Mustapha Bayaro, Amina Shehu Abdullahi, & Philip Abayomi Olorunfemi (2026). Dynamic accent accommodation: Training learners for world Englishes using generative speech. *Academos Journal, Vol. 2, No.3*, 2026, 76-88

Original scientific paper

UDC: 811.111'34]:37.091.3:004.8

Received: 09.08.2026.
Revised: 25.08.2026.
Accepted: 04.09.2026.

DOI: https://doi.org/10.65921/ny1hcf55

Dynamic accent accommodation: Training learners for world Englishes using generative speech

Mustapha Bayaro, Amina Shehu Abdullahi, Philip Abayomi Olorunfemi

[1]Department of Curriculum Studies and Educational Technology, Usmanu Danfodiyo University Sokoto, Nigeria, mustapha.bayaro@udusok.edu.ng, https://orcid.org/0009-0004-6642-3425
[2]Department of Adult Education and Extension Services, Usmanu Danfodiyo University Sokoto, Nigeria, aminashika2020@gmail.com, https://orcid.org/0009-0005-3867-9594
[3]Department of English Language Education, Lakeside University College, Zambia
philipabayomi1@gmail.com, https://orcid.org/0009-0000-2092-2406

Abstract: Traditional English Language Teaching (ELT) and Computer-Assisted Language Learning (CALL) have largely privileged native-speaker pronunciation models, despite the fact that much international communication occurs among speakers from Outer and Expanding Circle English contexts. This conceptual paper proposes Dynamic Accent Accommodation (DAA), a theoretically grounded framework that integrates generative speech synthesis with World Englishes pedagogy to develop learners' receptive flexibility across diverse English accents. Drawing on World Englishes theory, Communication Accommodation Theory, and the Lingua Franca Core, the framework introduces a three-phase Graduated Exposure Model; Baseline Calibration, Controlled Phonological Shift, and Conversational Complexity supported by an adaptive human-in-the-loop speech generation architecture. The paper explains the technical logic of accent manipulation via an Acoustic Accent Vector (α), pedagogical task sequencing, and proposed evaluation metrics for measuring auditory processing efficiency, decoding accuracy, and communicative confidence. It also addresses ethical issues related to algorithmic bias, accent authenticity, voice privacy, and linguistic equity. The principal contribution is the formulation of a pedagogically operational architecture intended to guide future empirical research on AI-supported listening instruction and Global Englishes education. Instead of promoting accent imitation, DAA emphasizes intelligibility, perceptual adaptation, and inclusive multilingual communication.
Keywords: World Englishes; Generative Speech Synthesis; Dynamic Accent Accommodation; Communication Accommodation Theory; Computer-Assisted Language Learning.

Introduction

The landscape of global communication has changed substantially, yet many English language teaching practices continue to rely on monolithic native-speaker pronunciation models (Bayaro et al., 2026; Olorunfemi, 2026). English now functions as a global lingua franca across multilingual and multicultural settings, where successful communication depends increasingly on listeners' ability to understand diverse English varieties rather than approximate a single prestige accent. Although research in World Englishes and Global Englishes has challenged the native-speaker benchmark, this perspective has not been adequately operationalized within AI-supported listening pedagogy. The central gap addressed in this paper is the absence of an adaptive instructional framework that systematically develops receptive accommodation across multiple English accents using generative speech technologies (Blair, 2020; Rose et al., 2021; Wang et al., 2023).

When language learners are exposed exclusively to standardized, native-speaker auditory input, their perceptual systems undergo hyper-specific acoustic tuning. The human auditory cortex adapts to recognize phonemic boundaries, stress patterns, and intonational contours based entirely on the narrow statistical distributions of these standard varieties. When such learners enter real-world communicative spaces and encounter non-native regional varieties, this rigid perceptual calibration fails. Learners experience acute cognitive overload, manifested

as severe drops in listening comprehension, extended auditory processing latency, and elevated affective filters. The mental energy expended simply trying to decode unfamiliar segmental and suprasegmental variations depletes working memory capacity, leaving inadequate cognitive resources for higher-level semantic processing, pragmatic interpretation, and responsive turn-taking (Olorunfemi, 2025).

Traditional language learning environments have attempted to address this gap by incorporating authentic audio recordings of non-native speakers into listening curricula. However, static audio repositories such as pre-recorded audio tracks, podcasts, or video clips suffer from insurmountable structural limitations. They are expensive to curate, difficult to update, fixed in speed and complexity, and incapable of dynamic adaptation to individual learner needs. Furthermore, static audio tracks cannot provide targeted, real-time scaffolding that isolates specific phonological features for perceptual recalibration (Bayaro et al., 2026).

To overcome these limitations, language education requires a fundamental paradigm shift that marries advanced sociolinguistic theory with emerging computational capabilities. The research gap that motivates DAA can be decomposed into three interrelated dimensions. Pedagogically, existing CALL materials and accent-familiarisation programmes typically rely on static audio corpora or limited recorded samples; they cannot systematically vary the intensity of accent features, isolate individual phonological targets, or adapt difficulty in real time to a learner's Zone of Proximal Development. Thus, learners rarely receive graduated, feature-focused perceptual training that builds from low to high accent load while preserving intelligibility. Technologically, most generative speech and zero-shot voice-conversion systems have been optimised for production quality or automated pronunciation assessment against native-speaker norms; they have not been architected as pedagogically controllable engines that expose a continuous Acoustic Accent Vector (α), inject controllable environmental noise (SNR), and close the loop with learner performance analytics.

Theoretically, while World Englishes, Communication Accommodation Theory (CAT), and the Lingua Franca Core (LFC) each supply essential insights, no existing framework has integrated them into a single operational architecture that simultaneously justifies accent diversity (Kachru), specifies the cognitive mechanism of receptive accommodation (CAT), and defines the phonological boundaries of intelligibility (LFC). Cognitive Load Theory and second-language acquisition principles further inform the sequencing of the Graduated Exposure Model (GEM), ensuring that acoustic challenge remains within manageable cognitive limits (Sweller, 2011). Within DAA, the three theoretical pillars occupy distinct structural roles. Kachru's Concentric Circles model supplies the sociolinguistic mandate: the generative engine must be capable of authentic Outer- and Expanding-Circle varieties rather than treating them as deviations. CAT supplies the learner-side mechanism: receptive accommodation is conceptualized as a trainable capacity for real-time perceptual convergence. The LFC supplies the algorithmic constraint: the engine may freely manipulate non-essential regional features while preserving essential intelligibility parameters. Cognitive Load Theory and SLA principles govern the pedagogical sequencing of GEM, determining the progression of α values and task complexity across the three phases (Sweller, 2011). Together these components address the gap left by prior accent-familiarization and perceptual-training procedures, which have lacked adaptive generative control, explicit theoretical integration, and a clear separation of receptive from productive goals.

By operationalizing these theoretical pillars, the primary objective of this paper is to detail the technical, pedagogical, and evaluative architecture of Dynamic Accent Accommodation. This paper addresses three core conceptual research questions:

i. How can deep learning speech synthesis models be architected to dynamically generate authentic, pedagogically controlled multi-accented audio across World Englishes contexts?

ii. What instructional frameworks and micro-pedagogical task designs effectively operationalize receptive Communication Accommodation Theory to accelerate auditory processing speed and comprehension?

iii. What standardized quantitative, qualitative, and sociolinguistic criteria are required to evaluate the efficacy, ethical integrity, and equity of generative speech interventions in language education?

Materials and Methods

Generative Speech System Architecture

The architectural and pedagogical model presented below was derived through a purposive synthesis of three literature streams. First, core theoretical sources on World Englishes (Kachru, 1992; Rose et al., 2021; Blair, 2020), Communication Accommodation Theory (Giles & Edwards, 2025; Derwing & Munro, 2015), and the Lingua Franca Core (Jenkins, 2000) were selected for their direct relevance to intelligibility and receptive accommodation. Second, recent technical literature on neural text-to-speech, diffusion-based generation, and zero-shot voice conversion (Tan et al., 2024; Liu, 2024; Soleymanpour et al., 2024) was examined to identify controllable latent representations suitable for pedagogical accent manipulation. Third, cognitive-load and second-language listening research (Mattys et al., 2012; Pichora-Fuller et al., 2016) informed the design of graduated exposure and noise-masking parameters. Sources were summarised thematically and mapped onto the three functional layers of DAA (input processing, generative latent engine with Acoustic Accent Vector α, and adaptive scaffolding loop) and onto the three phases of the Graduated Exposure Model.

The practical implementation of Dynamic Accent Accommodation requires a speech generation infrastructure capable of producing authentic, controllable, and pedagogically adaptive accent variation. Recent advances in neural text-to-speech synthesis, diffusion-based speech generation, and zero-shot voice conversion have significantly improved the naturalness, prosodic control, and speaker adaptation capabilities of AI speech systems (Tan et al., 2024; Liu, 2024). However, most educational applications continue to prioritize pronunciation production and automated assessment rather than receptive adaptation to accent diversity. The proposed DAA architecture addresses this gap by integrating acoustic control, adaptive scaffolding, and learner analytics into a unified pedagogical system.

The Input Processing Layer accepts raw text inputs representing authentic communicative scenarios, such as international business negotiations, academic panel discussions, or informal social interactions. These scripts are processed through a linguistic analyzer that marks syntactic boundaries, nuclear stress points, and pragmatic intonation targets consistent with Lingua Franca Core parameters. The core of the architecture resides in the Generative Latent Engine. This engine utilizes deep latent feature manipulation, separating speaker identity, text content, and acoustic accent characteristics into decoupled vector spaces. Leveraging advanced neural architectures such as diffusion-based acoustic models and zero-shot latent voice cloning systems the engine can project target accent characteristics onto synthesized speech without requiring thousands of hours of region-specific training audio (Liu, 2024; Soleymanpour et al., 2024).

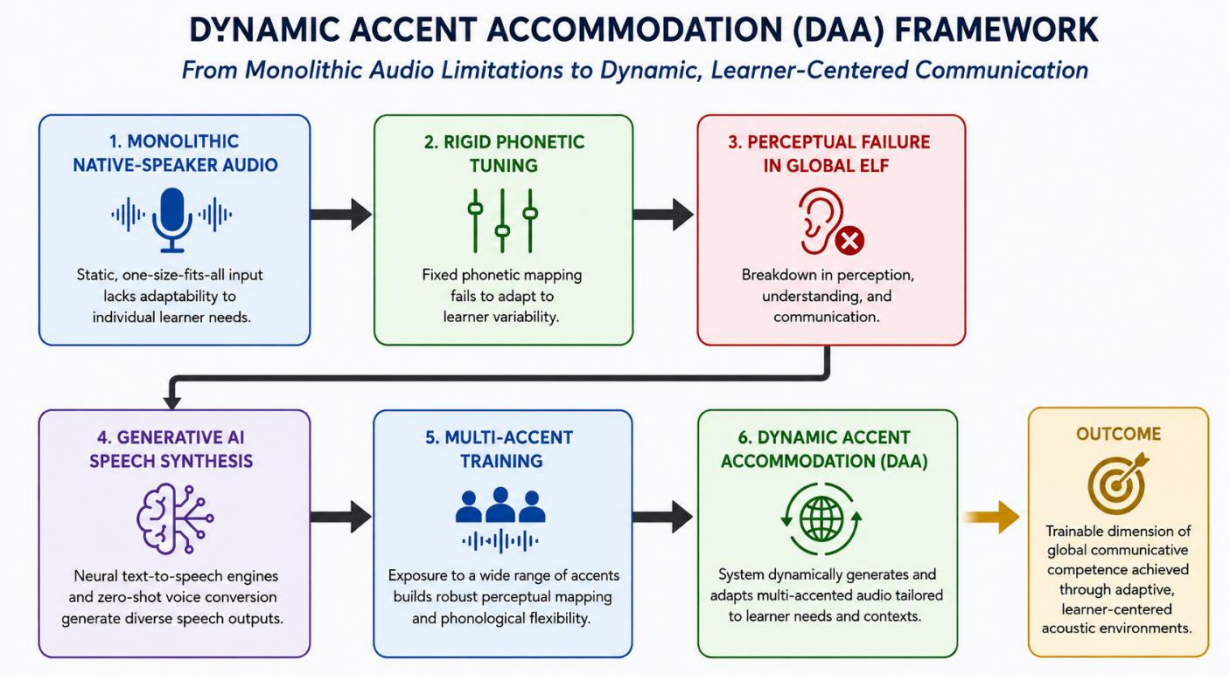

Figure 1. Dynamic Accent Accommodation Framework (Sources: Authors' own elaboration)

This framework contributes a pedagogically operational architecture that connects sociolinguistic theory with AI-enabled listening instruction and provides a foundation for future experimental validation. Kachru's Concentric Circles model provides the foundational sociolinguistic orientation for DAA. By categorizing English speakers into the Inner Circle (norm-providing), Outer Circle (norm-developing), and Expanding Circle (norm-dependent), Kachru (1992) established that regional varieties of English are not imperfect approximations of a singular standard, but are legitimate, rule-governed, and culturally vibrant linguistic systems. In the context of language learning technology, Kachru's model challenges the historical hegemony of Inner-Circle acoustic models. Recent scholars in Global Englishes Language Teaching emphasize that true communicative competence requires receptive flexibility across all three circles (Blair, 2020; Rose et al., 2021). The DAA framework operationalizes this theoretical stance by embedding the full spectrum of World Englishes directly into the acoustic generation engine.

While Kachru (1992) provides the sociolinguistic justification for diversity, Giles' Communication Accommodation Theory (CAT) provides the cognitive and behavioral mechanism governing listener adaptation. CAT posits that interlocutors continuously adjust their communicative behaviors through strategies of convergence, divergence, and maintenance to manage social distance and optimize communicative efficiency. While CAT historically focused on productive accommodation, recent extensions to listening psychology highlight the concept of receptive accommodation (Derwing & Munro, 2015; Mohammadreza & Okim, 2019). Receptive accommodation refers to a listener's willingness and cognitive capacity to adjust their perceptual filters in real time to decode unfamiliar speech variants without experiencing communicative breakdown. The DAA framework conceptualizes receptive accommodation as a plastic, highly trainability skill. By systematically exposing the auditory system to variable acoustic inputs, DAA trains the brain to execute real-time perceptual convergence, broadening the listener's phonemic acceptance bands and mitigating the shock of unfamiliar prosodic structures.

The operational boundaries of this perceptual training are defined by Jenkins' Lingua Franca Core (LFC). Formulated through extensive empirical observation of English as a Lingua Franca interactions, the LFC distinguishes between phonological features that are critical for

maintaining mutual intelligibility and those that are non-essential. Essential LFC features include consonant quantity, maintenance of vowel length distinctions, and correct placement of nuclear stress. Non-essential features, conversely, include the substitution of dental fricatives (such as pronouncing the voiced and voiceless "th" sounds as stops /t, d/) and regional pitch variations. Traditional ELT over-emphasizes the elimination of non-essential features, treating harmless regional pronunciations as critical errors. The generative engine is designed to manipulate non-essential regional variants to build perceptual tolerance, while preserving essential LFC parameters to ensure that synthesized speech remains fundamentally intelligible.

Environmental Noise Masking and Signal-to-Noise Ratio (SNR) Control

A critical oversight in legacy Computer-Assisted Language Learning (CALL) research is the assumption that acoustic processing occurs within pristine, acoustically isolated environments. In real-world international discourse, non-native accents are rarely encountered in studio-quality quiet; rather, they are embedded within noisy, dynamic environments such as international airport terminals, bustling industrial sites, or digital video conferences compromised by bandwidth compression and packet loss. When unfamiliar phonetic realization coincides with degraded signal quality, the listener's cognitive load increases exponentially, often precipitating complete communicative breakdown (Mattys et al., 2012; Strori et al., 2020). To prepare learners for these multi-layered acoustic challenges, the Generative Latent Engine incorporates an Environmental Noise Masking sub-module. This module dynamically injects controllable Signal-to-Noise Ratios (SNR) measured in decibels (dB) alongside ambient background interference (e.g., babble noise, HVAC hum, digital jitter) directly into the synthesized audio pipeline. By systematically lowering the SNR parameter in tandem with adjustments to the Acoustic Accent Vector (α), the system operationalizes what speech perception literature terms perceptual resilience (Pichora-Fuller et al., 2016; Keerstock & Smiljanic, 2019). This dual-stress training forces the learner’s auditory cortex to rely on top-down semantic constraints and robust phonemic restoration mechanisms, ensuring that dynamic accommodation skills remain functional under adverse, real-world listening conditions.

To achieve precise pedagogical control, the system introduces a continuous mathematical scalar known as the Acoustic Accent Vector (α), bounded between 0.0 and 1.0: $\alpha \in [0.0, 1.0]$ When $\alpha = 0.0$, the engine generates a baseline, highly standardized acoustic model (e.g., standard General American). As α increases toward 1.0, the engine progressively injects region-specific acoustic features associated with target World Englishes varieties, such as Nigerian English, Indian English, or Singaporean English. The manipulation of the Acoustic Accent Vector (alpha) governs three primary phonetic and prosodic transformations:

First, the system manages Segmental Realization shifts. At lower values of α, consonant clusters and vowel qualities remain aligned with standard global baselines. As α advances, the system introduces targeted segmental variations characteristic of specific outer-circle regions. For instance, in synthesizing a Nigerian English acoustic profile, the engine executes dental fricative stopping, shifting voiceless /θ/ to voiceless alveolar stop /t/ and voiced /ð/ to voiced alveolar stop /d/. Simultaneously, monophthongization is applied to specific diphthongs (e.g., shifting /eɪ/, /e/ and /oʊ/ /o/), reflecting authentic regional phonological patterns (Udofot, 2003; Gut, 2008).

Second, the engine executes Suprasegmental Rhythm Adjustments. Standard British and American Englishes are prototypically stress-timed, characterized by significant duration contrasts between stressed and unstressed syllables, alongside extensive vowel reduction in unstressed positions. Conversely, many Outer and Expanding Circle Englishes such as Singaporean, West African, and Indian varieties exhibit syllable-timed or mora-timed prosodic tendencies, where syllables recur at relatively equal temporal intervals and full vowel quality

is preserved across unstressed positions (Gardiner & David, 2020; Fuchs, 2023). The DAA latent engine dynamically modifies durational trajectories across synthesized speech segments. By dampening the duration contrast between stressed and unstressed vowels and reducing schwa substitutions, the engine transforms the rhythm from stress-timed to syllable-timed, training the learner's auditory system to process altered temporal pacing.

Third, the system modulates Pitch Contour and Intonational Trajectories. In many global varieties of English, intonational movement is realized through discrete tone shifts on nuclear syllables rather than the sweeping pitch contours typical of Inner-Circle varieties. The generative engine alters pitch contours across macro-rhythmic phrases, exposing learners to pitch movements that might otherwise be misinterpreted as abrupt or pragmatically inappropriate by listeners tuned exclusively to native standards.

Crucially, the architecture incorporates an Adaptive Scaffolding Loop that transforms the system from a static generator into a human-in-the-loop, responsive learning engine. As the learner interacts with synthesized audio prompts during listening tasks, the system continuously tracks performance metrics, specifically phonetic decoding accuracy and processing latency (response time). If the analytics engine detects a spike in decoding errors or cognitive delay when processing a specific phonological feature (e.g., syllable-timed prosody in South Asian English), the adaptive algorithm dynamically recalibrates the Acoustic Accent Vector (α). The algorithm lowers α to provide scaffolding, isolating the troublesome feature, and then gradually increments α across subsequent items as performance stabilizes. This closed-loop design ensures that the acoustic challenge remains within the learner's Zone of Proximal Development, preventing both cognitive boredom from overly simplified input and affective shutdown from overwhelming complexity.

The Graduated Exposure Pedagogical Model

Technology alone cannot transform learning outcomes without a structured instructional framework. To operationalize the system architecture described above, this paper formulates the Graduated Exposure Model (GEM), a pedagogical framework designed to systematically build receptive accommodation capacity in L2 learners. Grounded in cognitive load theory and second language acquisition principles, GEM structures learning activities into three progressive, highly targeted phases: Baseline Calibration, Controlled Phonological Shift, and Conversational Complexity.

Phase 1, termed Baseline Calibration, focuses on establishing stable cognitive anchor points while introducing low-level acoustic variation. In this phase, the Acoustic Accent Vector is set to a low intensity ($\alpha = 0.2$ to 0.3). Learners are exposed to audio prompts featuring local or highly familiar non-native varieties. The pedagogical goal of Phase 1 is not to challenge the auditory system with radical phonological shifts, but to desensitize the learner to the presence of non-native voice quality features. Listening tasks in this phase focus on high-frequency lexical items and familiar contextual themes. By ensuring that semantic content remains transparent, Phase 1 reduces affective anxiety and builds initial self-efficacy, establishing the foundational mindset required for dynamic accommodation.

Phase 2, designated Controlled Phonological Shift, advances the instructional challenge by isolating specific phonological deviations mapped directly to the Lingua Franca Core. In this phase, the system operates at moderate accent weights ($\alpha = 0.4$ to 0.6). Rather than presenting chaotic, unanalyzed audio, the GEM framework introduces systematic, targeted phonological shifts one feature at a time. For instance, a module may focus specifically on consonant cluster simplification in East Asian Englishes or vowel length neutralization in African Englishes.

The micro-pedagogies employed in Phase 2 utilize specialized auditory training techniques designed to recalibrate phonemic categories:

i. Perceptual Shadowing Tasks: Learners listen to short generative prompts exhibiting

targeted accent variations and immediately repeat the utterance aloud. This real-time auditory-to-articulatory mapping forces the brain to process the precise acoustic structure of the input, accelerating the formation of flexible mental representations for non-standard phonemes.

ii. Differential Minimal-Pair Cloze Exercises: Learners listen to sentences containing words where non-standard segmental realizations create potential lexical ambiguities (e.g., distinguishing between "thin" /θɪn/ synthesized as "tin" /tɪn/ in varieties exhibiting dental stop substitution). Learners must select the intended lexical item based on context, directly training the brain to rely on semantic constraint when phonetic cues diverge from standard norms.

iii. Auditory Recalibration Drills: Learners are presented with the same underlying sentence synthesized across an escalating gradient of accent vectors ($\alpha = 0.2 \longrightarrow 0.4 \longrightarrow 0.6 \longrightarrow 0.8$). This gradual acoustic morphing allows the auditory cortex to track structural transformations in real time, developing explicit cognitive awareness of how specific sounds shift across regional boundaries.

Discussion

Response to Research Questions

The conceptual architecture responds directly to the three research questions posed in the Introduction. Research Question 1 asked how deep-learning speech synthesis models can be architected to generate authentic, pedagogically controlled multi-accented audio. The proposed Generative Latent Engine, with its decoupled latent spaces and continuous Acoustic Accent Vector (α [0.0, 1.0]), together with the Environmental Noise Masking module, supplies a controllable mechanism for generating Outer- and Expanding-Circle varieties while preserving Lingua Franca Core intelligibility parameters. Research Question 2 asked what instructional frameworks and micro-pedagogical tasks operationalize receptive Communication Accommodation Theory. The Graduated Exposure Model (GEM). Baseline Calibration ($\alpha = 0.2$–0.3), Controlled Phonological Shift ($\alpha = 0.4$–0.6), and Conversational Complexity ($\alpha = 0.7$–1.0) together with perceptual shadowing, differential minimal-pair cloze, auditory recalibration drills, rapid-switching turn-taking, and noise-masked decoding, provides a sequenced answer grounded in cognitive-load and SLA principles. Research Question 3 asked what criteria are required to evaluate efficacy, ethical integrity, and equity. The proposed multi-dimensional matrix comprising APRT, DPDA, affective/WTC measures, and the ethical guidelines (authentic data sourcing, participatory design, compensation and data sovereignty, voice privacy and consent, voice-cloning safeguards, and sociolinguistic scaffolding) constitutes the evaluative framework. All metrics and mechanisms remain proposed and await empirical validation.

The Receptive-Productive Boundary: Preventing Unintended Phonetic Drift

A vital theoretical nuance that distinguishes Dynamic Accent Accommodation from conventional pronunciation frameworks is the strict conceptual separation between receptive processing capacity and productive phonetic modification. A known risk in intensive perceptual training is the phenomenon of uncontrolled phonetic drift or accidental productive over-accommodation, wherein a learner unintentionally begins to mimic the non-native acoustic, prosodic, or segmental features of their interlocutor (Chang, 2019; Piazza et al., 2023). In cross-cultural interactions, uncalibrated productive mimicry by an L1 or advanced L2 speaker can be misconstrued as patronizing, culturally insensitive, or mockingly performative (Giles & Edwards, 2025; Mohammadreza & Okim, 2019). The Graduated Exposure Model explicitly safeguards against this by establishing a firm Receptive-Productive Boundary. While interactive tasks (such as Perceptual Shadowing and Cloze Drills) require internal auditory-to-

articulatory mapping for perceptual recalibration, the system's automated evaluation metrics assess exclusively listener decoding accuracy and processing speed, never rewarding productive accent imitation. By decoupling receptive convergence from productive mimicry, the framework trains learners to effortlessly decode diverse World Englishes without compromising their own authentic, highly intelligible productive speech identity.

Phase 3, titled Conversational Complexity, represents the advanced operational stage of the GEM framework. In this phase, the system operates at high accent weights ($\alpha = 0.7$ to 1.0) and simulates authentic, unscripted global communicative environments. Real-world international interactions rarely involve a single speaker communicating at a measured cadence. Instead, they feature multi-party discussions, rapid speaker switching, overlapping speech, background noise, and varying degrees of pragmatic directness. To prepare learners for these complex environments, Phase 3 deploys multi-speaker dialogic simulations. The generative engine synthesizes real-time conversations involving multiple AI speech agents, each programmed with distinct World Englishes acoustic profiles (e.g., a multi-party business meeting featuring a Nigerian executive, an Indian software architect, a German logistics director, and a Brazilian marketing specialist).

The micro-pedagogies in Phase 3 emphasize macro-comprehension and real-time processing under pressure:

i. Rapid-Switching Turn-Taking Drills: The learning interface presents fast-paced dialogue exchanges where the active speaker shifts every few seconds, requiring the learner's auditory processor to rapidly recalibrate to a new acoustic accent profile without losing the thread of the conversation.

ii. Interactive Noise-Masked Decoding: To simulate real-world communicative conditions, the generative engine overlays synthesized multi-accent dialogues with ambient environmental noise (e.g., airport terminal chatter, conference hall reverberation). Learners must extract key semantic information despite competing acoustic distractions, testing the robust integration of their perceptual accommodation skills.

iii. Bidirectional Accommodation Simulations: Learners actively participate in dialogue simulations using their own voices. The generative AI speech agent evaluates the learner's input and adjusts its synthesized responses, creating a dynamic conversational loop where both human learner and machine agent continuously negotiate mutual intelligibility.

iv. By progressing sequentially through Baseline Calibration, Controlled Phonological Shift, and Conversational Complexity, the Graduated Exposure Model ensures that receptive accommodation capacity is systematically built, reinforced, and automated.

Sociolinguistic Implications

To transition Dynamic Accent Accommodation from a conceptual model into an empirically validated, ethically grounded educational framework, rigorous evaluation standards must be established. Traditional language learning assessments are themselves heavily compromised by native-speaker bias, evaluating learner success through standardized listening tests recorded exclusively by British or American voice actors. Such tests fail to measure receptive accommodation capacity, as they test only how well a learner decodes a singular, highly artificial acoustic standard. Evaluating DAA requires a multi-dimensional assessment framework combining quantitative processing metrics, qualitative affective scales, and sociolinguistic equity audits.

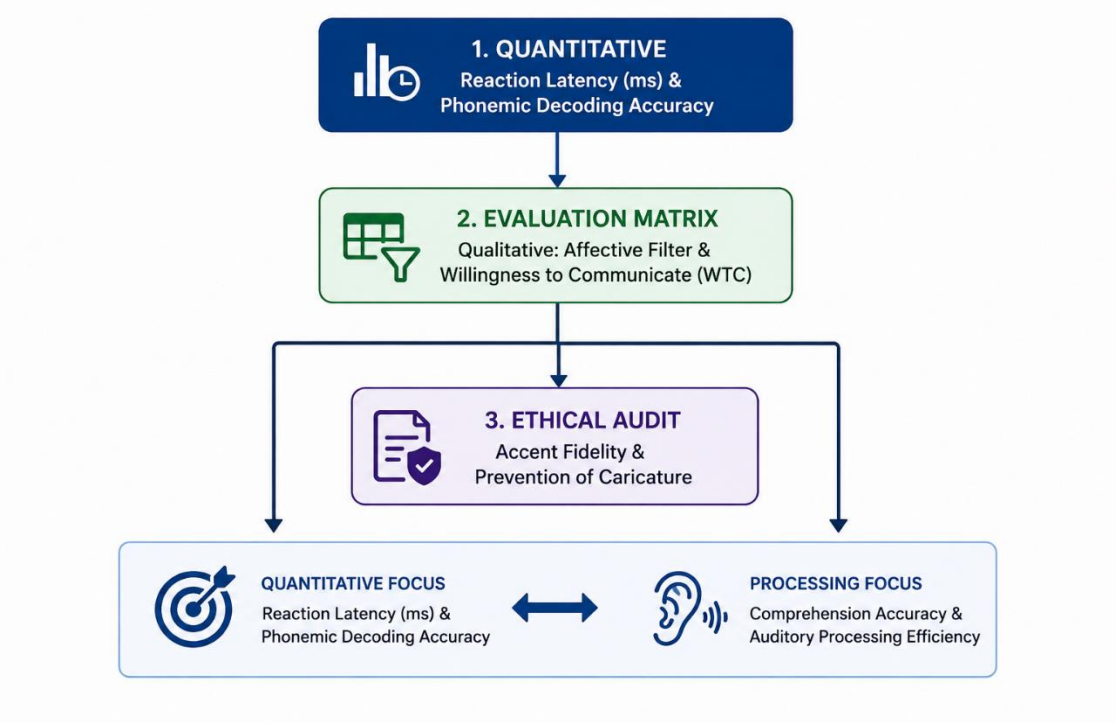

Figure 2. Evaluating DAA (Source: Authors' own elaboration)

Figure 2 summarizes the proposed multi-dimensional evaluation matrix, integrating quantitative processing metrics (APRT, DPDA), qualitative affective measures, and sociolinguistic equity audits. Quantitative evaluation of receptive accommodation must focus on measuring both comprehension accuracy and auditory processing efficiency. Processing efficiency is critical because a learner who understands a non-native accent only after ten seconds of intense mental reconstruction will fail in real-time, fast-paced international conversations.

The proposed assessment matrix establishes two core quantitative metrics:

First, Auditory Processing Reaction Time (APRT), measured in milliseconds (ms). APRT records the temporal latency between the precise acoustic offset of a multi-accented auditory prompt and the learner's correct selection of a semantic or phonological choice on a digital testing interface. A significant baseline latency indicates high cognitive friction and manual decoding effort. As receptive accommodation becomes automated through DAA training, APRT curves should demonstrate a steep negative slope, signaling that the auditory cortex is decoding non-standard input with minimal cognitive delay.

Second, Differential Phonemic Decoding Accuracy (DPDA), calculated as a percentage score across specific World Englishes categories. Rather than aggregating listening scores into a single monolithic grade, DPDA breaks performance down into region-specific sub-scores (e.g., West African English Decoding Accuracy, South Asian English Decoding Accuracy). This diagnostic granularity enables educators and adaptive algorithms to pinpoint specific perceptual blind spots in a learner's acoustic repertoire.

Qualitative evaluation must assess psychological and affective transformations within the learner. In alignment with MacIntyre (2007) Willingness to Communicate (WTC) framework, DAA evaluations must measure reductions in accent-induced communicative anxiety. When learners feel terrified of encountering unfamiliar accents, their affective filter rises, causing them to withdraw from global interactions. Standardized Likert-scale instruments, combined with semi-structured qualitative interviews, should track changes in learner self-efficacy, perceived listening ease, and readiness to engage in translingual contexts following DAA

intervention. Effective DAA training should manifest as a significant drop in accent anxiety and a corresponding surge in global WTC scores. Beyond technological and pedagogical metrics, the implementation of generative speech technology in language education carries profound sociolinguistic and ethical implications. Generative artificial intelligence systems are notorious for reflecting and amplifying the structural biases embedded within their training datasets (Weidinger et al., 2022; Bender et al., 2021). If an AI speech engine is trained on low-quality, limited, or biased voice samples, it risks synthesizing acoustic caricatures exaggerated, unnatural, or mocking representations of regional accents that reinforce harmful linguistic stereotyping. Auditory Processing Reaction Time (APRT) and Differential Phonemic Decoding Accuracy (DPDA) are proposed metrics; their reliability, validity, and sensitivity remain to be established through future empirical pilot studies.

Normative Assessment AI vs. Accommodative Pedagogy AI

To fully grasp the disruptive novelty of Dynamic Accent Accommodation, it must be positioned in direct theoretical opposition to the prevailing paradigm of commercial automated language assessment. Contemporary speech-evaluation AI embedded in mainstream language learning applications and high-stakes standardized testing platforms operates almost exclusively on Normative Assessment AI models. These systems utilize Automatic Speech Recognition (ASR) acoustic models trained primarily on native-speaker datasets, explicitly penalizing non-native stress patterns, vowel shifts, and consonant cluster reductions as "errors" or "deviations" (Galaczi, & Pastorino-Campos, 2025; Litman et al., 2018). This architectural bias reinforces linguistic hegemony by forcing non-native speakers to conform to an Anglo-American acoustic monolith while simultaneously training human listeners to view non-standard accents as deficient (Lippi-Green, 2012). In contrast, the framework proposed here establishes an Accommodative Pedagogy AI paradigm. Rather than using AI as a punitive gatekeeper that demands speaker conformity, Accommodative Pedagogy AI harnesses deep generative networks to expand human listener flexibility. By shifting the computational burden of adaptation from the speaker to the listener, this framework fundamentally reconfigures the sociolinguistic role of artificial intelligence in education moving from an instrument of linguistic assimilation to a tool for global linguistic equity.

To safeguard against these risks, DAA implementations must adhere to strict ethical guidelines:

i. Authentic Data Sourcing and Participatory Design: Generative accent models must be built using extensive high-fidelity voice data contributed directly by native speakers from Outer and Expanding Circle regions. Linguistic communities must be involved as co-designers, ensuring that synthesized acoustic profiles accurately represent authentic, prestigious regional varieties rather than offensive stereotypes.

ii. Fair Compensation and Data Sovereignty: AI developers must ensure fair financial compensation and clear data sovereignty rights for speakers whose voice data is utilized to train neural speech engines, preventing digital extraction and exploitation of global South voice talents.

iii. Explicit Sociolinguistic Scaffolding: DAA software must integrate explicit sociolinguistic commentary into the learner interface. The software should clearly inform learners that Outer and Expanding Circle varieties are legitimate, highly sophisticated linguistic systems integral to global trade and culture, explicitly counteracting historical narratives of linguistic deficit.

iv. By adhering to these ethical guardrails, Dynamic Accent Accommodation transforms generative AI from a potential tool of linguistic homogenization into a powerful engine for decolonizing language education. Historically, language teaching technology has enforced Western native-speaker dominance, compelling millions of learners worldwide to conform to a narrow linguistic ideal. DAA flips

this paradigm, using advanced computational power to celebrate, validate, and operationalize global linguistic diversity.

Limitations

As a conceptual contribution, this paper does not report empirical results; all mechanisms, metrics, and expected outcomes are proposed and require future validation. Several limitations must therefore be acknowledged. First, the authenticity of synthesized accents remains an open question. Even high-fidelity generative models may produce acoustic profiles that diverge from the statistical distributions of natural speech in a given variety; community validation and participatory design are essential but do not fully eliminate the risk of residual artificiality or stereotyping. Second, World Englishes varieties are internally heterogeneous. A single Acoustic Accent Vector continuum cannot capture the full range of sociolectal, regional, and individual variation within, for example, Nigerian or Indian English. The framework currently treats varieties at a relatively coarse level of granularity. Third, the validity of the continuous α parameter as a psychologically meaningful dimension of accent "strength" has not been established. Whether linear interpolation in latent space corresponds to linear changes in perceived accentedness or processing difficulty is an empirical question. Fourth, transfer of trained receptive skills to spontaneous, multi-party, real-world interaction cannot be assumed. Laboratory or platform-based gains in APRT or DPDA may not generalize to high-stakes communicative settings without additional ecological validation. Fifth, despite ethical safeguards, the risk of accent stereotyping persists if training data are unbalanced or if interface design inadvertently frames non-Inner-Circle varieties as marked or deficient. Sixth, access to the computational resources, bandwidth, and devices required for interactive generative-speech training is unevenly distributed; without deliberate attention to low-resource deployment, DAA risks reinforcing rather than reducing educational inequities.

Future Research

Future empirical work can operationalize the proposed metrics in small-scale pilot studies. For example, a within-subjects pilot might expose intermediate L2 learners ($N \approx 20–30$) to a two-week GEM sequence (Phases 1–2) targeting one Outer-Circle variety, measuring APRT and DPDA before and after training on matched pre-/post-test item banks. A second pilot could compare noise-masked versus clean-speech conditions at fixed α levels to test the contribution of Environmental Noise Masking to perceptual resilience. Qualitative interviews and WTC scales administered at both time points would complement the quantitative indices. Such pilots would supply the first empirical evidence of feasibility, effect-size estimates, and metric sensitivity before larger controlled trials.

Conclusion

Dynamic Accent Accommodation offers a theoretically grounded conceptual framework for aligning language education more closely with the multilingual realities of contemporary global communication. By integrating World Englishes, Communication Accommodation Theory, and the Lingua Franca Core with adaptive generative speech technologies, the framework shifts pedagogical emphasis from accent conformity toward receptive flexibility and intelligibility. The proposed architecture and Graduated Exposure Model provide a coherent basis for future empirical studies examining listening comprehension, processing efficiency, communicative confidence, and equitable AI-supported language learning. The paper's contribution is the formulation of a research agenda that connects sociolinguistic diversity, educational technology, and responsible artificial intelligence within a single pedagogical framework. Realization of the intended benefits remains contingent on rigorous empirical validation, community participation, and attention to the limitations outlined above.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bayaro, M, Olorunfemi, P.A, Michael, C.I, Adeleke, H. A. (2026) Beyond Native Norms: Decolonizing AI Speech Recognition for Accent Equity in Nigerian ESL Classrooms. *Center of Artificial Intelligence*, 2026;*1*(2), 111-121. https://doi.org/10.65591/CAI-206-2026

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21) (pp. 610–623). *Association for Computing Machinery.* https://doi.org/10.1145/3442188.3445922

Blair, A. (2020), Global Englishes for Language Teaching Heath Rose and Nicola Galloway. Cambridge, England: Cambridge University Press, 2019. Pp xix + 255. *TESOL* J, 54: 1144-1146. https://doi.org/10.1002/tesq.586

Chang, C. B. (2019). Language change and linguistic inquiry in a world of multicompetence: Sustained phonetic drift and its implications for behavioral linguistic research. *Journal of Phonetics,* 74, 96–113. https://doi.org/10.1016/j.wocn.2019.03.001

Derwing, T. M., & Munro, M. J. (2015). Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research (Vol. 42). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.42

Fuchs, R. (2023). Speech rhythm in recent research. In R. Fuchs (Ed.), Speech rhythm in learner and second language varieties of English (pp. 1–21). Springer.

Galaczi, E., & Pastorino-Campos, C. (2025). Ethical AI for language assessment: Principles, considerations, and emerging tensions. Annual Review of Applied Linguistics, 45, 294–314. https://doi.org/10.1017/S0267190525100081

Gardiner, I. A., & David, D. (2020). The features of Asian Englishes: Phonology. In K. Bolton, W. Botha, & A. Kirkpatrick (Eds.), The handbook of Asian Englishes. Wiley-Blackwell. https://doi.org/10.1002/9781118791882.ch8

Giles, H., & Edwards, A. L. (2025). Communication Accommodation Theory. In C. A. Chapelle & J. E. Bonnin (Eds.), The Encyclopedia of Applied Linguistics. https://doi.org/10.1002/9781405198431.wbeal20009

Gut, U. (2008). Nigerian English: Phonology. In R. Mesthrie, B. Kortmann, & E. W. Schneider (Eds.), A handbook of varieties of English: Vol. 4. Africa, South and Southeast Asia (pp. 35–54). De Gruyter Mouton. https://doi.org/10.1515/9783110208429.1.35

Jenkins, J. (2000). The phonology of English as an international language. Oxford University Press.

Kachru, B. B. (1992). Teaching world Englishes. In B. B. Kachru (Ed.), The other tongue: English across cultures (pp. 355–365). University of Illinois Press.

Keerstock, S., & Smiljanic, R. (2019). Clear speech improves listeners' recall. *The Journal of the Acoustical Society of America*, 146(6), 4604. https://doi.org/10.1121/1.5141372

Lippi-Green, R. (2012). English with an accent: Language, ideology and discrimination in the United States (2nd ed.). Routledge. https://doi.org/10.4324/9780203348802

Litman, D., Strik, H., & Lim, G. (2018). Speech technologies and the assessment of second language speaking: Approaches, challenges, and opportunities. *Language Assessment Quarterly*, 15(3), 294–309. https://doi.org/10.1080/15434303.2018.1472265

Liu, S. (2024). Zero-shot voice conversion with diffusion transformers. arXiv. https://doi.org/10.48550/arXiv.2411.09943

MacIntyre, P. D. (2007). Willingness to communicate in the second language: Understanding the decision to speak as a volitional process. *The Modern Language Journal*, 91(4), 564–576. https://doi.org/10.1111/j.1540-4781.2007.00623.x

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, 27(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

Mohammadreza, D., & Okim, K. (2019). Listener background in L2 speech evaluation. In IntechOpen. https://doi.org/10.5772/intechopen.89414

Olorunfemi, P.A, Nnaji, C.M & Michael, C.I. (2026). An Investigation into AI-Assisted Aspect-Based and Integrated English Teaching Approaches Using ADLL and CPCT Theories. Center of Artificial Intelligence. 1(2). 84-95. Doi: 10.65591/CAI-147-2026

Olorunfemi, P.A. (2025). Post-Structuralist Critiques of Authorship in Generative Artificial Intelligence (AI) Literature. *Advances Educational Innovation*, *1*(4), 159-168. https://analysisdata.co.id/index.php/aei/article/view/352

Olorunfemi, P.A. (2026). Enhancing Speaking Fluency Through Collaborative Learning: A Case Study of Senior Secondary School Students of Victory Academy Isua, Ondo State, Nigeria. *ERL Journal* - Volume 2025-2(14), *The Active Dimension of Language and of Linguistic Education* ISSN 2657-9774; https://doi.org/10.36534/erlj.2025.02.08

Piazza, G., Kalashnikova, M., & Martin, C. D. (2023). Phonetic accommodation in non-native directed speech supports L2 word learning and pronunciation. *Scientific Reports,* 13, 21282. https://doi.org/10.1038/s41598-023-48648-7

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M. M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, 37(1 Suppl), 5S–27S. https://doi.org/10.1097/AUD.0000000000000312

Rose, H., McKinley, J., & Galloway, N. (2021). Global Englishes and language teaching: A review of pedagogical research. *Language Teaching*, 54(2), 157–189.

Soleymanpour, M., Johnson, M. T., Soleymanpour, R., & Berry, J. (2024). Accurate synthesis of dysarthric speech for ASR data augmentation. *Speech Communication*, 164, 103112. https://doi.org/10.1016/j.specom.2024.103112

Strori, D., Bradlow, A. R., & Souza, P. E. (2020). Recognition of foreign-accented speech in noise: The interplay between talker intelligibility and linguistic structure. *The Journal of the Acoustical Society of America,* 147(6), 3765. https://doi.org/10.1121/10.0001194

Sweller, J. (2011). Cognitive Load Theory. In Psychology of Learning and Motivation (Vol. 55, pp. 37-76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Tan, X., Qin, T., Soong, F., & Liu, T.-Y. (2024). A survey on neural speech synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(6), 4234–4245.

Udofot, I. (2003). Stress and rhythm in the Nigerian accent of English: A preliminary investigation. *English World-Wide*, 24(2), 201–220. https://doi.org/10.1075/eww.24.2.04udo

Wang, Z., Zou, D., Lee, L. K., Xie, H., & Wang, F. L. (2023). A systematic review of generative artificial intelligence in language education. In Proceedings of the International Conference on Computers in Education. https://doi.org/10.58459/icce.2023.1199

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., Biles, C., Brown, S., Kenton, Z., Hawkins, W., Stepleton, T., Birhane, A., Hendricks, L. A., Rimell, L., Isaac, W., & Gabriel, I. (2022). Taxonomy of risks posed by language models. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 214–229). *Association for Computing Machinery*. https://doi.org/10.1145/3531146.3533088