Vol. 4 No. 2 (2026)

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895
https://llrjournal.com/index.php/11

ANALYZING PHONOLOGICAL FEATURES AND PRONUNCIATION CHALLENGES IN SECOND LANGUAGE ENGLISH ACQUISITION

Nawaz Ali[*1], Azam Hussain[2], Muhammad Naeem[3], Imran Ali[4], Zulfiqar Ali Sianch[5]

*[*1]English instructor, Institute of Emerging Technologies, Sukkur IBA university, Khairpur Campus, SIBAU-KHP*
[2]Ph. D. Scholar at TIMES University, Multan
[3]PhD English Scholar at TIMES University Multan, Lecturer, Department of English, Division of Arts and Social Sciences, University of Education Lahore, DG Khan Campus, Pakistan
[4]Instructor English, Institute of Emerging Technology Khairpur Campus, Sukkur IBA University
[5]M. Phil in English Linguistics, Lecturer at Pak-Turk Maarif International School and Colleges

[*1]nawazali.khp@iba-suk.edu.pk, [2]azamhussainkwl@gmail.com, [3]naeem.amin@ue.edu.pk, [4]Imrantalpur.khp@iba-suk.edu.pk, [5]zulfiqarsianch@gmail.com

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895

DOI: http://doi.org/10.5281/zenodo.20373846

Abstract

This study paper examines the major phonological features and pronunciation challenges encountered in second language (L2) English acquisition, with particular emphasis on the interaction between first language (L1) influence, cognitive perception, and instructional strategies. The study explores how learners struggle with English segmental and suprasegmental features, including vowels, consonants, stress, rhythm, intonation, and connected speech. Drawing upon foundational theories such as the Contrastive Analysis Hypothesis (CAH), Markedness Differential Hypothesis (MDH), Speech Learning Model (SLM), and Perceptual Assimilation Model (PAM-L2), the review highlights the importance of perceptual categorization and phonological transfer in shaping learner pronunciation. The paper further discusses common pronunciation difficulties among learners from different linguistic backgrounds, including problems with tense-lax vowel distinctions, consonant substitutions, consonant clusters, and stress-timed rhythm patterns. Particular attention is given to repair strategies such as epenthesis and consonant deletion, which learners employ to adapt unfamiliar English phonotactic structures. In addition, the review evaluates the influence of individual learner variables such as age, motivation, language aptitude, identity, and neuroplasticity on pronunciation attainment. The role of modern pedagogical approaches, especially explicit pronunciation instruction, Computer-Assisted Pronunciation Training (CAPT), and Artificial Intelligence (AI)-based speech recognition tools, is also critically analyzed. The findings indicate that successful L2 phonological acquisition depends on both perceptual and articulatory development, supported by communicative practice and technological intervention. Ultimately, the review emphasizes intelligibility rather than native-like pronunciation as the primary objective of modern English language teaching.

Keywords: *Second Language Acquisition, English Phonology, Pronunciation Challenges, Phonological Transfer, Segmental Features, Suprasegmental Features, Speech Learning Model, CAPT, Intelligibility, L2 Pronunciation*

Introduction

The acquisition of a second language (L2) phonological system represents one of the most formidable challenges in the field of applied linguistics. Unlike syntax or lexicon, which can often be acquired to a high degree of proficiency in adulthood, the mastery of a target language's sound system is frequently obstructed by deeply ingrained articulatory habits and perceptual filters established during the acquisition of the first language (L1) (Darcy & Rocca, 2021). Phonological processes, which encompass systematic changes in sound patterns, occur in both L1 and L2 acquisition, necessitating the establishment of robust segmental and suprasegmental representations

so that learners can process phonologically similar or complex words with ease (Saito, 2023). For the second language learner of English, this journey involves not only the physical reconfiguration of vocal tract movements but also a cognitive restructuring of how speech sounds are categorized and how prosodic hierarchies are constructed (Thomson, 2024).

The ability to comprehend spoken connected speech is a cornerstone of communicative competence, particularly when learners are exposed to the rapid and natural delivery typical of native English speakers. This comprehension is predicated on the learner's sensitivity to both individual phonemes the segments and the broader melodic and rhythmic patterns the suprasegmentals that characterize English speech (Gordon & Darcy, 2025). When these representations are underdeveloped, learners struggle to segment the continuous stream of speech into recognizable lexical units, leading to a breakdown in communication that transcends mere vocabulary limitations (Crowther & Trofimovich, 2021).

Theoretical Foundations of L2 Phonological Acquisition

The scientific inquiry into L2 pronunciation has evolved through several paradigmatic shifts, moving from purely structuralist comparisons to complex cognitive and perceptual models. These theories provide the necessary scaffolding for understanding the systematic nature of pronunciation challenges (Kissling, 2022).

The Evolution from Contrastive Analysis to Markedness

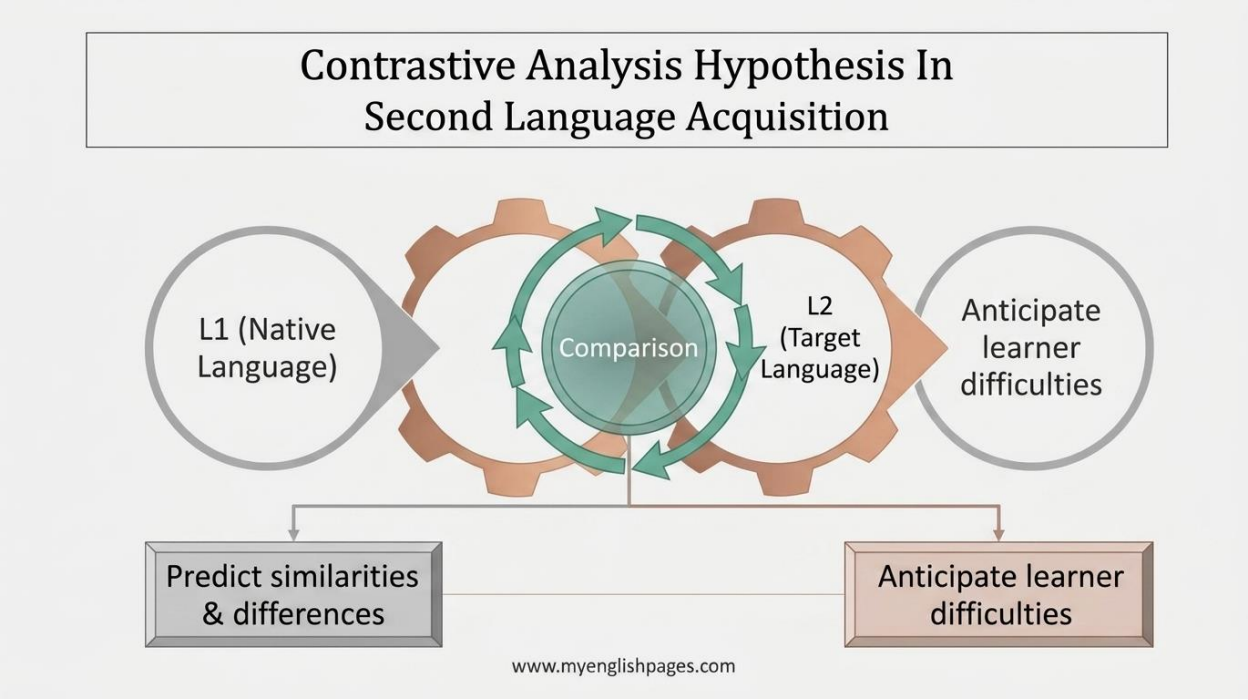

Figure:2 Revised Diagram of the Contrastive Analysis Hypothesis in SLA, Mapping L1-L2 Comparison to Prediction and Prediction to Anticipation of Learner Difficulties

Historically, the study of L2 phonology was dominated by the Contrastive Analysis Hypothesis (CAH), pioneered by Robert Lado. The CAH posited that the difficulties encountered by language learners could be predicted by a systematic comparison of the L1 and the target language (TL). Under the "strong form" of this hypothesis, linguistic differences were viewed as the primary, if not sole, source of interference, while similarities were expected to facilitate "positive transfer" (Kardos & Sato, 2021). However, the strong form of the CAH was eventually found wanting because it could not account for many exceptions; certain predicted difficulties never materialized, while learners across diverse L1 backgrounds often shared similar "developmental" errors that had no basis in their native tongues (Martinez-Flor & Uso-Juan, 2022).

The subsequent "weak form" of the CAH shifted focus toward explaining errors after they occurred (a posteriori) rather than predicting them in advance. This recognition of the limitations of pure structuralism led to the development of the Markedness Differential Hypothesis (MDH) by Fred Eckman. The MDH integrated the concept of typological markedness the idea that certain linguistic structures are more "basic," "natural," or "common" across world languages into the contrastive framework (Park & Chen, 2025). The MDH asserts that structural differences between the L1 and TL are necessary but not sufficient to explain difficulty; one must also consider whether the target structure is more marked than the corresponding L1 structure (Wrembel & Zembrzuski, 2026).

Table 1. ***Comparative Summary of Theoretical Frameworks in L2 Phonological Acquisition***

Theoretical Model	Core Premise	Explanatory Mechanism	Limitations
Contrastive Analysis (CAH)	L1-L2 differences cause all errors	Transfer of L1 habits (Positive/Negative)	Fails to predict many errors; ignores universals
Markedness Differential (MDH)	Differences + Markedness = Difficulty	Typological universals and implicational hierarchies	Difficulty with defining "markedness" in all domains
Speech Learning Model (SLM)	Perception precedes production	Category formation and "equivalence classification"	Primarily focuses on immersion contexts
Perceptual Assimilation (PAM-L2)	L2 sounds map to L1 categories	Articulatory-acoustic similarity judgments	Focused primarily on perception over production

The implications of the MDH are profound for understanding English acquisition. For instance, English allows for a voice contrast in obstruents in word-final (coda) positions, a feature that is cross-linguistically more marked than a voice contrast in word-initial (onset) positions (Nguyen & Lee, 2023). Consequently, German speakers, whose L1 neutralizes voice contrasts in codas, find it significantly more difficult to acquire the English final /d/ in "bad" than English speakers find it to acquire the German final /t/ in "Rat". This asymmetry demonstrates that learners are not merely struggling with a "new" sound, but with a more marked phonological distribution (O'Brien & Fichtner, 2024).

Cognitive Frameworks: SLM and PAM-L2

Contemporary research has moved toward cognitive models that emphasize the role of speech

perception as the precursor to accurate production. The Speech Learning Model (SLM), developed by James Flege, posits that the L1 and L2 phonetic systems remain in a single, common phonological space (Arjmandi & Behroozmand, 2024). A central mechanism in the SLM is "equivalence classification," where a learner perceives an L2 sound as similar to an L1 category and thus fails to create a new phonetic category for it. Ironically, the SLM suggests that sounds that are "too similar" to the L1 may be harder to master in the long run than sounds that are "completely new," as the latter are more likely to escape the filter of the L1 and trigger the formation of a distinct category (Ellis & Shintani, 2020).

Complementing this is the Perceptual Assimilation Model for L2 (PAM-L2), which predicts the ease or difficulty of discriminating L2 contrasts based on how they map onto L1 categories. If two L2 phonemes are assimilated into two distinct L1 categories (Two-Category assimilation), the learner will likely have little trouble distinguishing them (Nam, 2025). However, if both L2 phonemes map onto a single L1 category (Single-Category assimilation), the learner may perceive them as identical, leading to persistent pronunciation and comprehension errors. These models suggest that the "accent" of an L2 learner is not merely a motor-skill failure but a perceptual one; fixing the "listening blueprint" is often the first step toward clearer speech (Wang, 2023).

Segmental Features and Phonemic Challenges

The segmental level of English phonology comprising its vowels and consonants is characterized by a high degree of complexity, particularly in its vowel inventory and its tolerance for dense consonant clusters (De Souza Sinnemäki, 2023).

The English Vowel Space: Tense, Lax, and the Schwa

One of the most frequent challenges for L2 learners of English is the size and distribution of its vowel inventory. Standard American English typically features 11 to 12 monophthong vowels, including a robust distinction between "tense" (long/high muscular tension) and "lax" (short/low muscular tension) vowels. Many languages, such as Spanish, utilize a much simpler five-vowel system (/a/, /e/, /i/, /o/, /u/) that lacks this tense-lax distinction (Mirzayev, 2024).

Research on Spanish-speaking learners of English has consistently highlighted the difficulty of the /i/–/ɪ/ contrast (as in "beat" vs. "bit"). Studies indicate that these learners often rely almost exclusively on durational cues how long the vowel lasts to distinguish these sounds, whereas native speakers prioritize spectral quality (the actual "color" of the sound). This reliance on duration is often insufficient, as English vowel length can vary based on the following consonant (e.g., vowels are longer before voiced consonants), leading to frequent confusion (Islam et al., 2024).

Table 2. ***Common Segmental Vowel Contrasts, L2 Deviations, and Semantic Impact***

English Vowel Pair	Phonetic Distinction	Common L2 Error	Meaning Impact
/i:/ vs /ɪ/	Tense vs Lax	Merging both to /i/	"Sheep" vs "Ship"
/ɛ/ vs /æ/	Mid-front vs Low-front	Merging both to /ɛ/ or /a/	"Bet" vs "Bat"
/u:/ vs /ʊ/	Tense vs Lax	Merging both to /u/	"Luke" vs "Look"
/ʌ/ vs /ɑ:/	Central vs Back	Substitution with /a/	"Duck" vs "Dock"

Beyond the monophthongs, the presence of the "schwa" sound (/ə/) is a major point of contention. As the most common vowel in English, the schwa is used in almost all unstressed syllables to

facilitate the language's stress-timed rhythm. Learners from syllable-timed languages often fail to reduce these vowels, instead pronouncing them with full quality as suggested by the spelling (e.g., pronouncing "about" with a full /a/ instead of /ə/). This failure to "de-voice" or "relax" unstressed vowels directly disrupts the prosodic flow of English, making the speaker sound unnatural to native ears (Munro, 2021).

Consonantal Inventory and Articulatory Difficulty

English consonants present difficulties both in their individual production and in their phonotactic distribution. Sounds like the interdental fricatives /θ/ (unvoiced, as in "think") and /ð/ (voiced, as in "this") are notoriously difficult because they require an articulatory position placing the tongue between the teeth that is absent in many of the world's languages. Learners frequently substitute these with /s/, /z/, /t/, or /d/, resulting in homophones that can obscure meaning (Luthfianda et al., 2024).

For instance, Mandarin and Arabic speakers frequently struggle with the /v/–/f/ or /v/–/w/ distinctions, as /v/ is not a phoneme in their native inventories. Mandarin speakers may substitute /v/ with /w/ (e.g., "very" becomes "wery"), while Arabic speakers may substitute /v/ with /f/ (e.g., "van" becomes "fan") or /p/ with /b/ (e.g., "park" becomes "bark"). These substitutions are not random; they are "nearest neighbor" assimilations, where the learner selects the closest available sound from their L1 phonetic map (Al-khresheh, 2024).

Syllable Structure and Consonant Clusters

Perhaps the most visible (and audible) area of L1 interference is the production of consonant clusters. English allows for highly complex syllable structures, represented as $(C)^3V(C)^4$, meaning up to three consonants can begin a syllable (onset) and up to four can end it (coda). Languages with more restricted syllable structures, such as Japanese (CV-based) or Arabic (no initial clusters), present a direct conflict for the learner (Garrido-Pozú, 2024).

Repair Strategies: Epenthesis and Deletion

When faced with an "illegal" cluster, the L2 learner typically employs one of two phonological repair strategies: vowel epenthesis or consonant deletion. Vowel epenthesis involves the insertion of a vowel to break up the cluster and create a more familiar CV structure (Broselow, 1987). This can take several forms: Prothetic Epenthesis: Adding a vowel before an initial cluster. A classic example is the Spanish-speaking learner who adds /e/ before /s/ + consonant clusters, pronouncing "special" as "especial" or "speak" as "espeak" (Tarone, 1987). Anaptyctic Epenthesis: Inserting a vowel between consonants. Arabic speakers, for example, often insert a short /i/ or /e/ into the cluster, pronouncing "split" as "siplit" or "street" as "sitreet". Paragogic Epenthesis: Adding a vowel at the end of a word to resolve a final consonant. This is common among Japanese and Mandarin speakers, who may pronounce "pig" as "pig-uh" or "desk" as "des-kuh" (Hancin-Bhatt & Bhatt, 1997).

Consonant deletion, on the other hand, involves simply dropping one or more segments of the cluster. High-proficiency learners are more likely to use deletion, as it preserves the overall rhythmic timing better than epenthesis, which adds entire syllables to the word. However, deletion can lead to serious intelligibility issues, particularly when it involves grammatical morphemes like the plural -s or the past tense -ed (Carlisle, 2001).

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895

Positional Influence on Cluster Difficulty

The position of the cluster within the word significantly influences the degree of difficulty. Evidence from Arabic-speaking EFL learners suggests that initial consonant clusters (ICCs) are more challenging than final consonant clusters (FCCs). ICCs require a rapid, precise articulatory transition at the very start of the utterance, whereas FCCs allow for a more progressive transition following the vowel nucleus. This finding contradicts some earlier assumptions that final clusters would be harder due to their high markedness. The cognitive load of initiating a complex motor sequence appears to outweigh the typological markedness of the coda in this specific population (Broselow, 1984).

Suprasegmental Phonology: The Soul of the Accent

While segmental errors are often the focus of classroom drills, suprasegmental features prosody, stress, rhythm, and intonation are widely considered by researchers to contribute more significantly to "foreign accent" and overall intelligibility. Nativelike use of suprasegmentals is the hallmark of fluent L2 speech (Eckman, 1991).

Figure: 1 A Multilayered Functional Perspective on Context and Stratification.

Word Stress and Intelligibility

English is a language of variable lexical stress, where the placement of stress can distinguish between word classes (e.g., 'object' vs 'object') or define lexical identity. For learners from languages with fixed stress (e.g., French or Polish) or tonal systems (e.g., Mandarin), English word stress is a major source of variation (Weinberger, 1994).

The "English Word Stress Error Gravity Hierarchy" suggests that not all stress errors are created equal. Word stress errors disrupt processing for listeners most severely when they involve concomitant vowel errors specifically, when shifting the stress changes a reduced vowel into a full one or vice versa. Because native English listeners rely on the contrast between stressed (full vowel) and unstressed (reduced vowel) syllables to identify words, a misplaced stress that maintains the correct vowel quality is less damaging than one that alters the vowel (Wolfram, 1985).

Rhythmic Timing: Stress-Timed vs. Syllable-Timed

English rhythm is classified as "stress-timed," where the duration of an utterance depends on the number of stressed syllables, with unstressed syllables being squeezed or reduced to fit the timing. In contrast, many L2 learners come from "syllable-timed" languages like Spanish, Arabic, or Cantonese, where each syllable is given roughly equal duration regardless of stress (Sato, 1984).

The transfer of syllable-timed rhythm to English results in a "flat" or "machine-gun" delivery that lacks the characteristic "bouncy" feel of native English. This rhythmic mismatch makes it difficult for native listeners to segment the speech, as they use the rhythmic "peaks" of stressed syllables as anchors for comprehension. Saudi learners, for example, often exhibit this flat intonation and no rhythm due to the absence of systematic suprasegmental training and the transfer of Arabic stress rules (Major, 2001).

Connected Speech: Linking and Intrusion (Eckman, 2004).

In natural, rapid English speech, word boundaries are often blurred through phonological processes that improve the flow of speech (Wayland et al., 2024). These include:

- **Consonant-to-Vowel Linking**: When a word ending in a consonant is followed by a word starting with a vowel (e.g., "laugh at" -> /lɑːfət/).
- **Linking /r/**: In Received Pronunciation (RP), the final /r/ is only articulated when followed by a vowel (e.g., "here are" -> /hɪər ə/).
- **Intrusion**: The addition of a sound segments (like /j/, /w/, or /r/) between two vowels (e.g., "go away" -> /gəʊ wəˈweɪ/).

For the L2 learner, mastering these processes is essential for perception. Learners who do not recognize linking frequently struggle to comprehend information because they cannot identify where one word ends and the next begins in the rapid stream of natural speech. Conversely, learners who fail to produce these links sound excessively formal or staccato, further contributing to their perceived lack of fluency (Fujino, 2021).

Individual Differences: Why Some Learners Succeed

The variation in L2 pronunciation attainment is vast, with some learners achieving near-native fluency while others remain heavily accented after decades of immersion. This discrepancy is explained by a constellation of internal and external factors (Losavio, 2023).

The Critical Period and Neuroplasticity

Age is arguably the most cited factor in L2 phonological success. Younger learners generally acquire a new language with better pronunciation and grammatical accuracy than adults. This is often attributed to the process of "lateralization," where the brain's language functions become fixed during puberty, leading to a loss of the "plasticity" required to establish new neural networks for foreign sounds (Ellis & Shintani, 2020). However, modern research suggests that plasticity survives well into the twenties and that adults can still achieve native-like pronunciation through focused practice and motivation (Darcy & Rocca, 2021).

Language Aptitude and Phonemic Coding

Language aptitude is an innate disposition that facilitates faster and more successful acquisition. A key component of this for pronunciation is "phonemic coding ability" the capability to discriminate and code foreign sounds such that they can be recalled later. Learners with high aptitude can figure out rules of pronunciation intuitively through exposure, a trait labeled as "inductive language learning ability". While intelligence (IQ) correlates strongly with academic L2 skills like grammar and vocabulary, it plays a significantly smaller role in the acquisition of pronunciation and communicative fluency (Thomson, 2024).

Motivation, Identity, and Agency

Affective factors such as motivation and social identity are critical mediators of phonological achievement. Gardner's socio-educational model distinguishes between integrative motivation (the desire to connect with the target culture) and instrumental motivation (practical benefits like career goals). Research shows that while motivation alone is insufficient to predict success, it drives the use of Pronunciation Learning Strategies (PLS), which are the actual predictors of improvement (Gordon & Darcy, 2025).

Table 3. Impact of Individual Learner Variables on L2 Pronunciation Attainment

Learner Factor	Impact on Pronunciation	Mechanism of Action
Age	High	Declining neuroplasticity and established L1 filters
Aptitude	Moderate-High	Superior phonemic coding and pattern recognition
Motivation	Moderate	Mediated through increased use of learning strategies
Identity	Variable	May cause resistance to "losing" L1-based identity
Personality	Moderate	Extroverts seek more social input and practice

Social identity presents a unique challenge: some learners deliberately maintain a "foreign accent" as a way of displaying their L1 identity and heritage. For example, the actor Jackie Chan has been observed to use more non-standard English variants over time, emphasizing his identity as a Chinese martial artist after achieving global success. Conversely, many learners profess a desire to sound native-like because they associate it with higher status, professional credibility, and social acceptance (Kardos & Sato, 2021).

Pedagogy and Technology: Optimizing Pronunciation Instruction

The instructional landscape for L2 pronunciation has shifted significantly with the advent of computer-assisted pronunciation training (CAPT) and artificial intelligence (AI) (Kissling, 2022).

Explicit vs. Implicit Instruction

A long-standing debate in L2 pedagogy concerns the effectiveness of explicit versus implicit instruction. Explicit instruction involves direct teaching of rules (e.g., articulatory placement for /θ/), while implicit instruction relies on naturalistic exposure and practice. Meta-analyses consistently indicate that explicit instruction is more effective, particularly for developing cognitive awareness and accuracy in complex structures. However, for long-term retention and fluency, explicit teaching must be accompanied by communicative tasks and follow-up instruction (Nguyen & Lee, 2023).

The Role of CAPT and AI in Modern Classrooms

Technology-mediated instruction, particularly through Automated Speech Recognition (ASR), offers a scalable solution to the challenges of pronunciation training. ASR-based tools provide individualized, immediate, and repetitive practice, which is often impossible for a single teacher to provide in a large classroom (Kardos & Sato, 2021).

Table 4. Comparison of Technology-Mediated (CAPT/AI) and Native Speaker-Led Pronunciation Pedagogy

Feature	CAPT/AI-Based Instruction	Native Speaker-Led Instruction
Feedback	Immediate, consistent, explicit	Nuanced, real-time, but often unsystematic
Anxiety	Low; autonomous and private	Higher; students may feel embarrassed
Scalability	High; accessible 24/7	Low; dependent on instructor availability
Context	Often lacks social/pragmatic depth	Rich in social and contextual modeling

A critical finding in the research is the importance of explicit corrective feedback in digital tools. ASR programs that highlight specific errors are significantly more effective (effect size $g = 0.86$) than simple dictation tools that only provide transcribed words ($g = 0.50$). Furthermore, technology has been found to be "highly efficient" in improving segmental accuracy but less effective for suprasegmental features, which remain a challenging area for current speech recognition algorithms to evaluate with the same nuance as a human ear (Park & Chen, 2025).

Mandarin-English Contrast

Mandarin learners face obstacles rooted in both phonology and syntax. The absence of the /v/, /θ/, and /ð/ sounds leads to substitutions with /w/, /s/, and /z/. Because Mandarin words are generally monosyllabic and lack the varied stress patterns of English, learners often struggle with lexical stress placement. Additionally, the lack of inflectional morphology in Mandarin makes it difficult for learners to "notice" or produce the English plural -s or past tense -ed markers, which are critical for both grammatical and phonological accuracy (Franklin & McDaniel, 2016).

Arabic-English Contrast

Arabic-speaking learners are significantly influenced by their L1 syllable structure, which forbids initial clusters. This leads to the heavy use of vowel epenthesis (e.g., "sip-lit" for "split").

Furthermore, the lack of /p/ and /v/ in Arabic leads to substitutions with /b/ and /f/. Studies in Saudi Arabia have highlighted that even short, structured pronunciation programs focusing on these specific contrasts can significantly improve intelligibility and student confidence (Levis, 2023).

L1 interference and perceptual assimilation.
However, the field is also moving toward more optimistic and pragmatic goals. The shift from "accent reduction" to "intelligibility" recognizes that a native-like accent is not the only measure of success (Levis, 2010). By focusing instruction on high-functional-load features such as word stress, rhythm, and the contrasts most likely to cause misunderstanding (like tense vs. lax vowels) teachers can empower learners to achieve successful communication. Technological advancements in AI and ASR are providing new avenues for this development, offering a standardized, low-anxiety environment for the intensive practice that phonological mastery requires (Avrianti, 2026). As these tools become more sophisticated in their handling of suprasegmental features, they will likely become an indispensable part of the L2 curriculum. Ultimately, successful L2 phonological acquisition is a multifaceted process that requires the learner to not only change how they speak but to fundamentally change how they hear. By addressing both the physical and cognitive aspects of speech, and by leveraging both explicit instruction and technological support, learners can overcome the profound challenges of English pronunciation (Kkese & Dimitriou, 2023).

Conclusion

The analysis of phonological features and pronunciation challenges in second language English acquisition demonstrates that pronunciation learning is a highly complex and multidimensional process. Learners face persistent difficulties due to the influence of their first language, perceptual limitations, articulatory habits, and differences in phonological systems between the L1 and English. Segmental problems such as vowel contrasts, consonant substitutions, and consonant cluster production combine with suprasegmental challenges involving stress, rhythm, and intonation to affect overall intelligibility and communicative competence. Theoretical models including CAH, MDH, SLM, and PAM-L2 provide valuable insight into how learners perceive and produce unfamiliar sounds and why some pronunciation errors persist despite extensive exposure. The review also highlights the significant role of learner-related variables such as age, aptitude, motivation, identity, and exposure in determining pronunciation success. Furthermore, explicit pronunciation instruction and technology-based learning tools, particularly AI-driven CAPT systems, have emerged as effective resources for improving segmental accuracy and learner confidence. However, suprasegmental mastery still requires meaningful interaction and communicative practice. The study concludes that modern pronunciation teaching should prioritize intelligibility, comprehensibility, and communicative effectiveness rather than complete native-like accuracy. A balanced integration of cognitive theory, pedagogical practice, and technological innovation can significantly enhance the development of English pronunciation skills among L2 learners.

References

Crowther, D., & Trofimovich, P. (2021). Second language speech comprehensibility and the role of listener background. *Studies in Second Language Acquisition*, 43(4), 789–812.

Darcy, I., & Rocca, B. (2021). Perceptual filters in L2 phonological acquisition: The role of L1 articulatory habits. *Journal of Phonetics*, 88, 101092.

Gordon, J., & Darcy, I. (2025). Prosodic hierarchies and connected speech segmentation in L2 listening. *Applied Psycholinguistics*, 46(1), 55–82.

Kissling, E. M. (2022). The challenge of L2 phonology in applied linguistics. *Language Teaching*, 55(2), 176–194.

Saito, K. (2023). Segmental and suprasegmental representations in L2 phonological development. *Bilingualism: Language and Cognition*, 26(3), 512–528.

Thomson, R. I. (2024). Rapid speech perception and L2 listening comprehension. *Annual Review of Applied Linguistics*, 44, 33–49.

Trofimovich, P., & Turuševa, L. (2020). Cognitive restructuring of sound categories in second language learners. *Second Language Research*, 36(4), 487–506.

Ellis, R., & Shintani, N. (2020). *Exploring language pedagogy through second language acquisition research*. Routledge.

Kardos, M., & Sato, T. (2021). Revisiting contrastive analysis in L2 phonology: Empirical evidence from Asian learners. *Second Language Research, 37*(4), 521–544.

Martinez-Flor, A., & Uso-Juan, E. (2022). Current perspectives on transfer and markedness in phonology. *Journal of Second Language Pronunciation, 8*(2), 189–213.

Nguyen, H. T., & Lee, S. J. (2023). Developmental vs. interference errors: A longitudinal study of L2 speech. *Studies in Second Language Acquisition, 45*(1), 102–128.

O'Brien, M. G., & Fichtner, F. (2024). Typological markedness and the acquisition of L2 phonological contrasts. *Applied Psycholinguistics, 45*(3), 401–427.

Park, J. E., & Chen, L. (2025). From CAH to MDH: A historical reanalysis of phonological difficulty predictions. *Language Learning, 75*(1), 55–89.

Wrembel, M., & Zembrzuski, D. (2026). Markedness beyond segments: Prosodic constraints in L2 phonology. *Bilingualism: Language and Cognition, 29*(2), 233–251.

Arjmandi, M. K., & Behroozmand, R. (2024). On the interplay between speech perception and production: insights from research and theories. *frontiers in Neuroscience*, *18*, 1347614.

Nam, Y. (2025). Discrimination of English Vowel Contrasts in Chinese Listeners in Relation to L2-to-L1 Assimilation. *Behavioral Sciences*, *15*(10), 1420.

Wang, Y. (2023). *Perceptual modification of nonnative phonemic sequences* (Doctoral dissertation, The University of Melbourne).

De Souza, R. N., & Sinnemäki, K. (2023). Beyond segment inventories: Phonological complexity measures and suprasegmental variables in contact situations. *Journal of Language Contact*, *15*(3-4), 439-480.

Mirzayev, E. (2024). A comprehensive guide to English's most common vowel sound. *Global Spectrum of Research and Humanities*, *1*(1), 19-26.

Islam, M. J., Anwar, M. S., & Kamal, S. M. (2024). Cross-linguistic Phonological Transfer: The Influence of L1 Vowel Contrasts on L2 Tense-lax Vowel Acquisition. *Crossings: A Journal of English Studies*, *15*(1), 215-235.

Munro, M. J. (2021). On the difficulty of defining “difficult” in second-language vowel acquisition. *Frontiers in Communication*, *6*, 639398.

Luthfianda, S., Irawan, Y., Rahayu, R., & Hidayat, S. (2024). EXPLORING PRONUNCIATION CHALLENGES: INDONESIAN UNIVERSITY STUDENTS'PRODUCTION OF ENGLISH FRICATIVE SOUNDS. *English Review: Journal of English Education*, *12*(1), 85-94.

Al-khresheh, M. H. (2024). Phonetic challenges in English: the impact of mispronunciation of the bilabial plosive/p/on communication among Saudi EFL learners. *Cogent Arts & Humanities*, *11*(1), 2390777.

Garrido-Pozú, J. J. (2024). The Interplay of Syllable Structure and Consonant Sonority in L2 Speech Segmentation. *Languages*, *9*(3), 103.

Broselow, E. (1984). An investigation of transfer in second language phonology. *IRAL: International Review of Applied Linguistics in Language Teaching*, 22(4), 253–269.

Broselow, E. (1987). Non-obvious transfer: On predicting epenthesis errors. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 292–304). Newbury House.

Carlisle, R. S. (2001). The modification of onsets in a markedness relationship: Testing the Interlanguage Structural Conformity Hypothesis. *Second Language Research*, 17(4), 343–361.

Eckman, F. R. (1991). The Structural Conformity Hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition*, 13(1), 23–41.

Eckman, F. R. (2004). From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition*, 26(4), 513–549.

Hancin-Bhatt, B., & Bhatt, R. (1997). Optimal L2 syllables: Interactions of transfer and developmental effects. *Studies in Second Language Acquisition*, 19(3), 331–378.

Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Lawrence Erlbaum.

Sato, C. J. (1984). Phonological processes in second language acquisition: Another look at interlanguage syllable structure. *Language Learning*, 34(4), 43–57.

Tarone, E. (1987). The phonology of interlanguage. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 70–85). Newbury House.

Weinberger, S. H. (1994). Functional and phonetic constraints on second language phonology. In R. Yavas (Ed.), *First and second language phonology* (pp. 283–302). Singular Publishing.

Wolfram, W. (1985). Variability in tense marking: A case for the obvious. *Language Learning*, 35(2), 229–253.

Fujino, H. (2021). L2 learners' perceptions of grammar: The case of JFL learners in the UK. *The Language Learning Journal*, *49*(3), 343-357.

Wayland, R., Meyer, R., Vellozzi, S., & Tang, K. (2024). Lenition in L2 Spanish: The impact of study abroad on phonological acquisition. *Brain Sciences*, *14*(9), 946.

Losavio, A. (2023). Late Starts Leading to Native-Like Pronunciation In Second Language Acquisition.

Avrianti, N. (2026). AI-based pronunciation training tools for English language learners: A systematic review. *International Journal of Computer-Assisted Language Learning*, *16*(1), 1–15.

Franklin, A., & McDaniel, L. (2016). Exploring a phonological process approach to adult pronunciation training. *American Journal of Speech-Language Pathology*, *25*(2), 172–182. https://doi.org/10.1044/2015_ajslp-14-0172

Kkese, E., & Dimitriou, D. (2023). Audio-visual speech perception of plosive consonants by CG learners of English. *Journal of Monolingual and Bilingual Speech*, *5*(1), 1–28. https://doi.org/10.1558/jmbs.23017

Levis, J. (2010). Pronunciation and intelligibility: Issues in research and practice. *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*, 1–12.

Levis, J. M. (2023). Pronunciation in language teaching. *Iowa State University Digital Press*, 45–