



PERCEIVED EFFECTIVENESS OF MULTIMODAL AUDIO-VISUAL INSTRUCTION IN DEVELOPING ENGLISH PRONUNCIATION AMONG LIBYAN UNIVERSITY EFL LEARNERS

A Mixed-Methods Study at the University of Zawia

s.marimi@zu.edu.ly

<https://orcid.org/0009-0007-4619-9149>

Safia Marimi

English Language Department

Faculty of Arts University of Zawia

<https://orcid.org/0009-0007-4619-9149>

Received: 30/06/2026 - Revised: 18/06/2026 - Accepted: 28/06/2026 - Published: 1/8/2026

ABSTRACT

This study examines the perceived effectiveness of multimodal audio-visual instruction in supporting English pronunciation learning among EFL students at the University of Zawia, Libya. A mixed-methods design combined quantitative evidence from an eleven-item student questionnaire with qualitative evidence from semi-structured teacher interviews. Seventy-four valid questionnaires were analysed using frequencies, percentages, means, and standard deviations, while interviews with four English teachers were examined through thematic analysis. The questionnaire achieved acceptable internal consistency (Cronbach's $\alpha = .719$). The overall mean was 2.22 on a three-point scale, indicating a moderate level of perceived effectiveness. Students strongly associated audio and video with a more interesting classroom and with facilitated learning, but reported substantially less consistent classroom use by teachers. The interviews identified persistent challenges involving Arabic-language transfer, unfamiliar phonetic symbols, limited facilities, variable teacher readiness, and learner motivation. At the same time, teachers described audio-visual modelling, imitation, repetition, and out-of-class practice as useful mechanisms for raising phonological awareness and supporting clearer production. The findings reveal an implementation gap: learner acceptance and perceived pedagogical value are stronger than the systematic institutional use of audio-visual resources. The study recommends targeted teacher development, improved technical infrastructure, carefully designed pronunciation tasks, and future experimental research using objective pre- and post-measures of intelligibility, comprehensibility, segmental accuracy, and prosody.

Keywords: audio-visual aids; English pronunciation; multimodal learning; EFL; University of Zawia; Libya.

الملخص العربي

هدفت هذه الدراسة إلى تقييم الفاعلية المدركة للتدريس متعدد الوسائط القائم على الوسائل السمعية البصرية في دعم تعلم النطق الإنجليزي لدى طلبة اللغة الإنجليزية كلغة أجنبية في جامعة الزاوية. اعتمدت الدراسة منهجاً مختلطاً جمع بين البيانات الكمية المستمدة من استبانة مكونة من إحدى عشرة فقرة والبيانات النوعية الناتجة عن مقابلات شبه منظمة مع مدرسي اللغة الإنجليزية، وجرى تحليل أربع وسبعين استبانة صالحة باستخدام التكرارات والنسب المئوية والمتوسطات والانحرافات المعيارية، بينما حُللت مقابلات أربعة مدرسين تحليلًا موضوعيًا، بلغ معامل ألفا كرونباخ للاستبانة (0.719)، بما يشير إلى اتساق داخلي مقبول، أظهرت النتائج أن المستوى العام للفاعلية المدركة كان متوسطاً بمتوسط كلي بلغ (2.22 من 3)، مع ارتفاع واضح في إدراك الطلبة لدور الوسائل السمعية البصرية في تسهيل التعلم وزيادة تشويق الدرس، مقابل انخفاض انتظام استخدامها الفعلي داخل القاعات الدراسية، كما كشفت المقابلات عن صعوبات ترتبط بتأثير اللغة العربية، وتعقيد الرموز الصوتية، ونقص التجهيزات والتدريب، إلى جانب اتفاق المدرسين على أهمية النمذجة السمعية والبصرية والتقليد والتكرار في تحسين الوعي بالنطق والدافعية، وتخلص الدراسة إلى وجود فجوة تطبيقية بين قبول الطلبة لهذه الوسائل وبين دمجها المنهجي في التدريس، وتوصي بتطوير تدريب متخصص للمدرسين، وتحسين البنية التقنية، وتصميم مهام نطق تفاعلية مرتبطة بالأهداف التعليمية، وإجراء دراسات تجريبية تقيس التحسن الفعلي في الدقة والوضوح الصوتي.

الكلمات المفتاحية: الوسائل السمعية البصرية، النطق الإنجليزي، التعلم متعدد الوسائط، طلبة اللغة الإنجليزية، جامعة الزاوية، ليبيا.

1- Introduction

Pronunciation is a central component of spoken language because it mediates intelligibility, comprehensibility, interactional confidence, and the listener's ability to recover meaning from speech, Contemporary pronunciation pedagogy no longer treats the imitation of a native-speaker accent as its principal goal; rather, it prioritizes intelligible and comprehensible communication, strategic control of segmental and suprasegmental features, and the capacity to monitor one's own speech (Derwing & Munro, 2015; Saito & Plonsky, 2019), For university students learning English as a foreign language, these goals are especially important because oral participation, presentations, interviews, and professional communication depend on speech that listeners can process with reasonable ease.

Libyan EFL learners encounter a distinctive set of pronunciation challenges arising from the relationship between Arabic and English phonological systems, Differences in consonant inventories, vowel contrasts, syllable structure, stress, rhythm, and orthography may lead learners to substitute, delete, or insert sounds and to transfer Arabic timing and stress patterns into English, Such difficulties are not signs of inability; they are predictable consequences of cross-linguistic influence and limited exposure to varied models of spoken English, Effective instruction therefore needs to make sound contrasts perceptible, illustrate articulatory movement, provide repeated models, and allow learners to compare their production with accessible targets.

Audio-visual resources can respond to these needs by coordinating auditory input with visual information, Video can display lip, jaw, and tongue-related cues; captions can connect sound and orthography; waveform or spectrogram visualizations can make otherwise invisible acoustic properties visible; and mobile applications can provide repeated listening, recording, playback, and immediate feedback, These possibilities are consistent with the Cognitive Theory of Multimedia Learning, which proposes that well-designed combinations of verbal and pictorial information can support selection, organization, and integration when unnecessary cognitive load is controlled (Mayer, 2020), Recent reviews likewise indicate that mobile-assisted and computer-assisted pronunciation training generally produces beneficial outcomes, although effectiveness depends on feedback quality, task design, learner readiness, and teacher mediation (Amrate & Tsai, 2025; Metruk, 2024; Toyama & Hori, 2025).

The practical value of technology, however, cannot be assumed merely because devices are available, Multimedia becomes instructionally meaningful only when it is aligned with a clear pronunciation objective, embedded in active tasks, followed by guided practice, and supported by useful feedback. A classroom may contain a projector or smartphones and still rely on passive viewing that yields limited learning, Conversely, a simple mobile recording task can become powerful when students notice a target feature, rehearse it, receive feedback, revise their performance, and reflect on the difference, The present study therefore investigates not only whether students value audio-visual aids, but also whether such aids are used consistently and what teachers identify as the main enabling and constraining conditions.

1.1 Statement of the Problem

English pronunciation is widely recognized as difficult for Libyan university students, yet it is often taught with limited time, limited equipment, and insufficiently varied strategies, Students frequently experience recurring mispronunciations, while instructors may not always have a systematic repertoire for selecting and using audio-visual resources, The underlying problem is therefore twofold, First, learners need richer, more accessible models of speech and more opportunities for guided practice, Second, institutions need evidence about whether students and teachers perceive audio-visual instruction as useful, how frequently it is actually used, and what barriers prevent its consistent integration.

A further methodological issue concerns the interpretation of “effectiveness.” The research design measures student perceptions and teacher reports rather than experimentally measured changes in pronunciation performance. The term perceived effectiveness is therefore used to distinguish reported improvement from objectively demonstrated improvement. This distinction strengthens the scientific interpretation of the findings and prevents causal conclusions that the design cannot support.

1.2 Research Questions

- 1- To what extent are audio-visual aids used by teachers in pronunciation-related English classes and by students outside the classroom?
- 2- What are students’ preferences for audio, visual, and combined audio-visual resources, and how do they perceive their usefulness for language learning and pronunciation motivation?
- 3- What challenges, benefits, and contextual influences do teachers identify when using audio-visual resources for pronunciation instruction?
- 4- How do the quantitative and qualitative findings converge in explaining the implementation and perceived value of audio-visual pronunciation instruction at the University of Zawia?

1.3 Objectives of the Study

- 1- To determine the reported frequency of classroom and out-of-class audio-visual use.
- 2- To identify students’ modality preferences and their perceptions of facilitation, interest, and pronunciation motivation.
- 3- To examine teacher perspectives on pronunciation difficulties, instructional practices, resource constraints, and perceived learner improvement.
- 4- To integrate the two data strands and formulate evidence-based pedagogical and institutional recommendations.

1.4 Significance of the Study

The study is significant at contextual, pedagogical, methodological, and policy levels. Contextually, it contributes evidence from Libya, a setting that remains underrepresented in international research on technology-enhanced pronunciation instruction. Pedagogically, it identifies the specific forms of audio-visual support that learners value and the areas in which classroom implementation remains inconsistent. Methodologically, it combines student survey data with teacher interviews, enabling triangulation between reported learner experience and instructional reality. Institutionally, the findings can inform decisions concerning teacher development, language laboratories, classroom connectivity, resource selection, and the integration of guided mobile practice into university courses.

2- Conceptual and Theoretical Foundation

2.1 Pronunciation as Intelligible Communication

Pronunciation encompasses both segmental features, such as consonants and vowels, and suprasegmental features, such as lexical stress, sentence stress, rhythm, prominence, linking, and intonation. These systems jointly shape comprehensibility and intelligibility. A learner may produce every individual consonant accurately yet remain difficult to understand if stress and prominence are misplaced; similarly, fluent rhythm cannot fully compensate for a sound substitution that changes a high-frequency word. Balanced pedagogy therefore combines perception, controlled production, communicative transfer, and self-monitoring rather than reducing pronunciation to isolated repetition. These principles are consistent with established pronunciation pedagogy and recent syntheses of instructional effectiveness (Celce-Murcia et al., 2010; Lee et al., 2020; Thomson & Derwing, 2015).

For Arabic-speaking learners, instruction should be contrastive but not deterministic. Contrastive analysis can predict likely areas of difficulty, including English phonemes absent from Arabic varieties, distinctions represented differently in writing, consonant clusters, and the reduced vowels that characterize unstressed English syllables. Nonetheless, individual performance varies according to prior exposure, motivation, aptitude, dialect background, and opportunities for meaningful interaction,

The pedagogical goal is to use knowledge of Arabic-English differences to design targeted support without assuming that all learners will make identical errors. Such variation also reflects broader individual differences in language learning, including motivation and aptitude (Dörnyei, 2005).

2.2 Multimodal Learning and Cognitive Processing

Multimodal instruction presents complementary information through more than one mode. In pronunciation learning, auditory models provide timing, pitch, intensity, and spectral information, while visual representations can highlight articulatory position, stress placement, mouth movement, acoustic contours, or textual cues. Mayer's (2020) theory explains why the combination can be useful: learners process verbal and pictorial information through partially separate channels, each with limited capacity, and meaningful learning occurs when they actively select, organize, and integrate relevant information with prior knowledge.

The theory also warns against indiscriminate multimedia use. Decorative images, excessive text, long unsegmented videos, and simultaneous competing explanations can overload learners. Effective pronunciation materials should therefore be brief, focused, and synchronized. A useful sequence might involve listening to a short model, watching a close-up articulation, identifying the target contrast, repeating with recording, comparing the production, and applying the feature in a communicative sentence. The value lies in the instructional sequence rather than in the novelty of the device.

2.3 Computer- and Mobile-Assisted Pronunciation Training

Computer-assisted pronunciation training (CAPT) and mobile-assisted pronunciation training extend access to models and feedback beyond classroom time. Contemporary tools include recording and playback, visual displays of pitch and duration, automatic speech recognition, pronunciation dictionaries, captioned video, interactive minimal-pair practice, and social applications that support peer feedback. Systematic reviews report generally positive effects, but also substantial heterogeneity in interventions, outcome measures, and pedagogical quality (Amrate & Tsai, 2025; Ngo et al., 2024). Technology is most effective when feedback is interpretable, practice is repeated over time, and learners understand how the tool relates to communicative goals. Empirical work on pronunciation applications, automatic speech recognition, and mobile delivery further supports the importance of structured tasks and interpretable feedback (Fouz-González, 2020; García et al., 2020; Lan, 2022).

Mobile devices are particularly relevant in resource-constrained settings because many students possess smartphones even when language laboratories are limited. Metruk's (2024) systematic review of fifteen empirical studies found an overall beneficial influence of mobile learning on L2 pronunciation and generally positive learner attitudes. Yet the review also emphasized distraction, connectivity, inappropriate app selection, and inadequate pedagogical integration. These issues are directly relevant to Libyan universities, where mobile access can widen opportunity but cannot substitute for institutional planning and teacher guidance. Research on smartphone applications and app-based instruction likewise emphasizes that access must be accompanied by pedagogical guidance and appropriate task design (Kacatl & Klímová, 2019; Loewen et al., 2020; O'Brien et al., 2018).

2.4 Phenomenographic Sensitivity to Varied Experience

A phenomenographic perspective is useful for recognizing qualitatively different ways in which teachers and students may experience the same technology. One learner may view video as motivating and another as distracting; one teacher may regard mobile recording as feasible, while another may be constrained by connectivity or class size. Recognizing this variation prevents a one-size-fits-all interpretation and supports differentiated implementation.

3- Previous Studies

Recent research has increasingly moved from asking whether technology can support pronunciation to asking which combinations of mode, feedback, interaction, and pedagogy are most effective. Table 1 synthesizes studies that are particularly relevant to the present investigation.

PERCEIVED EFFECTIVENESS OF MULTIMODAL AUDIO-VISUAL INSTRUCTION
IN DEVELOPING ENGLISH PRONUNCIATION AMONG
LIBYAN UNIVERSITY EFL LEARNERS

Table 1, Selected recent studies on technology-enhanced pronunciation instruction

Study	Context and sample	Design / instruments	Main findings
Dai & Wu (2022)	10 Chinese university EFL learners; mobile peer-feedback task	Pre-, post-, delayed comprehensibility ratings; interaction recordings; interviews	Both feedback givers and receivers improved; peer interaction generated multiple feedback moves.
Hsieh, Yeh, & Chen (2025)	29 elementary EFL learners; robot and tangible-object system	Pronunciation tests, willingness-to-communicate questionnaire, video data	Pronunciation and willingness to communicate improved; learners showed repeated attempts and active help-seeking.
Metruk (2024)	Systematic review of 15 empirical studies (2015–2022)	PRISMA-based synthesis	Mobile learning generally benefited pronunciation and attracted positive attitudes, but pedagogical design and teacher preparedness remained crucial.
Benlaghrissi & Ouahidi (2024)	91 Moroccan secondary students; experimental and two control groups	Speaking pre/post tests and attitude survey	Mobile-assisted project-based learning significantly improved speaking, including pronunciation, and produced highly positive attitudes.
Ngo, Chen, & Lai (2024)	Meta-analysis of ASR-supported ESL/EFL pronunciation studies	Meta-analytic synthesis	Automatic speech recognition produced a positive overall effect, while design and outcome measures moderated interpretation.
Amrate & Tsai (2025)	Systematic review of empirical CAPT research	Pedagogy-focused review of CAPT studies	CAPT offers rich input, self-paced practice, and feedback, but research quality and pedagogical alignment vary.
Toyama & Hori (2025)	Mini-review of multimodal L2 pronunciation approaches	Review of gesture, speech visualization, and CAPT	Multimodal methods support awareness and engagement, but access, technical constraints, and classroom integration remain challenges.

The reviewed literature establishes a generally positive pattern but also reveals important qualifications. First, most strong claims about improvement come from experimental or quasi-experimental studies that directly assessed pronunciation, whereas perception surveys measure acceptability and reported benefit. Second, many studies were conducted in East Asian, European, or other non-Libyan contexts. Third, technology is frequently studied as a discrete intervention, while less

attention is given to the institutional gap between learner enthusiasm and routine classroom adoption, Finally, recent reviews call for mixed-methods work that explains not only outcomes but also teacher readiness, access, task design, and learner engagement.

3.1 Research Gap

There is limited empirical evidence concerning audio-visual pronunciation instruction in Libyan higher education, In particular, few studies have integrated student perceptions with teacher accounts to examine use inside and outside the classroom, modality preferences, L1-related challenges, infrastructure, and motivation within one explanatory framework, The present study addresses this contextual and methodological gap, It does not claim to establish causal pronunciation gains; instead, it offers a carefully bounded analysis of perceived effectiveness, implementation conditions, and the convergence between student and teacher evidence.

3.2 What Distinguishes the Current Study

- 1- It provides context-specific evidence from a Libyan public university rather than transferring conclusions uncritically from other educational systems.
- 2- It triangulates a student questionnaire with teacher interviews, allowing numerical patterns to be interpreted through classroom experience.
- 3- It identifies an implementation gap between positive learner perceptions and inconsistent teacher use.
- 4- It explicitly separates perceived effectiveness from objectively measured pronunciation improvement, thereby avoiding causal overstatement.
- 5- It translates the findings into a practical integration model involving focused input, guided noticing, recorded production, feedback, revision, and communicative transfer.

4- Methodology

4.1 Research Design

The study employed a mixed-methods design combining a descriptive quantitative survey with qualitative semi-structured interviews, The quantitative strand described student experiences, preferences, and perceptions across eleven items, The qualitative strand explored how teachers understood pronunciation difficulty, how they used audio-visual resources, what barriers they encountered, and whether they observed learner improvement, Integration occurred during interpretation, where the two strands were compared for convergence, complementarity, and discrepancy (Creswell & Plano Clark, 2018).

4.2 Setting, Population, and Sample

The research was conducted at the University of Zawia in Libya among students enrolled in English-language courses and teachers involved in English or phonetics instruction, Eighty questionnaires were distributed through random sampling procedures, Seventy-four usable responses were returned, yielding a valid response rate of 92.5%, Teacher participants were selected purposively because direct experience with pronunciation instruction was necessary for the interview questions, Although six teachers were initially approached, four completed the interviews and formed the qualitative sample.

Table 2, Study participants and data sources

Data source	Sample	Sampling	Purpose
Student questionnaire	80 distributed; 74 valid	Random sampling	Perceptions, use, preferences, motivation, perceived facilitation
Teacher interviews	4 completed interviews	Purposive sampling	Challenges, practices, facilities, cultural/L1 influences, reported improvement

PERCEIVED EFFECTIVENESS OF MULTIMODAL AUDIO-VISUAL INSTRUCTION
IN DEVELOPING ENGLISH PRONUNCIATION AMONG
LIBYAN UNIVERSITY EFL LEARNERS

Data source	Sample	Sampling	Purpose
Pilot study	10 students; additional pre- testing and review	Convenience / expert review	Clarity, feasibility, relevance, instrument refinement

4.3 Instruments

The student questionnaire contained eleven statements rated on a three-point frequency scale: Never (1), Sometimes (2), and Always (3). The items covered four domains: background experience of English classes (Items 1–3), audio-visual use inside and outside the classroom (Items 4–6), preferences for audio, visual, and combined resources (Items 7–9), and perceived facilitation and pronunciation motivation (Items 10–11). The three-point format was used to describe broad frequency patterns, although it offers less discrimination than a five- or seven-point scale.

The semi-structured interview guide asked teachers about the difficulty of teaching phonetics, reasons for learner pronunciation problems, frequency of audio-visual use, instructional and technical challenges, pedagogical usefulness, and observed improvement. Open questions allowed participants to explain context-specific experiences and provided qualitative depth that could not be captured by the fixed-response questionnaire.

4.4 Validity and Reliability

Face and content validity were supported through supervisor and expert review, pilot testing, and revision of unclear items. Construct relevance was addressed by aligning the questions with the study objectives and literature on pronunciation, multimedia, learner preference, and motivation. The final eleven-item questionnaire produced a Cronbach's alpha of .719. For an exploratory perception scale with a small number of heterogeneous items, this value indicates acceptable internal consistency, although it should not be interpreted as evidence that all items measure a single narrow construct.

4.5 Data Collection and Analysis

The questionnaire was administered between 5 and 30 March 2023, and teacher interviews were conducted during April 2023. Quantitative data were analysed in SPSS Version 25 using frequencies, percentages, arithmetic means, and standard deviations. Because the study did not include experimental groups, pre/post pronunciation tests, or inferential models, the analysis is descriptive. Qualitative interviews were transcribed and examined through thematic analysis involving familiarization, coding, theme development, comparison across participants, and interpretive synthesis (Braun & Clarke, 2006).

4.6 Ethical Considerations

Participation was voluntary. Participants were informed about the purpose of the study, the procedures, confidentiality, and their right to withdraw without academic consequences. Consent was obtained before data collection, recordings were used only for analysis, and identifying details were excluded from reporting. Institutional approval was obtained through the relevant postgraduate structures at the University of Zawia.

5- Results

5.1 Reliability and Overall Pattern

The questionnaire achieved Cronbach's $\alpha = .719$ across eleven items. Item means ranged from 1.84 to 2.43 on the three-point scale, and the overall mean was 2.22 (SD = .239). According to the adopted interpretation criterion, this result represents a moderate level. The pattern is more informative than the single overall mean: learner endorsement was highest for the pedagogical value of audio-visual resources, whereas teacher use inside the classroom was the weakest item. This contrast forms the central empirical finding of the study.

PERCEIVED EFFECTIVENESS OF MULTIMODAL AUDIO-VISUAL INSTRUCTION
IN DEVELOPING ENGLISH PRONUNCIATION AMONG
LIBYAN UNIVERSITY EFL LEARNERS

Table 3, Key quantitative results (N = 74)

Indicator	Mean	Never %	Sometimes %	Always %	Interpretation
Teacher uses audio-visual aids in class	1.84	33.8	48.6	17.6	Lowest mean; inconsistent classroom integration
Class is more interesting with audio/video	2.42	14.9	28.4	56.8	Strongest consistent endorsement of engagement
Uses audio-visual aids outside class	2.23	12.2	52.7	35.1	Substantial but mainly occasional autonomous use
Prefers audio aids	2.18	18.9	44.6	36.5	Moderate preference
Prefers visual aids	2.16	17.6	48.6	33.8	Moderate preference
Prefers combined audio-visual aids	2.39	5.4	50.0	44.6	Combined mode preferred over single modes
Audio-visual aids facilitate learning	2.43	4.1	48.6	47.3	Highest mean in the questionnaire
Audio-visual aids motivate pronunciation	2.34	6.8	52.7	40.5	Positive perceived motivational role

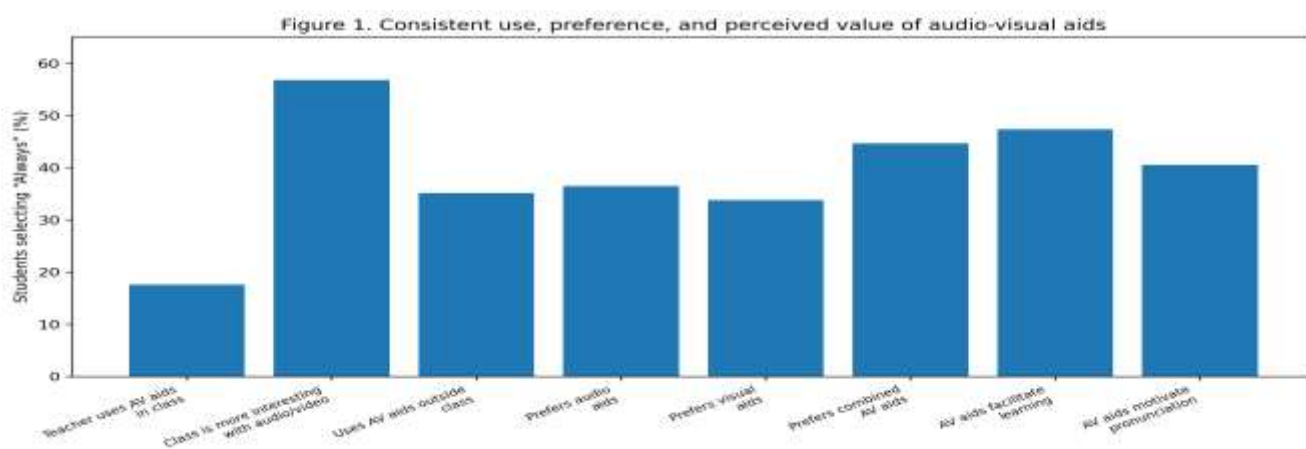


Figure 1, Percentage of students selecting "Always" for the main audio-visual indicators.

5.2 Pronunciation Difficulty and Classroom Climate

Seventy percent of respondents reported that they sometimes faced difficulty learning English pronunciation, and 24.3% reported always facing difficulty. Thus, 94.6% encountered difficulty at least

some of the time, The motivational climate was also uneven: 66.2% sometimes experienced a lack of encouragement to speak English and 20.3% always did so, Only 32.4% consistently described the classroom as enjoyable and interesting, These findings establish the need for interventions that address both phonological difficulty and affective engagement.

5.3 Use Inside and Outside the Classroom

Teacher use of audio-visual aids was the lowest-rated item, One third of students (33.8%) stated that teachers never used such aids, almost half (48.6%) reported occasional use, and only 17.6% reported consistent use, By contrast, 56.8% stated that audio and video always made the classroom more interesting, Out-of-class use was more common than regular teacher use: 35.1% always used audio-visual resources independently and 52.7% did so sometimes, This suggests that students may be compensating for limited classroom exposure through self-directed digital practice.

5.4 Learner Preferences

Combined audio-visual resources were preferred more strongly than either audio-only or visual-only materials, The “always” response was 44.6% for language laboratories and videos, compared with 36.5% for songs and recordings and 33.8% for handouts and slideware, Only 5.4% rejected combined audio-visual resources entirely, This pattern supports the argument that learners benefit from synchronized auditory and visual information rather than from isolated modalities.

5.5 Perceived Facilitation and Motivation

The highest mean (2.43) was recorded for the statement that audio-visual aids facilitate language learning, Nearly half of the respondents (47.3%) selected “always,” and another 48.6% selected “sometimes.” Likewise, 40.5% consistently perceived audio-visual resources as motivating them to improve pronunciation and 52.7% perceived this effect sometimes, The findings demonstrate broad acceptance, but because they are self-reports, they should be interpreted as perceived support and motivation rather than direct evidence of measured pronunciation gain.

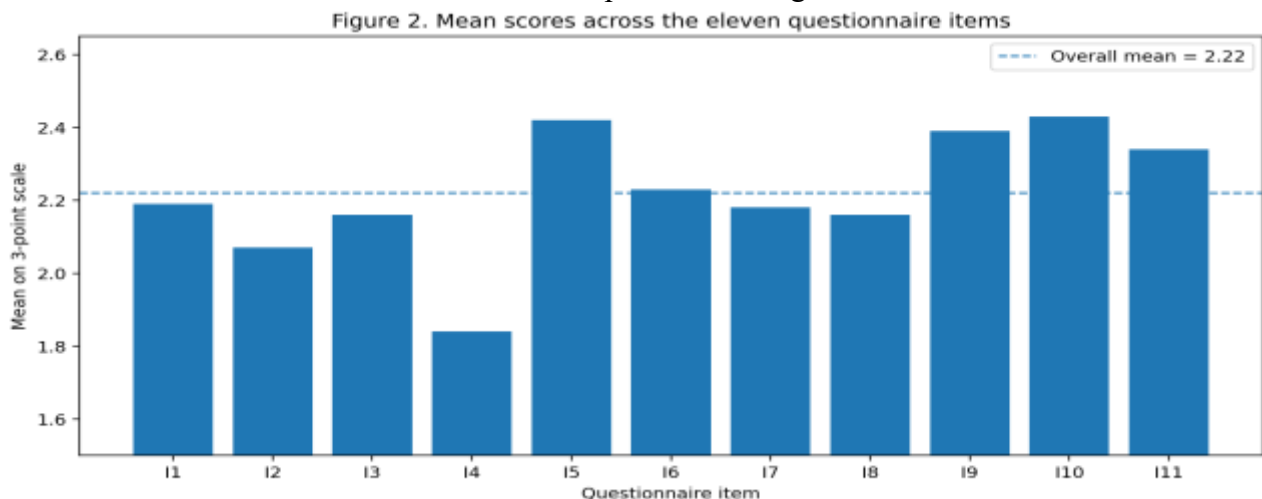


Figure 2, Questionnaire item means on the three-point scale; the dashed line indicates the overall mean.

5.6 Qualitative Themes

Table 4, Themes emerging from teacher interviews

Theme	Integrated interpretation
Teaching challenges in phonetics	Phonetic symbols and sound systems were perceived as unfamiliar and difficult for learners to remember.
Pronunciation difficulties for EFL learners	Teachers emphasized Arabic-language transfer, limited quality of models, and weaknesses in learner exposure and practice.
Usage of audio-visual aids	Practice ranged from non-use because of limited facilities to the use of board illustrations, videos, and other visual supports.

PERCEIVED EFFECTIVENESS OF MULTIMODAL AUDIO-VISUAL INSTRUCTION
IN DEVELOPING ENGLISH PRONUNCIATION AMONG
LIBYAN UNIVERSITY EFL LEARNERS

Theme	Integrated interpretation
Teaching with audio-visual aids	Teachers described the resources as helpful and interesting but identified equipment, access, and learner-specific problems.
Utility for pronunciation	Participants believed that well-used resources provide clearer models and enable imitation of intelligible production.
Reported pronunciation improvement	Teachers observed noticeable improvement among some students after exposure and repeated practice, but no objective test was administered.
Cultural and L1 factors	The Arabic phonological background shaped perception and production and required culturally and linguistically responsive explanations.
Student motivation	Engagement and willingness to practise were viewed as important conditions for improvement, especially outside class.

5.7 Mixed-Methods Integration

The two strands converge on three major points, First, pronunciation is perceived as difficult, and teachers attribute part of this difficulty to Arabic-English differences and unfamiliar phonetic concepts, Second, students strongly value audio-visual resources, while teachers acknowledge their usefulness, Third, implementation is inconsistent because access, training, and facilities vary, The qualitative evidence therefore explains the quantitative implementation gap: low teacher-use scores do not reflect rejection of the approach; they reflect practical constraints and uneven readiness.

There is also a productive tension between perceived benefit and evidence of learning, Students and teachers report facilitation, motivation, and improvement, but the design did not include objective pronunciation measures, This discrepancy does not invalidate the perceptions; it clarifies what they can establish, The study demonstrates acceptability, perceived usefulness, and contextual plausibility, while a future experiment is required to estimate actual learning effects.

6- Discussion

6.1 The Implementation Gap

The most important finding is the gap between high perceived value and low consistent classroom use, Students considered audio and video engaging and facilitative, yet only 17.6% reported that teachers always used audio-visual aids, This pattern corresponds with recent reviews showing that access to technology does not guarantee pedagogical integration, Teachers need confidence in selecting resources, aligning them with pronunciation targets, managing classroom time, and interpreting feedback (Amrate & Tsai, 2025; Metruk, 2024), Institutional investment should therefore combine equipment with professional development and curriculum design.

6.2 Why Combined Modalities Matter

The stronger preference for combined audio-visual materials is theoretically coherent, Pronunciation is simultaneously acoustic, articulatory, visual, temporal, and embodied, Learners need to hear a sound, notice how it differs from a familiar category, observe relevant articulatory cues, and practise producing it in meaningful language, Multimodal instruction supports this sequence by making hidden features more salient, Toyama and Hori (2025) similarly identify speech visualization, gesture, and CAPT as complementary approaches that can strengthen phonological awareness when carefully integrated.

However, the result should not be interpreted to mean that more media are always better, Audio-only work remains essential for developing perception without visual dependence, and visual displays can become confusing when learners lack guidance. A principled design alternates modes: visual support can introduce and clarify a feature, while later listening and communicative tasks test whether the learner can perceive and produce it without scaffolding.

6.3 Motivation, Autonomy, and Out-of-Class Practice

The substantial level of occasional and consistent out-of-class use indicates that students are willing to practise independently, This aligns with mobile-assisted pronunciation research showing that

portability, repeated access, and opportunities to record or exchange feedback can extend learning beyond limited class time (Dai & Wu, 2022; Metruk, 2024), The motivational benefit is particularly relevant because most respondents also reported some lack of encouragement to speak English in class.

Autonomy nevertheless requires structure, Unguided exposure to online videos may reinforce inaccurate interpretations, encourage imitation of unsuitable accents, or focus attention on entertainment rather than learning, Teachers can convert informal mobile use into purposeful practice by assigning brief weekly tasks: select a target feature, view a short model, record two attempts, compare them with a rubric, receive peer or teacher feedback, and submit a revised version, Such cycles preserve flexibility while introducing accountability and reflection.

6.4 Arabic-English Phonological Influence

The interview evidence confirms that L1 transfer is central to the local context, Effective materials should therefore include examples relevant to Arabic-speaking learners rather than relying exclusively on generic commercial applications, Targeted contrasts can address sound substitutions, consonant clusters, vowel length and quality, word stress, and reduced vowels, Visual articulatory cues and slowed high-quality models may be especially valuable when learners cannot easily perceive the difference between two English categories.

Culturally responsive instruction also requires a realistic target, The goal should be intelligible international communication, not the elimination of all traces of a Libyan accent, This orientation can reduce anxiety, increase willingness to communicate, and direct limited instructional time toward features that most strongly affect understanding.

6.5 Relationship to Recent Empirical Evidence

The positive perceptions in the current study are consistent with experimental findings elsewhere, Benlaghrissi and Ouahidi (2024) found that mobile-assisted project work improved speaking performance, including pronunciation, among Moroccan learners, Dai and Wu (2022) demonstrated that mobile peer feedback supported comprehensibility gains and rich interaction, Hsieh et al. (2025) reported improvement and repeated engagement in a multimodal robot-supported environment, At the synthesis level, Metruk (2024), Ngo et al. (2024), and Amrate and Tsai (2025) conclude that mobile and computer-assisted pronunciation training is generally beneficial, while cautioning that feedback design, pedagogical structure, and methodological quality determine the strength of the outcome.

The present study adds a complementary contribution, Rather than evaluating a single application, it maps the institutional conditions surrounding audio-visual instruction in a Libyan university, Its principal contribution is explanatory: students are receptive, teachers recognize the benefits, and independent use exists, but routine classroom adoption remains limited, This diagnosis provides a foundation for designing a subsequent controlled intervention.

7- Proposed Pedagogical Model for Audio-Visual Pronunciation Instruction

Based on the findings and contemporary research, the study proposes a six-stage model that can be implemented with basic classroom technology or smartphones, The model is designed to move learners from passive exposure to active, feedback-supported communication.

Table 5, Six-stage multimodal pronunciation cycle

Stage	Instructional action
1, Focus	Select one high-priority segmental or suprasegmental feature linked to intelligibility.
2, Multimodal noticing	Present a brief audio/video model with articulatory, caption, stress, or pitch cues; guide learners to notice the target.
3, Controlled production	Use minimal contrasts, words, phrases, and short sentences with immediate teacher or tool-supported feedback.
4, Record and compare	Students record their speech, compare it with a model or rubric, and identify one specific revision goal.

PERCEIVED EFFECTIVENESS OF MULTIMODAL AUDIO-VISUAL INSTRUCTION
IN DEVELOPING ENGLISH PRONUNCIATION AMONG
LIBYAN UNIVERSITY EFL LEARNERS

Stage	Instructional action
5, Interactive feedback	Peers and the teacher provide focused feedback on comprehensibility and the selected feature rather than on accent elimination.
6, Communicative transfer	Learners use the feature in a dialogue, presentation, role play, or project and reflect on intelligibility and confidence.

This sequence can be repeated in short cycles throughout a course, It also provides a framework for evaluating resources: a video or application is useful when it supports one or more stages of the cycle, not merely because it is technologically sophisticated, Teachers can begin with freely available tools and gradually incorporate speech visualization or automatic recognition as infrastructure improves.

8- Conclusions

The study concludes that audio-visual pronunciation instruction is broadly accepted by Libyan university EFL learners and is perceived as useful for facilitating learning, increasing classroom interest, and motivating pronunciation practice, The overall level of perceived effectiveness was moderate, but this average conceals a strong contrast: pedagogical value received high ratings, whereas consistent teacher use was limited, Students also reported substantial independent use outside class, indicating both demand and an opportunity for structured mobile learning.

Teacher interviews reinforced the quantitative findings, Participants identified genuine pronunciation challenges involving Arabic-language transfer, unfamiliar phonetic symbols, learner resistance, and restricted opportunities for high-quality modelling, They regarded audio-visual resources as helpful, but described uneven access and facilities, The combined evidence therefore indicates that the main obstacle is not negative attitude; it is the absence of systematic, supported implementation.

The study does not establish causal improvement in pronunciation because it relies on perceptions and teacher observations rather than objective pre/post testing, Its strongest contribution is to identify readiness, barriers, and design priorities for a future intervention. A scientifically robust next phase should implement the proposed multimodal cycle and assess intelligibility, comprehensibility, segmental accuracy, stress, rhythm, and learner confidence over time.

8.1 Recommendations

- 1- Provide targeted professional development on pronunciation pedagogy, multimedia design, mobile-assisted practice, and focused feedback rather than limiting training to technical operation.
- 2- Equip pronunciation classrooms with reliable audio, projection, internet access, speakers, and simple recording facilities, while maintaining low-technology alternatives for continuity.
- 3- Integrate short audio-visual pronunciation tasks into regular courses instead of treating technology as an occasional enrichment activity.
- 4- Design materials specifically for Arabic-speaking learners, prioritizing contrasts and prosodic patterns that influence intelligibility.
- 5- Adopt the six-stage cycle of focus, multimodal noticing, controlled production, recording, interactive feedback, and communicative transfer.
- 6- Use captions, articulatory video, pitch or stress visualization, and automatic speech recognition selectively and explain how learners should interpret feedback.
- 7- Create structured out-of-class mobile assignments with clear targets, rubrics, deadlines, and opportunities for revision.
- 8- Evaluate pronunciation through intelligibility and comprehensibility as well as accuracy, and avoid presenting native-like accent as the only legitimate outcome.

- 9- Establish a shared university repository of short, reviewed pronunciation resources and lesson plans that teachers can adapt.
- 10, Conduct experimental and longitudinal research with larger samples, comparison groups, delayed post-tests, and trained external raters.

8.2 Limitations

The study has several limitations, It was conducted at one university, limiting generalizability to other Libyan institutions, The student questionnaire used a three-point scale and included only eleven items, reducing measurement precision, The qualitative sample consisted of four teachers, so thematic variation may be incomplete, Most importantly, pronunciation performance was not directly tested; therefore, findings concern perceived effectiveness and reported improvement, The study also did not analyse differences by proficiency level, gender, year of study, frequency of technology access, or specific pronunciation feature, These limitations should guide rather than diminish future research.

8.3 Suggestions for Future Research

1. A quasi-experimental study comparing multimodal pronunciation instruction with conventional teaching using pre-, post-, and delayed tests.
- 2- An intervention comparing audio-only, visual-only, and integrated audio-visual conditions for selected Arabic-English contrasts.
3. A study of automatic speech recognition accuracy and bias for Libyan-accented English.
- 4- Longitudinal research on how structured mobile practice influences autonomy, willingness to communicate, and sustained pronunciation development.
5. A multi-university study examining infrastructure, teacher digital competence, and institutional policy across Libya.

8.4 Implementation and Evaluation Matrix

For practical adoption, the recommendations should be implemented through a staged plan that links each action to an observable indicator, The following matrix offers a concise roadmap that departments can adapt to available resources while preserving the pedagogical priority of intelligible and comprehensible speech.

Table 6, Practical implementation and evaluation roadmap

Priority	Action	Success indicator	Timeframe
Teacher preparation	Deliver short workshops on pronunciation pedagogy, resource selection, and feedback.	At least 80% of pronunciation teachers complete training and submit one redesigned lesson.	First semester
Resource readiness	Audit classrooms and create a minimum equipment package with low-technology alternatives.	Each participating class has reliable playback, projection, and recording access.	First semester
Curriculum integration	Embed one structured multimodal pronunciation cycle in each instructional unit.	Course outlines and lesson plans document recurring, assessed pronunciation practice.	Second semester
Learner practice	Assign brief mobile recording, comparison, and revision tasks outside class.	Students submit at least two recorded attempts and a reflection for each target feature.	Ongoing

PERCEIVED EFFECTIVENESS OF MULTIMODAL AUDIO-VISUAL INSTRUCTION
IN DEVELOPING ENGLISH PRONUNCIATION AMONG
LIBYAN UNIVERSITY EFL LEARNERS

Priority	Action	Success indicator	Timeframe
Assessment quality	Use analytic rubrics covering intelligibility, comprehensibility, accuracy, stress, and rhythm.	Pre/post samples are double-rated and inter-rater agreement is reported.	End of course
Programme review	Review usage data, learner outcomes, access barriers, and teacher feedback.	An annual report identifies gains, unresolved constraints, and revisions for the next cycle.	Annually

The matrix also prevents technology adoption from being evaluated only by the number of devices or applications available, Success is defined by teacher capability, repeated learner practice, measurable speech outcomes, and a documented process of institutional improvement.

References

- Amrate, M., & Tsai, P.-H. (2025), Computer-assisted pronunciation training: A systematic review, *ReCALL*, 37(1), 22–42, <https://doi.org/10.1017/S0958344024000181>
- Benlaghrissi, H., & Ouahidi, L. M. (2024), The impact of mobile-assisted project-based learning on developing EFL students' speaking skills, *Smart Learning Environments*, 11, Article 18, <https://doi.org/10.1186/s40561-024-00303-y>
- Braun, V., & Clarke, V. (2006), Using thematic analysis in psychology, *Qualitative Research in Psychology*, 3(2), 77–101, <https://doi.org/10.1191/1478088706qp063oa>
- Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010), *Teaching pronunciation: A course book and reference guide* (2nd ed.), Cambridge University Press.
- Creswell, J. W., & Plano Clark, V. L. (2018), *Designing and conducting mixed methods research* (3rd ed.), SAGE.
- Dai, Y., & Wu, Z. (2022), Mobile-assisted peer feedback on EFL pronunciation: Outcome effects, interactional processes, and shaping factors, *System*, 111, 102953, <https://doi.org/10.1016/j.system.2022.102953>
- Derwing, T. M., & Munro, M. J. (2015), *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*, John Benjamins.
- Dörnyei, Z. (2005), *The psychology of the language learner: Individual differences in second language acquisition*, Lawrence Erlbaum.
- Eisay, A. A. (2026). Code-Switching from Arabic to English among TV Presenters as an Indicator of Linguistic Competence and Linguistic Dominance. *Al-Farooq Journal of Sciences*, 2(3), 227-235.
- Fouz-González, J. (2020), Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app, *Language Learning & Technology*, 24(1), 62–85, <https://doi.org/10.64152/10125/44709>
- García, C., Nickolai, D., & Jones, L. (2020), Traditional versus ASR-based pronunciation instruction: An empirical study, *CALICO Journal*, 37(3), 213–232, <https://doi.org/10.1558/cj.40379>
- Hsieh, W.-M., Yeh, H.-C., & Chen, N.-S. (2025), Impact of a robot and tangible object (R&T) integrated learning system on elementary EFL learners' English pronunciation and willingness to communicate, *Computer Assisted Language Learning*, 38(4), 773–798, <https://doi.org/10.1080/09588221.2023.2228357>
- Kacatl, J., & Klímová, B. (2019), Use of smartphone applications in English language learning—A challenge for foreign language education, *Education Sciences*, 9(3), 179, <https://doi.org/10.3390/educsci9030179>
- Lan, E.-M. (2022). A comparative study of computer and mobile-assisted pronunciation training: The case of university students in Taiwan, *Education and Information Technologies*, 27(2), 1559–1583, <https://doi.org/10.1007/s10639-021-10647-4>
- Lee, B., Plonsky, L., & Saito, K. (2020), The effects of perception- vs, production-based pronunciation instruction, *System*, 88, 102185, <https://doi.org/10.1016/j.system.2019.102185>
- Loewen, S., Isbell, D. R., & Sporn, Z. (2020), The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability, *Foreign Language Annals*, 53(2), 209–233, <https://doi.org/10.1111/flan.12454>
- Latif, M. B., & Saeid, S. A. M. (2026). The Impact of Teaching English Literature as an Elective Course to Fourth-Year ESL Students in a University Context. *Al-Farooq Journal of Sciences*, 2(3), 1166-1179. <https://doi.org/10.65405/mfh93672>
- Mayer, R. E. (2020), *Multimedia learning* (3rd ed.), Cambridge University Press.

- Metruk, R. (2024), Mobile-assisted language learning and pronunciation instruction: A systematic literature review, *Education and Information Technologies*, 29, 16255–16282, <https://doi.org/10.1007/s10639-024-12453-0>
- Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2024), The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis, *ReCALL*, 36(1), 4–21, <https://doi.org/10.1017/S0958344023000113>
- O'Brien, M. G., Derwing, T. M., Cucchiarini, C., Hardison, D. M., Mixdorff, H., Thomson, R. I., Strik, H., Levis, J. M., Munro, M. J., Foote, J. A., & Levis, G. M. (2018), Directions for the future of technology in pronunciation research and teaching, *Journal of Second Language Pronunciation*, 4(2), 182–207, <https://doi.org/10.1075/jslp.17001.obr>
- Saito, K., & Plonsky, L. (2019), Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis, *Language Learning*, 69(3), 652–708, <https://doi.org/10.1111/lang.12345>
- Thomson, R. I., & Derwing, T. M. (2015), The effectiveness of L2 pronunciation instruction: A narrative review, *Applied Linguistics*, 36(3), 326–344, <https://doi.org/10.1093/applin/amu076>
- Toyama, M., & Hori, T. (2025), Technology-enhanced multimodal approaches in classroom L2 pronunciation training, *Frontiers in Education*, 10, 1552470, <https://doi.org/10.3389/feduc.2025.1552470>