Journal of Foreign Language Teaching and Applied Linguistics
ISSN 2303-5528 (Print) · e-ISSN 2303-6397 (Online)
j-fltal.org/j
Volume 13 · Issue 2 · 2026
Pages 969–981

Prospects for Digital English Language Teaching and Learning: Pronunciation and Speaking between Phonetic Standards and Linguistic Diversity

Mohamed Sayah[1], Belkacem Boudenna[2]

[1] University of Saida- Dr. Moulay Tahar, Algeria

[2] University Center of El Bayadh, Algeria

Email: sayah_2009@live.fr, b.boudenna@cu-elbayadh.dz

Received: 7 June 2026 | **Accepted:** 11 October 2026 | **Published:** 24 October 2026

Abstract

The rapid acceleration of digital technologies, artificial intelligence (AI), automatic speech recognition , and computer-assisted pronunciation training has fundamentally transformed English language teaching and learning. However, these technological innovations bring to the fore a profound theoretical and pedagogical dilemma: should digital pronunciation and oral communication instruction enforce strict adherence to traditional canonical native-speaker standards (such as Received Pronunciation or General American), or should it embrace an inclusive, intelligibility-driven paradigm that recognizes English as a Lingua Franca and the legitimacy of global World Englishes? Grounded in an epistemology of critical applied linguistics and digital sociophonetics, this paper synthesizes recent empirical and theoretical scholarship to examine the dynamic reconstruction of pronunciation standards in technology-mediated environments. Moving beyond the instrumental view of technology as a mere delivery mechanism, this paper investigates how multimodal learning ecosystems, conversational agents, and acoustic feedback systems reshape communicative competence, interactional agency, and learner identity. It proposes a five-stage pedagogical cycle—diagnosis, targeted practice, authentic interaction, differentiated feedback, and reflective reassessment—that harmonizes technological precision with linguistic inclusivity. By distinguishing communication-inhibiting phonological breakdowns from legitimate accent variation, this framework offers evidence-based curricular and technological principles for researchers, teacher educators, curriculum designers, and educational developers.

Keywords: Digital ,English Language Teaching, Phonetic Norms, Linguistic Diversity, AI Conversational Agents.

Introduction

The global landscape of English language teaching and learning is currently undergoing a profound ontological and pedagogical reconfiguration, catalyzed by the ubiquitous integration of Information and Communication Technologies , cloud-based multimodal platforms, and generative artificial intelligence (AI). As documented in recent baseline investigations, contemporary educational systems have transitioned rapidly from conventional teacher-fronted delivery to digitally mediated ecosystems (Hughes et al., 2023, p. 2; Narendran et al., 2025, p. 1507; Yusuf et al., 2024, p. 21). This transition, while accelerated by global institutional disruptions during the COVID-19 pandemic, reflects a deeper structural paradigm shift in applied

linguistics: language learning is no longer confined to isolated classroom drills or static textbooks, but unfolds within complex, interactive, and digitally augmented social networks (Babu, 2025, p. 166; Boukersoul & Sadoune, 2024, p. 781; López-Hidalgo et al., 2025, p. 4).

Pronunciation and Speaking as Interdependent Dimensions of Communicative Competence

Within this evolving digital architecture, oral language proficiency—encompassing pronunciation, fluent speaking, and interactional competence—occupies a central yet uniquely challenging position. Historically, pronunciation instruction has often been marginalized or treated as an isolated mechanical sub-skill, reduced to repetitive phonetic drills and decontextualized minimal pair repetitions. However, contemporary applied linguistics demonstrates that pronunciation and speaking are intrinsically interdependent dimensions of communicative competence (Hughes et al., 2023, p. 3; Pérez-Jorge et al., 2025, p. 2). Pronunciation provides the indispensable acoustic, segmental, and suprasegmental scaffolding that renders spoken discourse mutually intelligible, while speaking provides the dynamic, discursive, and pragmatic vehicle through which phonological forms acquire communicative meaning and social relevance (Boukersoul & Sadoune, 2024, p. 783; Celis & Cardona, 2026, p. 4).

The Tension Between Established Pronunciation Standards and Global Englishes

Crucially, the expansion of digital pronunciation tools brings to the surface an unresolved socio-pedagogical tension: the ideological conflict between the 'nativeness principle'—which demands unyielding conformity to prestigious inner-circle phonetic standards such as Received Pronunciation or General American —and the 'intelligibility principle', which foregrounds communicative effectiveness across linguistically and culturally diverse global populations (Amrate & Tsai, 2025, p. 23; Jenkins, 2000, p. 124; Zanke, 2018, p. 2). In the contemporary globalized economy, the overwhelming majority of English interactions occur between non-native speakers in international settings where English functions as a Lingua Franca . Consequently, measuring learner success against the unattainable or discriminatory benchmark of native-accent perfection is theoretically obsolete and pedagogically counterproductive (Bui, 2022, p. 2; Narendran et al., 2025, p. 1509).

Research Problem, Questions, Objectives, and Original Contribution

Despite widespread adoption of automated speech recognition (ASR), mobile applications, and AI conversational agents in language classrooms, educational practice remains caught in a structural paradox. Commercial software engines and algorithmic evaluators are predominantly trained on standardized native speech corpora, thereby penalizing legitimate non-native phonological variations and reinforcing standard-language ideologies (Babu, 2025, p. 167; Hughes et al., 2023, p. 11; Yusuf et al., 2024, p. 25). Conversely, unguided communicative activities in virtual environments often fail to provide the focused phonetic feedback required to eliminate genuine intelligibility barriers (López-Hidalgo et al., 2025, p. 15; Pérez-Jorge et al., 2025, p. 10).

To address this empirical and conceptual gap, this study examines five interrelated research questions:

1. How do digital environments and AI technologies reshape the conceptualization and development of English pronunciation and speaking competence?

2. What are the theoretical and pedagogical limitations of native-standard-centered automated speech assessment compared to an intelligibility-oriented framework?

3. How can multimodal phonetic tools, acoustic visualization, and pluralistic speech corpora be deployed to support pronunciation without imposing accent uniformity?

4. In what ways do interactive digital tasks, conversational AI, and collaborative podcasting cultivate learner communicative agency and intercultural accommodation?

5. What comprehensive, evidence-based model can guide curriculum design, teacher education, and algorithmic fairness in digital oral language education?

By adopting a scientific-philosophical and critical applied linguistic methodology, this monograph synthesizes primary empirical investigations, sociophonetic theory, and instructional frameworks. Its original contribution lies in formulating a dual-orientation model that reconciles high-precision technological diagnostics with linguistic inclusivity, providing concrete pedagogical protocols and policy recommendations for 21st-century language education.

I. Conceptual and Theoretical Foundations of Digital Pronunciation and Speaking Pedagogy

1.1. Digital Language Pedagogy: From Technological Delivery to Digitally Mediated Communication

A rigorous understanding of digital pronunciation and speaking instruction requires transcending the early technocentric view of educational technology as a mere delivery mechanism or electronic conduit for traditional grammar-translation and audiolingual drills. As Boukersoul and Sadoune (2024, p. 781) and Hughes et al. (2023, p. 2) observe, the evolution of digital language learning reflects a fundamental ontological transition toward digitally mediated communication, where digital tools function as semiotic ecosystems that restructure how learners perceive, interact with, and produce language.

In this sociocultural and socio-constructivist perspective (Boukersoul & Sadoune, 2024, p. 788; Vygotsky, 1978), learning occurs not through passive exposure to digitized content, but through active mediation, collaborative problem-solving, and participatory discourse in digital spaces (López-Hidalgo et al., 2025, p. 7; Narendran et al., 2025, p. 1508). Digital tools provide specific 'affordances'—action possibilities perceived and actualized by learners within their learning environment (Hughes et al., 2023, p. 6). When applied to oral language acquisition, digital platforms do not merely reproduce classroom speech; they offer novel modalities for acoustic visualization, asynchronous self-reflection, cross-cultural telecollaboration, and interactive dialogue with intelligent agents (Babu, 2025, p. 166; Pérez-Jorge et al., 2025, p. 2).

1.2. Pronunciation, Intelligibility, Comprehensibility, and Accentedness

To construct an equitable and scientifically rigorous pedagogical model, applied linguistics relies on the foundational tripartite distinction formulated by Derwing and Munro (2005, p. 385) and corroborated across contemporary empirical studies (Amrate & Tsai, 2025, p. 24; Celis & Cardona, 2026, p. 3; Zanke, 2018, p. 1):

• Intelligibility: The extent to which an acoustic speech signal is actually recognized and decoded by a listener (i.e., whether the listener accurately identifies the words and message uttered).

• Comprehensibility: The listener's subjective perception of the degree of difficulty or cognitive effort required to understand the spoken utterance.

• Accentedness: The degree to which a speaker's phonetic, phonological, and prosodic patterns diverge from a specific reference norm or regional variety, as judged by a listener.

The empirical separation of these constructs is of paramount importance for digital pedagogy. Robust psycholinguistic research has established that a speaker can exhibit a noticeable regional or non-native accent (high accentedness) while remaining perfectly and effortlessly understandable (high intelligibility and high comprehensibility) (Derwing & Munro, 2005, p. 387; Jenkins, 2000, p. 130). Consequently, conflating accent elimination with communicative competence represents an epistemological fallacy that misdirects instructional energy and causes acute foreign language speaking anxiety (Babu, 2025, p. 167; Yusuf et al., 2024, p. 24).

1.3. Communicative Competence, Interactional Competence, and the Integration of Pronunciation with Speaking

Pronunciation instruction attains pedagogical validity only when it is systematically embedded within broader models of communicative competence (Canale & Swain, 1980; Celce-Murcia et al., 1995) and interactional competence (Kramsch, 1986; Markee, 2008). Communicative competence encompasses linguistic/grammatical, sociolinguistic, discourse, and strategic competencies. Pronunciation constitutes the

physiological and phonological bedrock of linguistic competence, directly informing discourse cohesion through suprasegmental phrasing and prosodic prominence (Boukersoul & Sadoune, 2024, p. 783; López-Hidalgo et al., 2025, p. 9).

Furthermore, in spoken interaction, pronunciation cannot be separated from interactional competence—the co-constructed ability to manage turn-taking, repair conversational breakdowns, negotiate meaning, and signal pragmatic stance through intonation and pitch modulation (Hughes et al., 2023, p. 3; Pérez-Jorge et al., 2025, p. 11). A robust digital pedagogical architecture must therefore orchestrate a three-part developmental sequence:

1. Phonetic Perception: Developing acoustic awareness to discriminate critical phonemic contrasts, lexical stress, rhythm, and intonational contours.

2. Controlled Production: Scaffolding articulatory and prosodic accuracy through targeted, digitally assisted practice and acoustic feedback.

3. Spontaneous Interaction: Deploying phonological features dynamically in communicative tasks, collaborative problem-solving, and authentic digital dialogue.

1.4. English as a Lingua Franca, World Englishes, and the Sociolinguistic Legitimacy of Accent Diversity

The sociolinguistic reality of contemporary global communication is characterized by the framework of World Englishes (Kachru, 1985) and the paradigm of English as a Lingua Franca (Jenkins, 2000; Seidlhofer, 2011). In Kachru's Three Concentric Circles model—comprising the Inner Circle (native varieties like UK, USA), Outer Circle (post-colonial institutionalized varieties like India, Nigeria, Singapore), and Expanding Circle (EFL contexts like Spain, China, Brazil, Ecuador)—the Expanding Circle accounts for the vast majority of English users worldwide (Narendran et al., 2025, p. 1508; Yusuf et al., 2024, p. 22).

Jenkins' (2000, pp. 123–158) seminal research on the Lingua Franca Core (LFC) identified the specific phonological features essential for mutual intelligibility in international communication, alongside non-core features that do not cause communicative failure despite diverging from native norms. As synthesized in Table 1 below, core features include most consonant contrasts, vowel quantity distinctions (duration), nuclear stress placement, and the preservation of initial consonant clusters. Conversely, non-core features—such as dental fricatives (/θ, ð/), vowel quality nuances (e.g., precise RP vowel formant positions), and pitch trajectory directionality—can freely vary without impairing international intelligibility (Amrate & Tsai, 2025, p. 26; Zanke, 2018, p. 2).

1.5. Digital Sociophonetics and the Implications of Technology-Mediated Speech Analysis

Digital sociophonetics investigates the intersection of acoustic phonetics, sociolinguistic variation, and computational speech processing. Modern computational tools—such as Praat acoustic analyzers, automated formant tracking, and fundamental frequency (F0) contour extractors—provide unprecedented precision in visualizing acoustic parameters (Amrate & Tsai, 2025, p. 25). However, digital sociophonetics also demonstrates that speech is inherently variable, indexical of social identity, and context-dependent. Digital speech tools that treat phonetic norms as static, monolithic constants ignore the natural phonetic variability of human speech communities and impose an artificial homogeneity on diverse learner populations (Hughes et al., 2023, p. 11; Yusuf et al., 2024, p. 25).

II. Digital Technologies and the Reconstruction of Pronunciation Standards

2.1. Automatic Speech Recognition and Computer-Assisted Pronunciation Training: Affordances and Limitations

Computer-Assisted Pronunciation Training and Automatic Speech Recognition (ASR) systems have revolutionized oral language instruction by enabling autonomous, individualized, and high-frequency

Table 1. Sociolinguistic Phonological Matrix: Lingua Franca Core vs. Non-Core Features in Digital Instruction

Phonological Domain	Lingua Franca Core (Critical for Intelligibility)	Non-Core Features (Legitimate Accent Variation)	Pedagogical & Digital Focus
Consonants	Preservation of all consonant contrasts (except /θ, ð/); aspiration of voiceless plosives (/p, t, k/); consistent rhoticity (/r/).	Substitution of /θ, ð/ with [t, d] or [f, v]; allophonic variations of /l/ (clear vs. dark [ɬ]).	Prioritize functional phonemic oppositions in ASR training; avoid penalizing dental fricative substitutions.
Consonant Clusters	Maintenance of initial word clusters without deletion (e.g., 'string' → [strɪŋ], not [trɪŋ]); epenthesis preferable to deletion.	Simplification of medial and final clusters according to native-like elision rules (e.g., 'facts' → [fæks]).	Train acoustic recognition to differentiate cluster repair strategies; prevent communicative breakdown in initial positions.
Vowels & Quantity	Vowel length/quantity distinctions (contrast between long and short vowels, e.g., /iː/ vs. /ɪ/, /uː/ vs. /ʊ/).	Precise native vowel quality and formant frequencies; exact centralization of unstressed vowels (schwa /ə/).	Utilize visual vowel duration charts and spectrograms rather than demanding exact native-speaker formant positioning.
Suprasegmentals & Prosody	Nuclear stress (tonic placement) indicating information focus; correct division of speech into tone units (chunking).	Tone directionality (rise vs. fall nuances for subtle attitudinal signaling); stress-timed rhythm vs. syllable-timed rhythm.	Implement F0 pitch contour visualization for prominence marking and syntactic chunking during oral discourse.

Note. Synthesized from Jenkins (2000, pp. 123–158), Amrate & Tsai (2025, pp. 24–28), and Zanke (2018, pp. 2–4).

phonological practice. As highlighted by Babu (2025, p. 166), Narendran et al. (2025, p. 1509), and Pérez-Jorge et al. (2025, p. 2), modern CAPT platforms provide immediate acoustic feedback, speech rate calculation, and mispronunciation detection, relieving classroom time constraints and allowing learners to practice in low-anxiety, self-paced environments.

Despite these substantial affordances, ASR-based systems exhibit critical theoretical and technical limitations. Most commercial engines rely on Hidden Markov Models or deep neural networks trained on acoustic data from a narrow demographic of standard native speakers (predominantly standard American or southern British dialects). Consequently, when a non-native speaker produces an utterance that is fully intelligible to human listeners but exhibits L1-influenced phonetic variation, the ASR algorithm categorizes it as an error, generating misleading diagnostic scores and learner demotivation (Hughes et al., 2023, p. 11; Yusuf et al., 2024, p. 25; Zanke, 2018, p. 3).

2.2. Artificial Intelligence, Speech Synthesis, and Conversational Agents in Oral-Language Development

Recent advances in Generative Artificial Intelligence (GenAI), Large Language Models (LLMs), and neural text-to-speech (TTS) synthesis have extended digital pedagogy beyond isolated acoustic scoring to interactive conversational agents (Babu, 2025, p. 167; Celis & Cardona, 2026, p. 5). AI-powered conversational systems (e.g., ChatGPT Voice, conversational avatars, intelligent dialogue systems) simulate naturalistic conversational exchanges, adapting their linguistic complexity to the user's proficiency level and providing real-time communicative practice (Pérez-Jorge et al., 2025, p. 8).

Neural speech synthesis engines now generate highly naturalistic, human-like voice models with authentic prosody, emotional inflection, and varied speaking rates. When systematically integrated into speaking pedagogy, conversational agents foster oral fluency, lower foreign language anxiety, and scaffold the pragmatic management of conversational turns (Boukersoul & Sadoune, 2024, p. 787; López-Hidalgo et al., 2025, p. 14).

2.3. Accent Representation, Algorithmic Bias, and the Risks of Standard-Centred Automated Feedback

A critical applied linguistic critique reveals that digital technologies are never ideologically neutral. When automated scoring algorithms are calibrated exclusively against standard native varieties, they institutionalize what sociolinguists term 'algorithmic linguistic discrimination' or standard-language bias (Amrate & Tsai, 2025, p. 25; Zanke, 2018, p. 3). Learners whose regional accents deviate from GA or RP are consistently penalized, reinforcing the false belief that their biological accent is a deficit that must be eradicated.

Furthermore, as Yusuf et al. (2024, p. 24) and Hughes et al. (2023, p. 13) emphasize, such standard-centred feedback induces linguistic insecurity, inhibits willingness to communicate (WTC), and alienates learners from their local multilingual identities. Educational technology developers and language educators

must therefore actively audit speech recognition architectures to ensure they reflect global acoustic diversity rather than perpetuating monocentric normativity.

2.4. Multimodal Learning Environments: Audio, Visual Phonetics, Speech Recording, and Acoustic Feedback

To optimize phonetic development, digital pedagogy leverages multimodal learning environments that combine acoustic, visual, and kinesthetic sensory channels (Hughes et al., 2023, p. 3; Pérez-Jorge et al., 2025, p. 9). Visual phonetics tools provide dynamic sagittal-plane cross-sections of the vocal tract, 3D animated articulatory models, and ultrasound biofeedback illustrating precise tongue and lip positioning during phonation (Amrate & Tsai, 2025, p. 26).

Simultaneously, acoustic feedback platforms render real-time pitch (F0) contours, spectrograms, and waveform envelopes. When learners record their speech and visually compare their pitch track against model utterances, they develop metacognitive awareness of suprasegmental features—such as prominence, rhythm, and tone unit boundaries—that are notoriously difficult to perceive through auditory input alone (López-Hidalgo et al., 2025, p. 11; Narendran et al., 2025, p. 1510).

2.5. Pedagogical Conditions for Integrating Technological Precision with Linguistic Inclusivity

Reconciling technological precision with linguistic inclusivity requires establishing clear pedagogical boundary conditions (Boukersoul & Sadoune, 2024, p. 798; Hughes et al., 2023, p. 14). Digital tools must be configured not as authoritarian evaluators of accent purity, but as diagnostic mirrors that support intelligibility. This pedagogical integration demands that teachers possess critical digital literacy (Babu, 2025, p. 167; Narendran et al., 2025, p. 1510) and that speech technologies be re-engineered to distinguish between acoustic distortions that obstruct semantic decoding and non-standard features that reflect legitimate sociolinguistic variation.

Figure 1. Empirical Assessment of Digital Competence Dimensions in Language Pedagogy

↕

	Oral Interaction & Pronunciation	Mediation & Intercultural Skills	Digital Assessment & Feedback	Receptive Multimodal Skills	General Tool Literacy
Mean Importance / Need Score (1-5 Scale)	4.45	4.3	4.15	3.85	3.6

Figure 1: Empirical distribution of digital pedagogical training needs in language education (Synthesized from Hughes et al., 2023, pp. 9–11; Yusuf et al., 2024, pp. 27–29).

III. An Applied Framework for Digital Pronunciation Instruction

3.1. The Digital Phonetic Laboratory: Segmental Contrasts, Word Stress, Rhythm, and Intonation

An applied digital pronunciation curriculum is structured around the concept of a 'digital phonetic laboratory'—a flexible instructional space that integrates software tools to address both segmental and suprasegmental domains (Boukersoul & Sadoune, 2024, p. 785; Celis & Cardona, 2026, p. 12). At the segmental level, the laboratory focuses on high-functional-load phonemic oppositions (e.g., /p/ vs. /b/, /iː/ vs. /ɪ/, /l/ vs. /r/) using interactive minimal-pair sorting algorithms and waveform comparison tools (Amrate & Tsai, 2025, p. 26; Jenkins, 2000, p. 132).

At the suprasegmental level, instructional software targets lexical stress placement (critical for English word recognition), sentence-level rhythm, and intonational phrasing. By utilizing digital pitch visualizers, learners observe how manipulating fundamental frequency (F0) signals new information versus given information, thereby mastering the communicative mechanics of English prosody (López-Hidalgo et al., 2025, p. 11; Narendran et al., 2025, p. 1509).

3.2. A Pluralistic Digital Speech Corpus: Exposure to Diverse English Accents and Communicative Contexts

To dismantle monocentric bias and prepare learners for international communication, digital pronunciation instruction must replace single-dialect audio models with pluralistic digital speech corpora (Hughes et al., 2023, p. 5; Zanke, 2018, p. 3). Platforms such as the Speech Accent Archive, YouGlish, and curated open-access international corpora expose learners to authentic speech samples produced by speakers from Inner, Outer, and Expanding Circle backgrounds.

Systematic exposure to accent diversity achieves two crucial psycholinguistic objectives: it expands learners' perceptual phonetic boundaries, enhancing their ability to comprehend unfamiliar accents during international encounters (intercultural listening comprehension), and it validates their own linguistic identity, demonstrating that effective English communication does not require adopting an Anglo-American persona (Babu, 2025, p. 167; Yusuf et al., 2024, p. 23).

3.3. Intelligibility-Oriented Assessment: Evaluating Phonetic Features According to Communicative Significance

Assessment practices in digital pronunciation must undergo a paradigm shift from 'accent eradication' to 'intelligibility-oriented assessment' (Derwing & Munro, 2005, p. 388; Jenkins, 2000, p. 155). Rather than deducting marks for every deviation from canonical British or American norms, evaluative rubrics must weight phonological errors based on their functional impact on communicative clarity.

Errors that disrupt the Lingua Franca Core—such as shifting lexical stress to an incorrect syllable, deleting initial consonant clusters, or collapsing vowel duration contrasts in minimal pairs—receive primary diagnostic focus. In contrast, accent features that do not impede comprehension—such as realizing /θ/ as [t] or [f]—are recognized as legitimate dialectal variation and are not penalized in oral proficiency scores (Amrate & Tsai, 2025, p. 27; Zanke, 2018, p. 4).

3.4. The Digital Pronunciation Portfolio: Longitudinal Evidence, Self-Assessment, Peer Feedback, and Teacher Guidance

To foster learner autonomy and metacognitive reflection, the applied framework incorporates the 'Digital Pronunciation Portfolio' (Boukersoul & Sadoune, 2024, p. 797; Hughes et al., 2023, p. 12; López-Hidalgo et al., 2025, p. 14). Within an e-portfolio platform (e.g., Padlet, Flip, Canvas ePortfolios), learners curate audio and video recordings of their oral performance across an academic semester. The portfolio integrates three complementary evaluative layers:

• Self-Assessment: Learners transcribe their own recordings, analyze acoustic waveforms, and reflect on their intelligibility strengths and weaknesses.

• Peer Feedback: Structured peer-review protocols where classmates evaluate mutual comprehensibility and negotiate meaning asynchronously.

• Teacher Guidance: Instructors provide targeted qualitative feedback on high-priority communicative features, guiding individual learning trajectories.

3.5. Differentiated Feedback: Distinguishing Communication-Obstructing Difficulties from Legitimate Accent Variation

A cornerstone of the applied framework is the implementation of 'differentiated feedback' (Amrate & Tsai, 2025, p. 28; Hughes et al., 2023, p. 11). Teachers and automated diagnostic tools must explicitly distinguish between two categories of spoken production:

1. Communication-Obstructing Breakdowns: Phonetic substitutions, cluster deletions, or misplaced nuclear stress that lead to lexical ambiguity, listener confusion, or complete message failure.

2. Legitimate Accent Variations: Phonetic modifications, allophonic substitutions, and intonational preferences that preserve semantic integrity and reflect the

speaker's regional or cultural linguistic background.

By channeling corrective feedback exclusively toward communication-obstructing features, instructors prevent cognitive overload, reduce speaking anxiety, and foster communicative confidence (Babu, 2025, p. 167; Yusuf et al., 2024, p. 24).

Table 2. Applied Diagnostic Matrix for Differentiated Phonetic Feedback in Digital ELT

Speech Phenomenon	Observed Utterance	Acoustic / Linguistic Status	Impact on Intelligibility	Pedagogical Action / Feedback
Cluster Deletion	'I study at the *univer-tity* / *es-chool*'	Deletion of /s/ in /st/ or initial s-cluster modification.	High / Disruptive (causes lexical misidentification).	Targeted Intervention: Deploy visual spectrogram and articulatory animation to practice cluster preservation.
Lexical Stress Shift	'We need to *pro-JECT* (intended noun: project)'	Stress placed on ultimate syllable instead of penultimate syllable.	High / Disruptive (causes grammatical/semantic confusion).	Targeted Intervention: Acoustic F0 tracking and duration practice to contrast noun/verb stress patterns.
Vowel Duration Collapse	'Please *fit* the sheep (intended: feed)'	Failure to contrast tense long vowel /iː/ and lax short vowel /ɪ/.	High / Critical (causes semantic minimal-pair confusion).	Targeted Intervention: Duration waveform stretching exercises to calibrate vowel length contrasts.
Dental Fricative Substitution	'I *think* so → [aɪ tɪŋk soʊ] / [aɪ fɪŋk soʊ]'	Realization of /θ/ as dental stop [t] or labiodental [f].	Low / Negligible (context easily resolves meaning in ELF).	No Intervention: Accept as legitimate regional accent variation; do not penalize in scoring.
Rhoticity Variation	'The car was parked [ðə kɑː wəz pɑːkt]'	Non-rhotic pronunciation of post-vocalic /r/ in rhotic target.	Zero / Non-existent (both rhotic and non-rhotic models are standard).	No Intervention: Fully acceptable variation across global English varieties.

Note. Formulated based on Derwing & Munro (2005, p. 385), Jenkins (2000, pp. 135–140), and Amrate & Tsai (2025, pp. 27–30).

IV. Digital Speaking, Interaction, and the Development of Communicative Agency

4.1. Scenario-Based Digital Speaking: Interviews, Presentations, Discussions, and Collaborative Problem-Solving

Transitioning from phonetic control to authentic oral proficiency requires immersive, scenario-based digital speaking environments (Boukersoul & Sadoune, 2024, p. 786; López-Hidalgo et al., 2025, p. 10; Pérez-Jorge et al., 2025, p. 12). Digital platforms enable the creation of high-fidelity professional and academic simulations—such as virtual job interviews, conference panel presentations, boardroom negotiations, and international project meetings.

In these scenario-based tasks, learners must synthesize pronunciation accuracy, lexical precision, and pragmatic appropriateness in real time. Working in digital breakout rooms or synchronous virtual spaces (e.g., Zoom, Microsoft Teams, VR immersive environments), students engage in collaborative problem-solving that demands active negotiation, clarification requests, and discursive argumentation (Babu, 2025, p. 166; Yusuf et al., 2024, p. 23).

4.2. AI Conversational Partners: Adaptive Dialogue, Fluency Development, and Reflective Feedback

The integration of Generative AI conversational partners represents a major frontier in oral fluency development (Babu, 2025, p. 167; Celis & Cardona, 2026, p. 8). Unlike human interlocutors whose availability is constrained, AI dialogue agents offer unlimited, round-the-clock conversational immersion. These agents can be prompted to assume specific personas (e.g., an inquisitive international journalist, a technical client, an academic peer), scaffolding interactive speech within defined communicative registers.

Importantly, conversational AI can be programmed to provide immediate, post-interaction reflective feedback. Following a ten-minute spoken dialogue, the system generates a diagnostic summary analyzing the learner's vocabulary range, grammatical accuracy, turn-taking patterns, and perceived phonological clarity. This reflective feedback loop enables learners to identify recurring speech bottlenecks and immediately retry communication strategies (Hughes et al., 2023, p. 6; Pérez-Jorge et al., 2025, p. 9).

4.3. The Multi-Voice Podcast: Authentic Production, Audience Awareness, and Collaborative Oral Communication

Podcasting has emerged as one of the most effective pedagogical tools for cultivating oral communication, voice modulation, and audience awareness (Boukersoul & Sadoune, 2024, p. 787; Hughes et al., 2023, p. 5; López-Hidalgo et al., 2025, p. 17). In the 'multi-voice podcast' task model, student teams script, record, edit, and publish digital audio episodes addressing contemporary scientific, professional, or cultural topics.

Producing a podcast requires students to master acoustic delivery: adjusting speech tempo, articulating clearly for a remote audience, employing expressive intonation to sustain listener interest, and synchronizing conversational handovers between co-hosts. Because podcasts are published to authentic digital audiences (e.g., course channels, Spotify, SoundCloud), students develop genuine communicative agency and communicative investment (Narendran et al., 2025, p. 1509; Yusuf et al., 2024, p. 24).

4.4. Intercultural Digital Dialogue: Accommodation, Negotiation of Meaning, and Communicative Repair

Digital telecollaboration and virtual exchange programs connect language learners across geographical and linguistic borders, creating authentic contexts for intercultural digital dialogue (Hughes et al., 2023, p. 5; Pérez-Jorge et al., 2025, p. 6). In these multicultural interactions, success depends fundamentally on speech accommodation—the sociolinguistic capacity to adjust one's speaking rate, vocabulary, and phonetic clarity to match the interlocutor's comprehension needs (Giles et al., 1991; Jenkins, 2000, p. 165).

Learners develop crucial strategic competencies in conversational repair: identifying acoustic misunderstandings, rephrasing ambiguous statements, using comprehension checks ('Do you know what I mean?'), and employing confirmation checks. Through repeated intercultural exchanges, students realize that mutual intelligibility is a collaborative, co-constructed achievement rather than a unilateral native-speaker standard (Bui, 2022, p. 2; Narendran et al., 2025, p. 1510).

4.5. Integrated Assessment of Oral Performance: Intelligibility, Fluency, Interactional Effectiveness, and Pragmatic Appropriateness

Comprehensive assessment of oral performance in digital environments requires an integrated, multi-componential evaluative framework. As outlined in Table 3, a robust digital assessment rubric must synthesize four core dimensions:

1. Phonetic Intelligibility & Comprehensibility: Core phonological clarity, nuclear stress placement, and ease of listener decoding.

2. Spoken Fluency: Temporal flow, speech rate, pause distribution (distinguishing natural grammatical pauses from hesitant lexical searches)(boudenna,2025,p383)

3. Interactional Effectiveness: Active listening, appropriate turn-taking, responsive conversational management, and collaborative repair.

4. Pragmatic Appropriateness: Register choice, politeness markers, modal hedging, and socio-cultural contextual fit.

Table 3. Multi-Componential Rubric for Integrated Digital Oral Assessment

Competence Dimension	Exemplary Performance (Level 4)	Competent Performance (Level 3)	Developing Performance (Level 2)	Beginning Performance (Level 1)
Phonetic Intelligibility & Prosody (Weight: 30%)	Effortlessly intelligible; nuclear stress precisely marks information focus; vowels/consonant core well-maintained; accent is authentic and clear.	Fully intelligible with minor effort; occasional stress shifts that do not obscure meaning; core phonemes intact.	Partially intelligible; frequent core phoneme confusions or cluster deletions requiring listener concentration.	Severe intelligibility breakdowns; pervasive phonemic collapses and misplaced stress obstructing message decoding.
Spoken Fluency & Speech Rate (Weight: 25%)	Smooth, continuous delivery (130-160 wpm); natural pause placement at syntactic boundaries; minimal disruptive hesitation.	Generally continuous flow; minor hesitation during complex lexical search; natural pacing maintained.	Noticeably fragmented speech; frequent mid-clause pauses and false starts impairing discourse continuity.	Extremely halting delivery (<80 wpm); pervasive long pauses and fragmentation preventing coherent expression.

Table 3. Multi-Componential Rubric for Integrated Digital Oral Assessment *(continued)*

Competence Dimension	Exemplary Performance (Level 4)	Competent Performance (Level 3)	Developing Performance (Level 2)	Beginning Performance (Level 1)
Interactional Effectiveness & Repair (Weight: 25%)	Seamless turn management; initiates and responds dynamically; actively executes communicative repair and accommodation.	Effective turn-taking; responds appropriately to partner; resolves misunderstandings when prompted.	Passive interaction; relies heavily on interlocutor to drive conversation; struggles to execute conversational repair.	Fails to maintain interactional flow; unresponsive or unaware of partner's communicative signals.
Pragmatic Appropriateness (Weight: 20%)	Flawless alignment with task register; sophisticated use of modal hedging, politeness markers, and conversational conventions.	Appropriate register maintained throughout; clear and polite phrasing suited to the communicative scenario.	Occasional register lapses; overreliance on informal phrasing in formal tasks; limited hedging.	Inappropriate register; blunt or misaligned pragmatic choices causing potential communicative friction.

Note. Developed from Hughes et al. (2023, pp. 3–4), López-Hidalgo et al. (2025, pp. 19–21), and Council of Europe CEFR Companion Volume (2020).

Section V. Toward an Inclusive and Evidence-Based Model of Digital Pronunciation and Speaking Instruction

5.1. The Principles of a Dual-Orientation Model: Phonetic Comprehensibility and Linguistic Diversity

Synthesizing the theoretical, technological, and empirical dimensions analyzed throughout this study, we propose the 'Dual-Orientation Model of Digital Pronunciation and Speaking Instruction' (Figure 2). This model operates on two complementary pillars:

• Pillar 1: High-Precision Phonetic Comprehensibility. Deploying advanced digital diagnostic tools, acoustic visualization, and ASR feedback to eliminate genuine intelligibility bottlenecks and master core phonological features.

• Pillar 2: Sociolinguistic Inclusivity and Accent Legitimacy. Validating multilingual learner identities, exposing students to global Englishes, and cultivating intercultural accommodation without demanding native-speaker mimicry.

5.2. A Staged Pedagogical Cycle: Diagnosis, Targeted Practice, Authentic Interaction, Feedback, and Reassessment

The model is operationalized through a recurring five-stage instructional cycle designed for blended, virtual, and hybrid learning environments:

Stage 1: Multi-Accent Acoustic Diagnosis. Learners complete an automated speech diagnostic that evaluates phonemic contrasts, lexical stress, and speech rate across diverse accent reference models, identifying specific individual intelligibility priorities.

Stage 2: Multimodal Targeted Practice. Students engage with interactive visual phonetic tools, 3D articulatory models, and waveform pitch trackers to practice high-priority core features in controlled and semi-controlled tasks.

Stage 3: Scenario-Based Authentic Interaction. Learners deploy their phonetic skills in spontaneous speaking tasks, collaborative digital problem-solving, AI conversational simulations, or multi-voice podcast recordings.

Stage 4: Differentiated & Peer-Reflective Feedback. Students receive differentiated diagnostic feedback from instructors, structured peer reviews, and automated metric summaries, focusing strictly on communication-obstructing features.

Stage 5: Longitudinal Reassessment & Portfolio Synthesis. Learners archive their spoken productions in their digital portfolio, conduct acoustic self-analyses against earlier recordings, and set new communicative goals.

5.3. Teacher and Learner Agency in AI-Mediated Language Learning

A critical condition for the success of this model is the preservation and enhancement of human agency (Hughes et al., 2023, p. 14; Yusuf et al., 2024, p. 30). AI and automated speech tools must serve as cognitive

scaffolds rather than autonomous pedagogical decision-makers. Teachers must exercise critical evaluative judgment, overriding algorithmic misclassifications and guiding learners to interpret automated feedback through an ELF lens.(boudenna,2025 p386).

Simultaneously, learners must be empowered as active linguistic agents who view their pronunciation not as an imperfect copy of an idealized native norm, but as an authentic, expressive voice for international participation (Babu, 2025, p. 167; Narendran et al., 2025, p. 1511).

5.4. Implementation Challenges: Digital Access, Privacy, Teacher Preparation, and Algorithmic Fairness

Implementing this comprehensive model entails addressing several structural and institutional challenges documented in recent literature (Boukersoul & Sadoune, 2024, p. 799; Hughes et al., 2023, p. 13; López-Hidalgo et al., 2025, p. 5; Yusuf et al., 2024, p. 29):

• Digital Divide & Infrastructure Inequity: Unequal access to high-speed broadband, modern acoustic hardware, and premium educational software across different socio-economic and regional contexts.

• Data Privacy & Biometric Security: Protecting learner biometric voice data, student audio recordings, and confidential profiles when utilizing cloud-based commercial AI platforms.

• Pre-Service & In-Service Teacher Training: Overcoming the documented deficit in teacher training regarding digital phonetic tools, acoustic analysis, and ELF pedagogical frameworks.

• Algorithmic Fairness: Advocating for transparent, ethically audited ASR models trained on diverse global English speech corpora.

5.5. Evaluation Criteria and Recommendations for Institutions, Curriculum Designers, and Technology Developers

To ensure sustainable and ethical integration, we offer specific evidence-based recommendations across three educational tiers:

1. For Educational Institutions & Policy Makers: Institutionalize digital language competencies within national curricula, upgrade digital laboratory infrastructure, and mandate comprehensive professional development in digital pedagogy for language faculty (Hughes et al., 2023, p. 14; Yusuf et al., 2024, p. 31).

2. For Curriculum & Materials Designers: Transition syllabus benchmarks from native-accent mimicry to intelligibility-driven communicative outcomes; integrate pluralistic speech corpora and multi-voice collaborative tasks into coursebooks and digital platforms (Boukersoul & Sadoune, 2024, p. 798; López-Hidalgo et al., 2025, p. 21).

3. For EdTech & AI Developers: Train speech recognition engines on multi-dialectal and non-native speech datasets; implement customizable diagnostic thresholds that prioritize Lingua Franca Core features while disabling penalization of non-core regional accent markers (Babu, 2025, p. 168; Narendran et al., 2025, p. 1511).

Figure 2. Structural Component Allocation in the Dual-Orientation Pedagogical Model

↕

	Authentic Interaction & Scenarios	Differentiated Feedback & Repair	Targeted Acoustic Practice	Acoustic Diagnosis & Auditing	Portfolio Synthesis & Reflection
Curricular Time & Weight Allocation (%)	35	20	20	15	10

Figure 2: Proposed curricular time and pedagogical weighting across the five stages of the Dual-Orientation Model.

Conclusion

This monograph has conducted a comprehensive, multi-dimensional inquiry into the prospects, affordances, and challenges of digital English pronunciation and speaking instruction. By bridging the gap between

technological innovation and sociolinguistic reality, our analysis demonstrates that digital tools—far from being mere automated drill-masters—possess the transformative capacity to empower learners as confident, autonomous, and intelligible global communicators. However, realizing this potential requires abandoning the outdated ideological commitment to native-speaker accent conformity and fully embracing an intelligibility-oriented, pluralistic pedagogical paradigm.

The reconstruction of pronunciation standards in the digital age demands a fundamental realignment of pedagogical values. Pronunciation instruction must be repositioned as an empowering enabler of mutual comprehension rather than a gatekeeping mechanism of social prestige. Digital assessment systems must evaluate speech based on its communicative functionality, interactional agility, and pragmatic nuance, treating accent diversity as an inherent, legitimate feature of a vibrant global language.

The Dual-Orientation Model formulated in this paper provides a concrete, theoretically grounded blueprint for harmonizing technological precision with linguistic inclusivity. Nonetheless, future empirical research must validate and refine this framework across diverse educational ecosystems. Longitudinal intervention studies are needed to measure the long-term impact of ASR-mediated ELF instruction on learner intelligibility, communicative anxiety, and identity development. Additionally, ongoing collaborative research between applied linguists and computer scientists is imperative to develop next-generation, ethically aligned AI speech architectures that celebrate rather than suppress global phonetic diversity.

References

Aljurijeer, S. (2025). The pedagogical integration of computer-assisted pronunciation tools in tertiary education. Modern Language Journal, 109(1), 65–82.

Amrate, M., & Tsai, Y. (2025). Digital sociophonetics and the intelligibility paradigm: Re-evaluating accent in automated speech recognition. Journal of Second Language Pronunciation, 11(1), 22–42. https://doi.org/10.1075/jslp.24012.amr

Assassi, T. (2021). Needs analysis in ESP curriculum design: A pragmatic approach to professional communication. Arab World English Journal, 12(3), 145–160.

Babu, M. P. (2025). The future of English language teaching: Innovations in digital learning. Integrated Journal for Research in Arts and Humanities, 5(1), 166–168. https://doi.org/10.55544/ijrah.5.1.20

Boukersoul, F., & Sadoune, F. (2024). Teaching English for specific purposes in a digital environment. El Mohtaref Journal of Sports Science, Social and Human Science, 11(3), 780–803.

Bui, T. (2022). English as a Lingua Franca and the pedagogical legitimacy of regional accents in digital language learning. TESOL Quarterly, 56(4), 1–18.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics, 1(1), 1–47.

Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. Issues in Applied Linguistics, 6(2), 5–35.

Celis, R., & Cardona, M. (2026). Artificial intelligence and speech technology in language education: A critical review of pedagogical affordances. Language Learning & Technology, 30(1), 1–28.

Council of Europe. (2020). Common European framework of reference for languages: Learning, teaching, assessment – Companion volume. Council of Europe Publishing.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. TESOL Quarterly, 39(3), 379–397. https://doi.org/10.2307/3588486

Giles, H., Coupland, N., & Coupland, J. (1991). Contexts of accommodation: Developments in applied sociolinguistics. Cambridge University Press.

Hughes, S. P., Corral-Robles, S., & Ortega-Martín, J. L. (2023). Let's get digital: ICT training needs in pre-service language teaching. Education Sciences, 13(12), Article 1238, 1–16. https://doi.org/10.3390/educsci13121238

Hutchinson, T., & Waters, A. (1987). English for specific purposes: A learning-centred approach. Cambridge University Press.

Jenkins, J. (2000). The phonology of English as an international language: New models, new norms, new goals. Oxford University Press.

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), English in the world: Teaching and learning the language and literatures (pp. 11–30). Cambridge University Press.

Kramsch, C. (1986). From language proficiency to interactional competence. The Modern Language Journal, 70(4), 366–372.

Lebert, M. (2011). Del libro impreso al libro digital. Project Gutenberg / NEF, Universidad de Toronto. http://www.gutenberg.org/ebooks/34091

López-Hidalgo, D. B., Mazo-Atiaja, A. P., & Flores-Ubidia, J. L. (2025). Teaching English through active methodologies in virtual and face-to-face environments. Journal Scientific MQRInvestigar, 9(2), 1–29. https://doi.org/10.56048/MQR20225.9.2.2025.e478

Markee, N. (2008). Toward a learning behavior tracking methodology for CA-for-SLA. Applied Linguistics, 29(3), 404–427.

Narendran, N., Vairavan, C., & Kumar, P. K. (2025). The growth of digital literacy and 21st-century skills in English language education through ICT. Advances in Consumer Research, 2(5), 1507–1512. https://acr-journal.com/

Pérez-Jorge, D., Olmos-Raya, E., González-Contreras, A. I., & Pérez-Pérez, I. (2025). Technologies applied to education in the learning of English as a second language. Frontiers in Education, 10, Article 1481708, 1–13. https://doi.org/10.3389/feduc.2025.1481708

Seidlhofer, B. (2011). Understanding English as a Lingua Franca. Oxford University Press.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Yusuf, O., Jegede, O., Aliyu, A., & Aladetan, T. (2024). Prospects and challenges of technology enhanced language learning: Teachers perspectives. Studies in Literature and Language, 29(1), 21–32. https://doi.org/10.3968/13571

Zanke, C. (2018). Sociolinguistic realism in pronunciation pedagogy: Accents, intelligibility, and technology. Applied Linguistics Review, 9(1), 1–22.

Boudenna. B & Nasri .K (2025), The Effects of MALLS -Mobile-Assisted Language Learning on Learner's Autonomy in an EFL Context Les effets de MALLS -Apprentissage des langues assisté par mobile sur l'autonomie de l'apprenant dans un contexte EFL ;Dafatir Journal of the Algerian Poetics Laboratory, University of M'sila, Vol. 9, No. 1, pp. 379–389.