers" local accents.

The final inquiry in the survey (Question 15) was an open-ended question, delving into students' perspectives on the potential threats posed by AI to the humanistic essence of the student-teacher dynamic. This query garnered responses from 17 participants, reflecting a spectrum of length and depth in their insights. Through a thematic qualitative analysis, it was discerned that merely two out of these 17 respondents expressed concerns about AI-powered applications and software for pronunciation learning. Among these, one student highlighted the risk stemming from negative feedback and criticism delivered by teachers whose pronunciation does not serve as an exemplary model. She articulated this concern as follows:

> It may threaten. But students would opt for apps when there are negative feedback and critics and also and especially when the teachers' pronunciations are not even a model for the student to learn pronunciation from them.

The second student briefly conveyed her belief that AI could indeed pose a threat to the humanistic dimension of teacher-student relationships, emphasizing the need for smart utilization. However, she did not elaborate on the specific reasons or mechanisms through which this threat might manifest.

The remaining 15 responses uniformly rejected the notion that AI could jeopardize the humanistic essence of teacher-student relationships. They expressed their disagreement using phrases like *"No, I don't think so"*, *"I don't believe"*, *"I don't agree"*, and *"I disagree"*. Common themes in their responses included affirmations that AI-powered tools cannot undermine or supplant the human connection between teachers and students. They emphasized the indispensability of human interaction in the learning process and asserted that humans must interact with each other, not solely with machines. Moreover, they viewed AI as a facilitator of pronunciation learning and a supportive tool for teachers, enhancing efficiency and expediting the learning process. These responses reflect a nuanced understanding among students regarding the role of AI-powered technology tools in pronunciation learning and teaching. The response of one of the participants was articulated this way:

> I disagree because these apps lack the empathy, comprehension, and individualized guidance that human teachers offer. While they can supplement learning with extra resources and practice, they can't replace the special bond and assistance that teachers provide.

Another response was articulated as follows:

> I do not think so because I feel like human beings will always crave human interaction. I think people will want to practice their speaking skills with each other rather than a software. The role of teacher here is to come up with creative activities to encourage the students to be more active. The humanistic aspect will not fade away as long as there is something unique if teaching embedded in class.

The participants demonstrated a dual awareness: while they acknowledged the unique and irreplaceable nature of the teacher–student relationship, these digital natives also recognized the imperative for teachers to adapt to technology and master its effective use. This was expressed in one of the respondents' comments: *"No, AI would never threaten the teacher-student relationship. However, teachers should be up to speed with the new technology and use it in their classes."* Students seem to be cognizant of the truth

that although AI-powered systems and technology can offer an advanced experience in improving pronunciation, they cannot undertake the role of the human teacher. One of the participants reported to the teacher (outside the course time) that the *ELYSAI* tool struggled to understand her accent and she had to repeat several times, which was frustrating. The teacher's feedback and response in this case would be more effective as humans do not rely on voice traces only to identify and process meaning. Facial expressions and kinesics are also tools that help us process words and understand human speech. Yet, teachers have to be capable of using technology effectively (as suggested by the participants) and know its limitations to address them. The general conclusions of this action research are provided in the following section.

4. Discussion and conclusions

There are different types of teacher-action research and quite dissimilar methodologies to respond to the problems raised. In this study, which was carried out in the Tunisian EFL context, the teacher-researcher did not tackle a specific skill or offer a specific treatment to a group of students to remedy a proficiency problem. Rather, the teacher-researcher designed a course that aimed at preparing MA students to become efficient researchers and sought to raise their awareness of the important complementarity between several scientific disciplines. The ongoing discourse surrounding the potential risks posed by AI in the realm of education has motivated the teacher to design a course that bridges the humanities with scientific principles, with a particular focus on AI. The course is titled "Advanced Prosody" since it engages students with advanced aspects of phonetic science, which are acoustic phonetics and mainly the analysis of suprasegmental features. It also introduces them to CAPT systems and AI-powered pronunciation apps. Throughout the course, students explored the mutual exchange of knowledge and skills between the linguist, representing the humanities sector, and the speech technologist, embodying the realm of hard science. They developed an understanding that AI-powered pronunciation systems have foundations in experimental phonetics and speech science technology.

Their engagement with the course content was verified through an online questionnaire. The responses to the questions revealed a prevalent interest in the course among the majority of the students. Students appeared to be more interested in specific components of the course, such as the prosody component, learning pronunciation, and the teaching component, than in the course as a whole. This connection may inform their choice of specific components as research topics for their MA theses, which they are expected to write in the subsequent semester following the completion of the course. When asked about the course's contribution to raising their awareness about the interplay between phonetic science and computer science in the development of speech analysis software and pronunciation learning apps, students affirmed that the course contributed to their discovery and recognition of this collaboration and complementarity. Indeed, students

realized that research in experimental phonetics could provide speech technologists with fine-grained information on what to include as functionalities in a software or pronunciation application, depending on the L1 background of different speakers and the type of difficulties they may encounter. For instance, specific pronunciation apps could be developed for speakers of Arabic, focusing on their particular difficulties, such as vowel quantity and quality, in learning English pronunciation. This reinforces the significance of linguistics and humanities in both maintaining the functionality of these systems and propelling their advancement.

The second part of the questionnaire surveyed the prevailing beliefs and attitudes among Tunisian EFL learners regarding the utilization of AI-powered pronunciation applications and CAPT systems. The majority of the respondents reported a favorable view regarding the utilization of those systems, similar to EFL learners from other contexts (Arini et al., 2022). They advocated the autonomous learning opportunities these apps offer, the feedback, and the native speaker accent. However, they acknowledged their preference for learning pronunciation through various ways, including the classroom setting with the teacher. The others reported that they preferred the warm interaction with the teacher in pronunciation classes and think that the teacher provides more constructive feedback. Interestingly, the majority of the respondents were confident enough that AI-powered teaching and learning systems do not represent any threat to the warm interaction between teachers and students if used intelligently. However, most students believed that teachers should be able to cope with technological development and use it effectively as a teaching aid.

In most domains where AI technology is used, such as medicine, AI systems should never be fully autonomous because this may result in people losing control over these systems. Similarly, in the education sector, AI-powered systems should not be fully autonomous. Education should always be assisted by humans since learning is in itself a social activity par excellence. Previous research (Couper, 2003; Derwing & Munro, 2005; Lord, 2005; Saito & Lyster, 2012) has shown that explicit pronunciation instruction by the teacher in regular classroom settings improves learners' pronunciation. Therefore, students' use of pronunciation apps and CAPT systems should be guided by the teacher who would provide warmer and more authentic feedback. AI-powered apps lack the emotional intelligence and empathy the real teacher has. Furthermore, the human instructor would not only focus on fixing the pronunciation of students but would also work on boosting their self-confidence and motivation to learn.

Among the practical issues debated over the use of AI systems in various life domains is the future of work. Numerous professions face the threat of obsolescence due to the instantaneous growth in automation, driven by AI. Among these, the profession of pronunciation teacher could be at risk

unless instructors rapidly embrace technology training and prioritize refining their accents to serve as exemplary models for their students. In addition, stakeholders should be aware of these risks, provide the required training and assistance in technology use to teachers, and save the noble mission of the teacher.

The course has achieved its goals of raising students' awareness regarding the interplay between phonetic science and the front-line technologies of speech science and AI-driven pronunciation systems. Humanity students realized the important roles they could play as future teachers and researchers in the development of similar systems. Through the course and the survey addressed to these students in an EFL context characterized by limited resources and meager access to technology among learners and institutions, this teacher-action research would hopefully contribute to the debate about technology and AI use in the education sector, mainly in the field of teaching English with technology.

This teacher-action research has its limitations such as the researcher being the instructor and its resultant effect on students' responses, in addition to the small sample size of the population. These factors do not allow generalization of the findings. More research on larger samples at the undergraduate level, where there are more groups with larger numbers of students taught by different teachers, would perhaps provide deeper insights into these issues.

Acknowledgment

I would like to thank the participants who voluntarily agreed to share their experiences and screenshots.

Submitted: December 09, 2024 EET

REFERENCES

Arini, D. N., Hidayat, F., Winarti, A., & Rosalina, E. (2022). Artificial intelligence (AI)-based mobile learning in ELT for EFL learners: The implementation and learners' attitudes. *International Journal of Educational Studies in Social Sciences*, *2*(2), 88–95. https://doi.org/10.53402/ijesss.v2i2.40

Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer* (Version 6.0.37) [Computer software].

Bouchhioua, N. (2008a). Duration as a cue to stress and accent in Tunisian Arabic, Native English, and L2 English. *Proceedings of the Fourth International Conference on Speech Prosody*. https://doi.org/10.21437/SpeechProsody.2008-120

Bouchhioua, N. (2008b). The role of vowel quality in cuing stress and accent in Tunisian Arabic, Native English, and L2 English. *Proceedings of the Fourth International Conference on Speech Prosody*. https://doi.org/10.21437/SpeechProsody.2008-121

Cauldwell, R. (2012). *Cool Speech app*. http://www.speechinaction.org/cool-speech-2

Celce-Murcia, N., Brinton, M. D., & Goodwin, J. M. (1996). *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. Cambridge University Press.

Couper, G. (2003). The value of an explicit pronunciation syllabus in ESOL teaching. *Prospect*, *18*, 53–70.

Davis, K. H., Biddulph, R., & Balashek, S. (1952). Automatic recognition of spoken digits. *The Journal of the Acoustical Society of America*, *24*(6), 637–642. https://doi.org/10.1121/1.1906946

Delmonte, R., Petrea, M., & Bacalu, C. (2004, January 9). *SLIM: A Prosodic Model for Learning Activities in a Foreign Language*. http://www.byron.cgm.unive.it

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

Forgie, J. W., & Forgie, C. D. (1959). Results obtained from a vowel recognition computer program. *The Journal of the Acoustical Society of America*, *31*(11), 1480–1489. https://doi.org/10.1121/1.1907653

Ghazali, S., & Bouchhioua, N. (2003). The learning of English prosodic structures by speakers of Tunisian Arabic: Word stress and weak forms. In *Proceedings of the 15th ICPhS* (pp. 961–964).

Kawai, G., & Hirose, K. (2000). Teaching the pronunciation of Japanese double-mora phonemes using speech recognition technology. *Speech Communication*, *30*, 131–143. https://doi.org/10.1016/S0167-6393(99)00041-2

Levy, M. (1997). *CALL: Context and Conceptualization*. Oxford University Press.

Lim, F. V., & Toh, W. (2024). Apps for English language learning: A systematic review. *Teaching English with Technology*, *24*(1), 79–98. https://doi.org/10.56297/FSYB3031/GAQR358

Lord, G. (2005). (How) Can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania*, *88*, 557–567. https://doi.org/10.2307/20063159

Olson, H. F., & Belar, H. (1956). Phonetic typewriter. *The Journal of the Acoustical Society of America*, *28*(6), 1072–1081. https://doi.org/10.1121/1.1908561

Peng, X., Chen, H., Wang, L., & Wang, H. (2018). Evaluating a 3-D virtual talking head on pronunciation learning. *International Journal of Human-Computer Studies*, *109*, 26–40. https://doi.org/10.1016/j.ijhc

Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer-Assisted Language Learning*, *12*(5), 427–440. https://doi.org/10.1076/call.12.5.427.5693

Rogerson-Revell, P. M. (2021). Computer-assisted pronunciation training (CAPT): Current issues and future directions. *RELC Journal*, *52*(1), 189–205. https://doi.org/10.1177/0033688220977406

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

Appendices

Appendix A

ASSIGNMENT 2

From the recording of your first assignment:

a) Select and extract the following words:

Permit/ comment/ contact/ object/ content/

b) Save them as .wav files (example: permit.wav)

c) Check the stressed syllables in those words in native English according to their grammatical categories.

d) Check if you have produced stress correctly on those words

e) Find how these same words are correctly pronounced in native English speech by checking them on an online pronunciation dictionary or from the CD accompanying our pronunciation book (page 42 L.1.6 in the CD)

f) Extract the target words from a native speaker's file and save them as WAV files in *Praat*.

g) Create text grids for your pronunciation of those words and create text grids for the native words

h) Create a table like the one below and fill it in with the values you obtained:

	My pronunciation of the stressed syllable		The native pronunciation of the stressed syllable	
Permit	Duration (in ms)		Duration (in ms)	
	F0 (in Hz)		F0 (in Hz)	
	Intensity (in dB)		Intensity(in dB)	

i) Finally, take screenshots of the text grid of each word and provide them with the table

j) Measure the duration and the formant values (F1 & F2) of the vowels in all the function words "for/ to/ of" in the text. Report their values in a table. What do you notice?

Appendix B

QUESTIONNAIRE

This questionnaire is addressed to you to survey your responses to and engagement with the optional course "Advanced Prosody". It is anonymous and your responses will be kept confidential and used only for research and educational purposes. It will take you about 5 min to answer the questions.

1. Are you

Male
Female

2. How old are you?

20-25
More than 25

3. Which languages do you speak?

Arabic
French
English

Other, please specify.

4. How did you find the Advanced Prosody course?

Interesting
Boring
Too difficult
Too easy

5. If your answer was “interesting”, what factors made it such for you:

The prosody component (stress, intonation, rhythm, juncture, etc.)
The acoustic-phonetic aspect (measurement of duration, intensity, F0, etc)
The use of Praat and its functionalities (recording, making text grids, etc)
The pronunciation learning and teaching aspect
The combination of all the aspects mentioned above

6. If your answer was " too difficult", please provide reasons that made it such for you

7. Did you know before taking the Advanced Prosody course that acoustic phonetics has a role in the development of computer systems such as Praat and other pronunciation-learning software?

Yes
No

8. If your answer to the question above is “No”, how did this course change your awareness about the interaction of phonetic science and computer science for the development of such systems?

9. Have you ever used a speech analysis software or any App to learn English pronunciation before taking the course?

Yes
No

10. If yes, which of these have ever used?

Tell me More
My ET
Duolingo
Praat
ELSA
Other, please specify.

11. According to you, what are the advantages of using pronunciation software or language apps?

More time for practice
Unlimited opportunities to review my speech
Neutrality
Autonomous learning
Technical feedback

12. How do you like to learn pronunciation?

With the teacher in the classroom
With software or an App
Independently (through my own means of exposure to English)
I don't like learning pronunciation at all
I have no preferences

13. If you prefer learning pronunciation with software, what are the reasons?

The software is non-judgmental
I can practice at any time I choose
I can practice at my own pace
I receive individualized feedback
My classmates will not comment on/criticize my pronunciation
I prefer the native speaker accent of the software/App
The software is faster
The software provides a visual display of my utterances
Other

14. If you prefer learning pronunciation with your teacher in the classroom, what are the reasons?

The teacher gives me better feedback
My local teacher's accent is a better model for me to follow
I have warmer interaction with the teacher

I feel more confident with the teacher
I trust my teacher more than the machine
Other

15. Taking into consideration that most language apps and pronunciation learning software are built through Artificial intelligence, do you believe that their use threatens the humanistic aspect of the teacher-student relationship?