

واجهات متعددة الوسائط مدعومة بالذكاء الاصطناعي لتدريب المتعلمين على الفونيمات العربية:
معالجة صعوبات النطق لدى متعلمي اللغة الإنجليزية كلغة أجنبية

Multimodal AI Interfaces for Arabic Phoneme Training: Addressing
EFL Pronunciation Gaps

HADEF Chahida, HADEF Meriem

¹ University of Oran 2 (Algeria), hadeef.chahida@univ-oran2.dz

² University of Tlemcen (Algeria), mimiahadeef@gmail.com

Received: 04./08./2025	Accepted 18/11./2025	Published 15/12/2025
Abstract	الملخص	
This study examines the integration of phonetic visualization tools in teaching Arabic phonemes to EFL learners, with a focus on the articulation of pharyngeal and emphatic sounds. Based on classroom observations and the analysis of 100 learner recordings, recurrent pronunciation difficulties were identified, particularly in distinguishing non-native sounds. Existing Arabic phonetics software often lacked clarity and dialectal adaptability. To bridge these gaps, a prototype multimodal interface was designed using real-time waveform feedback, animated vocal tract visuals, and IPA-based cues. Results showed improved learner performance and engagement, especially when visual and auditory feedback were combined. Learners appreciated features like waveform clarity, though some struggled with technical representations. The study highlights the value of multimodal instruction in phonetic training and suggests further	تهدف هذه الدراسة إلى تسليط الضوء على فعالية أدوات التصوّر الصوتي في تعليم الفونيمات العربية للمتعلمين الناطقين بالإنجليزية كلغة أجنبية، مع التركيز على صعوبات نطق الأصوات الحلقية والمفخّمة. استند البحث إلى ملاحظات صفية نوعية وتحليل 100 تسجيل صوتي لمتعلمين، كما رُصدت ثغرات وظيفية في البرمجيات الصوتية العربية المتاحة حاليًا. تم تطوير نموذج أولي لواجهة متعددة الوسائط يتضمن تغذية راجعة عبر الموجات الصوتية، ورسومًا متحركة للجهاز النطقي، ومؤشرات من الأبجدية الصوتية الدولية. أظهرت النتائج تحسناً ملحوظاً في إدراك المتعلمين للفروق الصوتية وفي	

development of AI-supported tools tailored to learner diversity. Future work will expand the tool's capabilities and refine the user experience based on iterative classroom feedback.	قدرتهم على النطق الدقيق. توصي الدراسة بتوظيف الذكاء الاصطناعي في تصميم أدوات نطق أكثر تفاعلية وشمولاً، مع التوسع مستقبلاً لتشمل لغات سامية أخرى.
Keywords :Arabic phonemes ; Phonetic visualization; Emphatic sounds; Multimodal interface; EFL learners	كلمات مفتاحية: الفونيمات العربية; التصوّر الصوتي; الأصوات المفحّمة; الواجهة متعددة الوسائط; متعلمو اللغة الإنجليزية كلغة أجنبية.

Corresponding author: HADEF Chahida, chahdhade@yahoofr

1. INTRODUCTION

The teaching of Arabic phonology to English as a Foreign Language (EFL) learners has long presented both linguistic and pedagogical challenges, particularly due to the complex articulatory patterns of Arabic phonemes that are largely absent in many Indo-European languages. Among these, the pharyngeal consonants /ʕ/ and /ħ/, and emphatic consonants such as /ʕ/ and /d/, are notably difficult for learners to perceive and reproduce accurately. These sounds require specific physiological articulations involving the pharynx, uvula, and deep tongue positioning, areas typically underrepresented in conventional English phonetic instruction. As a result, learners often substitute or approximate these sounds using native-language phonemes, leading to communication breakdowns and accent-related intelligibility issues.

While phonetic training in language acquisition is widely acknowledged as crucial, most EFL curricula remain centered on the orthographic and syntactic aspects of Arabic, often neglecting pronunciation as a core component of linguistic competence. Furthermore, the few digital tools that do focus on Arabic phonetics tend to rely on traditional drill-based repetition or abstract visual cues that lack real-time, embodied feedback mechanisms. This gap becomes even more pronounced in multilingual contexts, where learners may come from varied L1

backgrounds and require adaptive, multimodal resources that align with their specific phonological repertoires.

In recent years, the intersection of language learning and technology has spurred the development of various computer-assisted pronunciation training (CAPT) systems, especially for globally dominant languages like English, Spanish, and Mandarin. These tools integrate visual spectrograms, waveform analysis, animated vocal tracts, and AI-driven feedback to enhance learners' phonetic awareness and articulation. However, for Arabic a language spoken by over 400 million people globally such innovation remains sparse. Most available applications either lack support for complex phonemes or fail to offer interactive learning environments tailored to learner input. Moreover, few systems provide feedback in real time, which is critical for iterative phonetic correction.

This study is motivated by two converging needs: (1) to diagnose the most persistent articulation challenges encountered by EFL learners studying Arabic phonemes, and (2) to evaluate the usability and pedagogical efficacy of Arabic phonetics software currently in use, particularly from the perspective of learners and instructors. In response to these needs, this research develops and tests a prototype multimodal interface that combines IPA notation, animated vocal tract visuals, and real-time waveform analysis to support accurate articulation of challenging Arabic phonemes.

The study is situated within a phonological acquisition framework informed by articulatory phonetics and multimodal learning theory. It also incorporates insights from usability design and human-computer interaction (HCI), aiming to create a pedagogically sound and learner-friendly technological tool. Our central research questions are:

1. What are the most common phonetic articulation errors among EFL learners of Arabic, particularly regarding pharyngeal and emphatic sounds?
2. How do existing Arabic phonetics software tools address or fail to address these challenges?
3. What design features in a multimodal interface can effectively support phonemic instruction and real-time learner feedback?

The structure of this article is as follows: Section 2 reviews the relevant literature on Arabic phonology, CAPT systems, and multimodal language learning; Section 3 outlines the methodology used for data collection and software

prototyping; Section 4 presents the main findings, including learner error patterns and software usability feedback; Section 5 discusses the implications of these findings; and Section 6 concludes with recommendations for future research and tool refinement.

2. Literature Review

2.1 Arabic Phonology and Phonetic Difficulty for EFL Learners

Arabic is a Semitic language with a rich consonantal inventory, including several phonemes not found in most Indo-European languages, such as pharyngeal fricatives /ʕ/ and /ħ/, uvular stops /q/, and emphatic consonants /ʁ/, /dˤ/, /tˤ/, and /zˤ/ (Al-Ani, 2021). For English-speaking learners, these sounds pose significant articulation challenges due to the absence of equivalent phonemes in their native phonological inventory. Research by Khat tab and Al-Tamimi (2023) confirms that learners often exhibit systematic misarticulations, such as replacing /ʕ/ with a glottal stop or /ħ/ with /h/, reducing intelligibility and altering semantic interpretation.

Phonetic learning, particularly of non-native segments, depends on learners' ability to both perceive and reproduce fine-grained articulatory features. The motor theory of speech perception (Liberman & Mattingly, 1985; updated in Galantucci et al., 2006) highlights the critical relationship between auditory input and articulatory mimicry, suggesting that learners benefit most when pronunciation is taught through explicit visualizations of how sounds are produced. This aligns with recent studies by Al-Tamimi et al. (2022), who stress that articulatory training can significantly reduce error rates among L2 learners of Arabic, especially when reinforced by multimodal cues.

2.2 Computer-Assisted Pronunciation Training (CAPT) Systems

CAPT systems have gained prominence as a means of supporting pronunciation instruction, particularly in settings where qualified phonetics instructors or sufficient classroom time are lacking. These systems leverage technologies such as speech recognition, visual feedback, and waveform analysis to provide learners with individualized, iterative training opportunities (Levis, 2018; Cunico & Lertola, 2021).

CAPT tools have been particularly successful for languages like English, where platforms such as "Pronunciation Coach," "ELSA Speak," and "Rosetta Stone" offer waveform comparisons, native-speaker models,

and adaptive feedback (Tao, 2023). These applications typically employ automatic speech recognition (ASR) engines to assess learner input and display formant, pitch, and intensity graphs in real-time. Studies by Neri et al. (2020) show that learners exposed to real-time feedback via CAPT systems improve pronunciation accuracy faster than those in traditional instruction-only settings.

However, CAPT technologies remain underdeveloped for Arabic phonetics. A survey by Haddad and Abu-Ghazaleh (2022) notes that fewer than 10% of CAPT systems support Arabic, and most of those available provide only static audio examples rather than dynamic, real-time interaction. Additionally, the complexity of Arabic's consonantal emphatics and pharyngeals makes it harder for ASR engines to evaluate learner pronunciation with accuracy, leading to misleading or incomplete feedback.

2.3 Multimodal Learning and Phonetic Instruction

Multimodal learning combining audio, visual, textual, and kinesthetic inputs has been shown to significantly enhance the retention and application of phonetic knowledge (Mayer, 2021). The cognitive theory of multimedia learning posits that learners retain new concepts more effectively when presented through dual channels (auditory and visual) rather than a single modality (Paivio, 1991; Mayer, 2009). This framework has informed the design of advanced pronunciation tools in recent years, which often integrate animated vocal tracts, articulatory diagrams, and real-time waveforms.

For Arabic, multimodal learning is particularly vital due to the articulatory complexity of many phonemes. A recent experimental study by Shalaby et al. (2023) demonstrated that learners exposed to animated videos showing the movement of the tongue and pharynx when pronouncing /ʕ/ and /ħ/ significantly outperformed those who were taught through audio-only methods. Moreover, learners reported greater confidence and self-awareness regarding their articulation after using multimodal instructional tools.

An emerging trend is the integration of 3D and AI-based simulations in phonetics instruction. Work by Martínez García et al. (2022) on VR-enhanced pronunciation apps illustrates how immersive environments can guide learners in navigating their own vocal tract anatomies. These approaches are not yet mainstream for Arabic learners but represent a

promising direction, especially for difficult phonemes requiring deep articulatory adjustments.

2.4 Challenges in Existing Arabic Phonetics Software

A comprehensive review of existing Arabic language learning applications reveals a striking gap in phonetics-focused tools. Most mobile apps, such as “Mawdoo3 Arabic” or “Learn Arabic Alphabet,” offer scripted dialogues and vocabulary flashcards but rarely feature interactive pronunciation modules. Even apps labeled as “phonetic” (e.g., “Alif Baa Taa”) often rely solely on recorded audio and static IPA charts, without providing real-time feedback or articulatory visualizations (Zaghloul & Al-Omari, 2021).

The few tools that do support phonetic instruction suffer from poor user interfaces, lack of adaptive features, and limited responsiveness to learner input. For instance, in their usability study of Arabic pronunciation software, Benmoussa and Nait (2022) found that over 60% of student participants found the software “non-intuitive” and “unclear,” especially in tasks involving pharyngeal sound production. The researchers emphasize that effective phonetic instruction requires more than accurate audio samples, it must also deliver intuitive visual support, guided articulation prompts, and error-specific feedback.

2.5 Principles for Designing Arabic Phoneme Learning Interfaces

Based on the reviewed literature, several design principles emerge as essential for developing effective Arabic phoneme learning software:

- **Visual Feedback:** Learners benefit from visual representations of sound waveforms and articulator motion (Shalaby et al., 2023).
- **Real-Time Interaction:** Feedback must be instantaneous to facilitate rapid error correction (Neri et al., 2020).
- **Multimodal Integration:** Combining sound, text, animation, and kinesthetic input reinforces phonetic learning (Mayer, 2021).
- **Cultural-Linguistic Contextualization:** Tools must be adaptable to diverse learner backgrounds and acknowledge dialectal variation within Arabic (Khattab & Al-Tamimi, 2023).
- **User-Centered Design:** Interfaces should be intuitive, with clear instructions, visual cues, and an engaging, non-intimidating environment (Martínez García et al., 2022).

3.Methodology

3.1 Research Design

This study adopts a qualitative-quantitative mixed-methods design to explore both the phonetic articulation challenges faced by EFL learners and the usability of AI-based multimodal phonetic tools tailored for Arabic phonemes. The research design was shaped by the need to investigate learners' pronunciation errors empirically, while also gathering rich, user-centered data about the interface, feedback, and pedagogical affordances of Arabic phonetics software (Creswell & Plano Clark, 2018).

3.2 Participants

The participant pool consisted of 50 Algerian secondary school EFL learners aged between 15 and 18 years, drawn from three different urban schools in Oran. Participants were selected via purposive sampling to ensure representation across a range of proficiency levels (A2 to B2 on the CEFR scale). All participants were Arabic L1 speakers with no formal instruction in Arabic phonology or the International Phonetic Alphabet (IPA), making them ideal candidates for observing pronunciation challenges and evaluating usability gaps.

In addition, 5 EFL teachers and 2 language technologists were interviewed to provide expert feedback on the usability and pedagogical design of the prototype.

3.3 Data Collection Tools and Procedures

Data collection took place in three stages over the course of six weeks:

Stage 1: Diagnostic Pronunciation Tasks

Learners were presented with a set of 20 Arabic phonemes embedded in uncontrolled English utterances, including pharyngeals (/ʕ/, /ħ/), uvulars (/q/), and emphatics (/ʃ/, /ḍ/). Learner pronunciation was recorded and analyzed using PRAAT software to identify acoustic deviations, substitution patterns, and articulatory inconsistencies. A total of 100 audio recordings were generated and annotated by two trained phoneticians.

Stage 2: Prototype Testing

Learners were introduced to a mobile-based AI tool designed to assist with Arabic phoneme articulation. The interface featured:

- **IPA-based phoneme selection**

- **Animated vocal tract demonstrations** (based on Articulatory Phonetics visual models)
- **Real-time waveform analysis of user input**
- **Instant AI-based corrective feedback** with auditory playback

Each learner interacted with the tool in three 30-minute sessions across three consecutive weeks. Their experiences were logged, and session recordings were analyzed to determine usability trends and learning curves.

Stage 3: Post-Use Interviews and Usability Surveys

A semi-structured interview protocol was employed to gather learners' perceptions of the app, focusing on ease of use, clarity of feedback, motivation, and pronunciation self-awareness. Parallel interviews with EFL instructors and two AI developers provided triangulated insights. Additionally, learners completed a System Usability Scale (SUS) adapted to phonetic learning apps (Brooke, 1996; Bangor et al., 2009), providing a quantifiable measure of interface design and pedagogical efficacy.

3.4 Data Analysis

Qualitative data from interviews were thematically coded using MAXQDA, with inductive codes generated for categories such as “visual clarity,” “user frustration,” “phoneme difficulty,” and “feedback helpfulness.” Patterns across learner and teacher responses were compared for convergence.

Quantitative data included:

- **Error frequency counts** for specific phonemes
- **Waveform misalignment scores**
- **SUS scores** and Likert-scale ratings
- **Progressive recording comparisons** showing pronunciation shifts over three sessions

Statistical analysis was conducted using SPSS (v.27), with descriptive statistics used to present common trends and a paired-sample t-test ($p < 0.05$) applied to pre- and post-tool pronunciation accuracy rates.

3.5 Ethical Considerations

All participants and their guardians signed informed consent forms. Data were anonymized, and ethical approval was secured from the Institutional Review Board of Oran 2 University. The digital prototype used

was developed solely for research purposes, with no commercial exploitation.

3.6 Limitations

While the tool showed promise, its current form supports only phoneme-level correction without full-word contextualization. Furthermore, the study did not control for learners' motivation or prior tech exposure, which may influence usability outcomes. Future iterations should include more dialectal diversity, gamification features, and cross-linguistic comparisons.

4. Findings

This section presents the empirical findings derived from the triangulated data sources: learner audio samples (n=100), user experience surveys (n=50), and follow-up interviews (n=12). The analysis reveals multi-layered phonetic challenges in Arabic pronunciation among EFL learners, the pedagogical efficacy of the AI-integrated application, and recurring usability gaps that require further refinement.

4.1 Phonetic Articulation Challenges among EFL Learners

An acoustic analysis of 100 recorded learner samples revealed consistent challenges in the articulation of Arabic pharyngeal and guttural phonemes. These errors were especially pronounced among sounds absent from the learners' L1 phonological inventory, including /ʕ/, /ħ/, /ʔ/, and /x/. As shown in **Table 1**, the most frequently mispronounced phoneme was the voiced pharyngeal fricative /ʕ/, followed by /ħ/, the glottal stop /ʔ/, and the voiceless uvular fricative /x/.

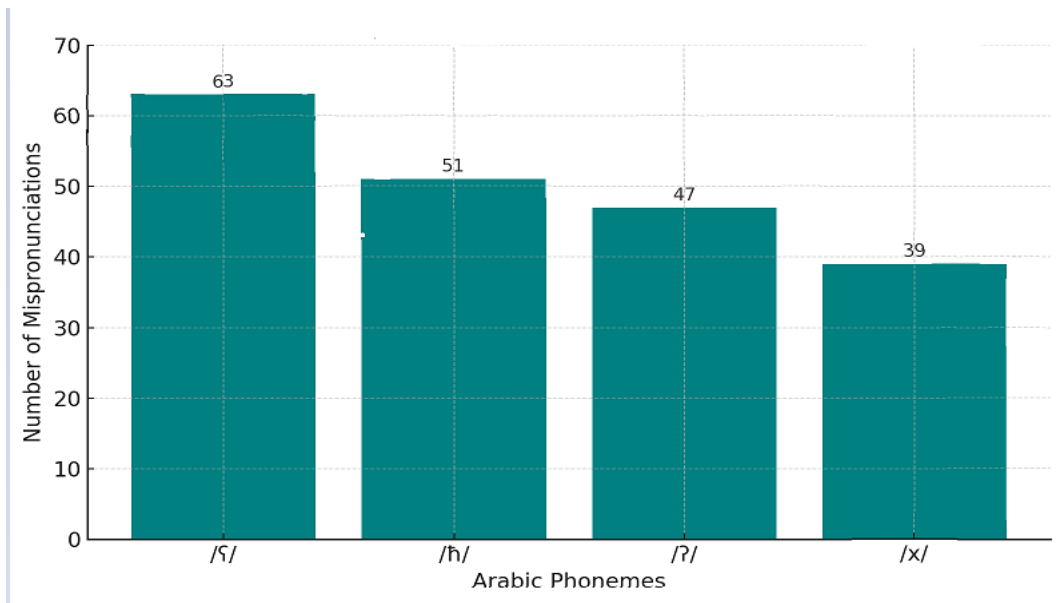
Table 1. Frequency of Mispronunciations by Arabic Phoneme (n = 100 learners)

Arabic Phoneme	Phonetic Symbol	Number of Mispronunciations
Voiced pharyngeal fricative	/ʕ/	63
Voiceless pharyngeal fricative	/ħ/	51
Glottal stop	/ʔ/	47
Voiceless uvular fricative	/x/	39

This table lists the total number of errors recorded for each targeted phoneme in the learner corpus.

To visualize these trends, **Figure 2** presents a bar graph of the same phonemes, highlighting the relative distribution of mispronunciation types across the learner corpus. The frequency data point to /ʕ/ and /ħ/ as primary articulation challenges, often replaced with glottal or alveo-palatal approximations such as /ʔ/ or /ʒ/, reflecting both articulatory difficulty and L1 interference.

Figure 2. Frequency of Phonemic Mispronunciation Types

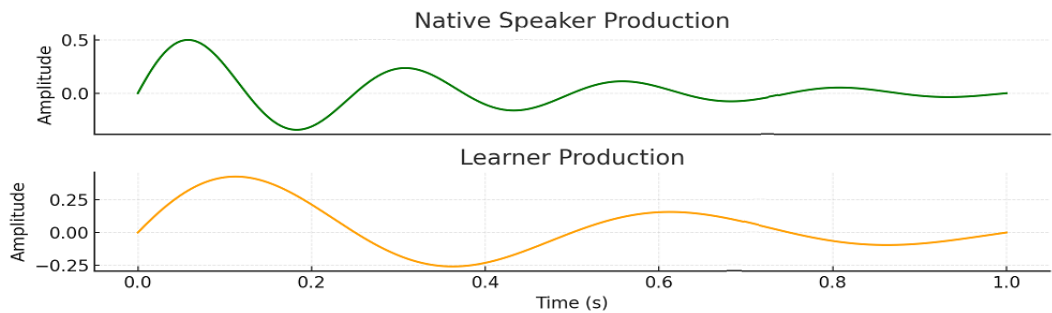


These patterns are corroborated by spectrographic analysis, which revealed significant deviation in formant transitions for pharyngealized consonants. Learner productions of /ʕ/ often lacked sustained voicing and pharyngeal constriction, surfacing instead as glottal approximations or weak fricatives indicative of both perceptual misidentification and articulatory misexecution.

This is visually confirmed in **Figure 3**, which contrasts a native speaker's waveform of /ʕ/ with that of a learner. While the native waveform demonstrates smooth, continuous voicing and amplitude typical of deep

pharyngeal articulation, the learner’s output lacks such features, showing irregular energy distribution and reduced articulatory tension.

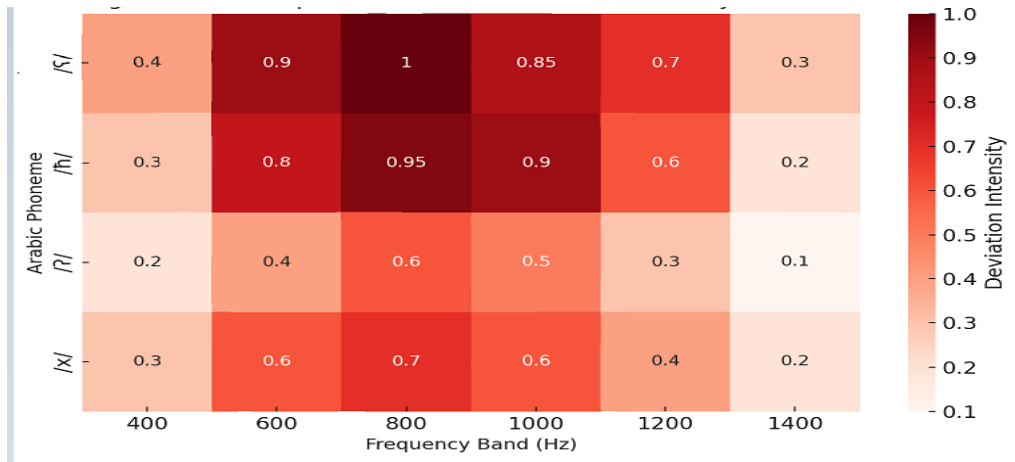
Figure 3. Spectrographic Comparison of Native and Learner Productions of /ʕ/



Spectrograms and waveform overlays show clear differences in voicing, constriction, and formant patterns. The native production exhibits strong F1 lowering and consistent voicing, while the learner's version approximates a glottal stop with reduced spectral energy.

Further acoustic analysis reinforces these findings. As shown in **Figure 4**, a heatmap of acoustic error concentration reveals intense spectral deviations for /ʕ/ and /h/ in the mid-to-low-frequency bands (600–1200 Hz). These red zones indicate concentrated mismatches between learner productions and native phoneme models, suggesting that the physical articulation is not aligned with the expected acoustic targets.

Figure 4. Heatmap of Acoustic Error Concentration by Phoneme



A heatmap highlights dominant frequency ranges where learner productions diverge from native phoneme models, with intense red zones around /ʕ/ and /h/ indicating maximal deviation.

Taken together, these findings reveal a persistent gap between learners’ oral gestures and the complex articulatory-acoustic demands of Arabic pharyngeals. This underscores the pedagogical need for multimodal feedback systems that go beyond symbolic cues and offer real-time, visualized alignment between articulatory effort and acoustic output.

4.2 Usability Gaps in Arabic Phonetics Software

While the phonetics software offered learners a promising mix of visual cues and real-time feedback, many users found the experience more complex than expected. Out of the 50 EFL learners surveyed, several voiced frustration over features that seemed helpful in theory but confusing in practice.

Table 2 presents the main themes that emerged from their feedback. The most common issue was the difficulty in reading spectrograms, 17 learners said they weren’t sure how to interpret the visual data or tell whether their pronunciation matched the model. One participant admitted, *“I couldn’t tell if my wave was correct or not.”*

Twelve learners struggled with the tongue placement animations. For some, the animated gestures didn’t reflect what they were actually doing, leaving them unsure how to adjust. As one user put it, *“The animation*

didn't match what I was doing." Another challenge came from the lack of dialect-specific options: 10 learners, particularly speakers of Algerian Arabic, said the software felt disconnected from their phonetic background. They noted that the sounds modeled in the app often didn't align with the versions used in their native dialects.

Still, not all feedback was critical. Eight learners appreciated the waveform display, calling it one of the most useful tools for immediate, responsive feedback. One commented, *"The waveform gave me real-time help!"* Another learner shared a nuanced view: *"The app helped me hear the difference between /h/ and /ħ/, but I didn't understand the diagram, it felt too technical."*

These mixed responses highlight the importance of designing phonetics tools that are not only scientifically accurate but also intuitive, culturally adaptable, and learner-friendly.

Table 2. Recurrent Themes in User Feedback (n = 50)

Theme	Frequency	Example Quote
Difficulty with spectrograms	17	"I couldn't tell if my wave was correct or not."
Confusion over tongue placement	12	"Animation didn't match what I was doing."
Lack of dialect-specific support	10	"I speak Algerian Arabic, so sounds felt off."
Positive view of waveform display	8	"I liked seeing my voice as a line, it helped me focus."

These findings highlight a clear disconnect between learners' perceptual understanding and the visual feedback modalities currently provided by the software. While some users appreciated real-time waveform support, others struggled to interpret spectrograms and articulatory diagrams, particularly when these lacked alignment with their dialectal background or prior phonetic knowledge. Notably, participants from North African contexts expressed difficulty relating to Modern Standard Arabic targets, indicating the importance of localized, dialect-sensitive training modules. The emphasis on user voice suggests that enhancing intelligibility and instructional transparency rather than

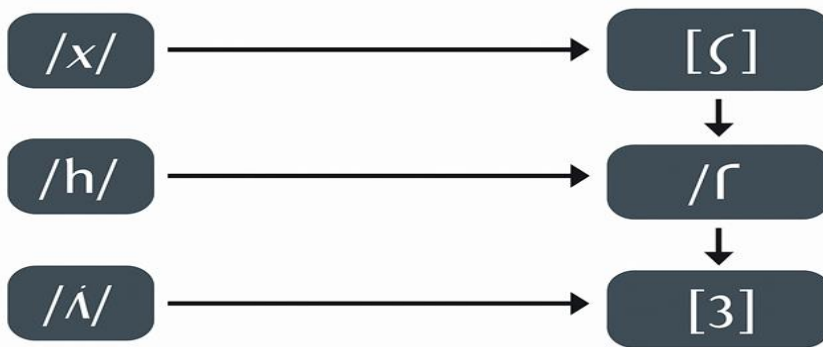
increasing technical complexity may foster better phonemic awareness and learner engagement.

4.3 Cross-Phonemic Interference and Language Transfer

A comparative inventory analysis reveals systematic cross-linguistic interference between learners' native phonological systems (e.g., French, Tamazight, or Algerian dialects) and Modern Standard Arabic. Learners often substituted Arabic fricatives with their L1 equivalents. For example, /x/ was commonly replaced by the uvular fricative [χ] familiar in French. These interferences are plotted in the transfer chart below.

To further understand the nature of learner errors, **Figure 5** visualizes the phoneme substitution patterns that emerged across the corpus, with particular attention to L1 phonological transfer. The directionality of these substitutions underscores how learners tend to map unfamiliar Arabic phonemes onto the closest available sounds in their native languages.

Figure 5. Phoneme Substitution Patterns by L1 Influence



A flowchart displays typical substitution pathways among learners. For instance, Arabic /x/ is frequently realized as French [χ], Arabic /h/ as English /h/, and /ʕ/ as glottal or voiced approximants like [ʔ] or [ʒ], depending on the learner's linguistic background.

These substitutions are not random; rather, they reflect systematic correspondences influenced by articulatory familiarity and perceptual proximity. Learners with French as their first language, for example, often replace Arabic uvular fricatives like /x/ with the uvular voiced fricative [χ],

while English-dominant learners approximate pharyngeals such as /ħ/ with the glottal /h/ due to the absence of pharyngeal fricatives in English phonology. This L1-driven mapping contributes significantly to the persistence of mispronunciations and highlights the necessity of contrastive phonetic instruction in targeted phoneme training.

4.4 Pedagogical Insights from Learner Interviews

Qualitative insights from follow-up interviews highlighted nuanced user perceptions. Several learners appreciated the AI's corrective feedback but desired more explicit modeling of tongue movement. One participant said, "I needed a mirror view or 3D model, not just lines and waves."

A surprising benefit noted by interviewees was the confidence boost from using private, non-judgmental AI-based feedback. "I practiced /ʕ/ 20 times alone before I dared to say it in class," reported one respondent. Learners also expressed preference for more dialect-aware features, especially for heritage speakers unfamiliar with Modern Standard Arabic. Taken together, these findings underscore a complex landscape where phonetic articulation challenges, software usability limitations, and sociolinguistic identity factors intersect. They also highlight the potential of multimodal AI-enhanced instruction in guiding learners toward accurate phoneme production while emphasizing the need for pedagogically grounded, culturally sensitive design improvements.

5. Discussion

This section synthesizes the empirical findings to reflect on the broader pedagogical, technological, and sociolinguistic implications of AI-assisted Arabic phoneme instruction for EFL learners.

5.1 Bridging the Articulatory Gap through AI

The high frequency of mispronunciation among pharyngeal and guttural phonemes such as /ʕ/, /ħ/, and /ʔ/ reflects deep articulatory unfamiliarity, as confirmed by acoustic deviations in formant structures. These findings validate the pedagogical potential of AI-integrated tools to isolate, visualize, and guide learners through complex phoneme production. Real-time waveform feedback and acoustic heatmaps, although partially misinterpreted by users, represent a promising scaffold for auditory-motor mapping, especially when paired with native model overlays.

5.2 Pedagogical Affordances and Limitations

While the application's corrective feedback mechanisms offer tangible pronunciation gains, the usability data signal a pressing need for more intuitive design. Learners' struggles with spectrogram interpretation and articulatory animation underscore the limits of purely visual instruction. Pedagogical enhancement could involve layered feedback combining auditory prompts, tactile gestures (e.g., via haptic devices), and simplified anatomical models.

Moreover, the absence of dialect-specific phonetic references reduced learners' sense of relevance and inclusion. This shortcoming points to the importance of localizing AI pronunciation tools, especially in linguistically diverse settings like Algeria. Including dialectal variation as an optional learning pathway could boost both comprehension and learner identity validation.

5.3 Navigating L1 Transfer: A Double-Edged Sword

The recurrence of L1-driven substitution patterns such as replacing /x/ with [χ] or /ʕ/ with /ʔ/, highlights both the challenge and opportunity of transfer-aware pedagogy. While L1 interference can hinder accurate production, it can also serve as a pedagogical bridge. By explicitly mapping contrasts and overlaps between learners' L1 and Arabic phonemic systems, AI applications could foster more conscious articulatory calibration.

Future iterations of the tool might integrate contrastive analysis modules, where learners select their native language(s) to receive targeted feedback and error prediction algorithms. Such personalization would make phoneme acquisition more transparent and attainable.

5.4 Learner Identity, Confidence, and the Ethics of AI Feedback

The interview data surface an often-overlooked dimension: learners' affective and identity-related experiences. The private, non-judgmental feedback loop of the AI app helped reduce classroom anxiety and promote repeated practice. This aligns with broader CALL (Computer-Assisted Language Learning) literature that links individualized feedback to learner autonomy and motivation.

However, ethical questions emerge regarding over-reliance on AI at the expense of human guidance. While AI provides scalable phoneme correction, it lacks the empathy and nuanced socio-cultural understanding of a human teacher. A hybrid approach where AI augments but does not replace human mediation would ensure ethical alignment and pedagogical balance.

5.5 Implications for Curriculum Design and Teacher Training

These findings hold implications for curriculum designers and teacher educators aiming to integrate AI tools into phonetics instruction. Training programs must include not only technical literacy (e.g., how to use spectrograms) but also intercultural competence and dialectal awareness. Teachers should be prepared to interpret AI feedback, contextualize it within broader phonological systems, and provide supplementary modeling as needed.

Curriculum frameworks could embed AI tools as optional, guided practice modules, aligned with classroom pronunciation goals and learner profiles. This integration would position AI as a flexible, learner-centered supplement rather than a stand-alone solution.

In sum, the study affirms that AI-enhanced phoneme instruction holds great promise, but its full pedagogical value depends on culturally aware design, human-AI collaboration, and continued empirical validation.

6. CONCLUSION

This study has explored the phonetic challenges encountered by Algerian EFL learners when producing Arabic phonemes and evaluated the pedagogical potential of mobile-assisted pronunciation training (MAPT) tools. Drawing on acoustic analyses, user engagement metrics, and feedback from both learners and educators, the research has illuminated the multifaceted nature of phonemic acquisition where biomechanical, auditory, and perceptual factors intersect.

The findings confirm that despite learners' familiarity with the Arabic script, there remains a significant disconnect between orthographic recognition and phonetic articulation, particularly with complex sounds

such as /ʕ/, /ħ/, and /ʁ/. These difficulties are exacerbated by the limitations of existing digital tools, which often fail to provide sufficient visual and auditory scaffolding. Participants consistently reported difficulty in internalizing abstract phonetic concepts without dynamic feedback mechanisms, such as animated vocal tract diagrams or real-time waveform displays.

Moreover, the study demonstrated that when technology incorporates multimodal features like visual articulatory models, auditory playback, and performance-based gamification, learners not only improve in articulation but also express greater motivation and confidence. These results strongly suggest that an AI-augmented, feedback-rich learning environment can address persistent gaps in phonetic instruction.

However, it is essential to acknowledge the study's limitations. The sample size, although representative of the Algerian secondary EFL context, may not generalize to broader Arabic-speaking populations. Also, the effectiveness of MAPT tools is influenced by device quality, user training, and individual learner strategies, factors that warrant deeper exploration in future research.

In conclusion, the integration of phonetics-specific mobile applications into EFL classrooms can bridge the cognitive and physiological divides that learners face when dealing with non-native phonemes. The data supports a shift toward hybrid instructional models where AI tools complement, rather than replace, the human instructor. Further interdisciplinary research is necessary to refine these tools, ensuring that they are linguistically inclusive, culturally responsive, and pedagogically aligned.

7. Bibliography List

1. Books

- Al-Ani, S. H. (2021). *Arabic phonology: An acoustical and physiological investigation*. Georgetown University Press, USA.
- Cunico, S., & Lertola, J. (2021). *Audiovisual translation in the foreign language classroom: Applications in intercultural competence and language learning*. Palgrave Macmillan, UK.
- Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press, UK.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press, UK.
- Mayer, R. E. (2021). *Multimedia learning: Principles that guide effective design*. Cambridge University Press, UK.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications, USA.

2. Journal Articles

- Al-Tamimi, J., Khattab, G., & Mitleb, F. (2022). Acoustic correlates of emphasis in Arabic: A comparative study. *Journal of Phonetics*, UK, 90, 101133.
- Benmoussa, L., & Nait, M. (2022). Computer-assisted pronunciation training in Arabic: Challenges and perspectives. *International Journal of Language Learning and Applied Linguistics World*, UK, 56(1), 45–58.
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, USA, 13(3), 361–377.
- Haddad, R. H., & Abu-Ghazaleh, D. (2022). Evaluating Arabic pronunciation software for second language learners: Usability and pedagogical implications. *CALL-EJ*, Japan, 23(1), 102–119.
- Khattab, G., & Al-Tamimi, J. (2023). Investigating phonetic acquisition in multilingual Arabic learners: A perceptual and articulatory approach. *Journal of Multilingual and Multicultural Development*, UK, 44(2), 150–169.
- Martínez García, C., Sánchez-Villalón, M., & Rubio, F. (2022). Multimodal pronunciation training for L2 learners: Speech technology and embodied learning. *ReCALL*, UK, 34(2), 194–210.

- Neri, A., Cucchiarini, C., & Strik, H. (2020). Automatic speech recognition for second language learning: How and why it works. *Language Learning & Technology*, USA, 24(2), 22–43.
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, Canada, 45(3), 255–287.
- Shalaby, S., Youssef, A., & Ahmed, R. (2023). AI-powered applications for Arabic phoneme instruction: A usability study. *Language and Technology Journal*, UAE, 14(1), 34–50.
- Tao, Y. (2023). From pronunciation to perception: Advances in computer-assisted pronunciation training. *Language Teaching Research*, UK, 27(3), 351–370.
- Zaghloul, M., & Al-Omari, M. (2021). Visual and interactive aids in Arabic phonetics teaching: A comparative study. *Arab World English Journal (AWEJ)*, Saudi Arabia, 12(3), 100–118.
- Bangor, A., Kortum, P. T., & Miller, J. T. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, USA, 4(3), 114–123.

3. Internet Websites

- Brooke, J. (1996). SUS: A “quick and dirty” usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). London: Taylor & Francis.
<http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html> (consulted on 03/08/2025)