Analysis of Pronunciation Errors and Correction Strategies in Second Language Acquisition of English

Tingzhen Zhao

Education University of Hong Kong, Hong Kong, China
zhaotzh417@163.com

Abstract. As demand for English as a Second Language grows globally, learners face native language interference, limited phonetic perception, and scarce resources, hence posing key barriers to intelligibility and cross-cultural communication. This study explores the types, causes, diagnosis, and intervention of English pronunciation errors across different native language backgrounds backgrounds, and examines intelligent technologies for personalized correction. By reviewing and analyzing relevant literature, this study summarizes research progress and current teaching practices in second language pronunciation errors. The results demonstrates that traditional teacher-dependent correction methods are limited by delayed feedback and insufficient support for individual differences, whereas emerging technologies such as acoustic analysis, virtual reality, biofeedback, and AI offer feasible approaches for real-time visual diagnosis and targeted correction, thus enhancing learners' pronunciation accuracy and speech fluency. This paper further points out that future ESL pronunciation teaching needs to deepen the integration of linguistic theory and technological tools, and build a multimodal and sustainable pronunciation correction ecosystem, to better meet the cross-cultural communication needs of different learners.

Keywords: Cross-cultural pronunciation teaching, Pronunciation errors and correction, Negative transfer from native language, Multidimensional pronunciation correction system

1. Introduction

With globalization, English has become an important lingua franca in international communication, education, and business, with approximately 1.5 billion people learning it as a second language (ESL) worldwide [1]. The comprehensibility of communication and learners' self-confidence are directly affected by pronunciation, which is an essential component of language learning. However, pronunciation has long been a major challenge for ESL learners, as non-native accents may lead to communication barriers while also influencing learners' sense of social identity. In recent years, research on pronunciation errors in ESL learners has made significant progress. The focus of such studies has shifted from early contrastive analyses emphasizing the influence of learners' native language on the target language to perspectives informed by sociolinguistics and cognitive science [2]. Findings indicate that pronunciation errors typically manifest in multiple dimensions, including segmental substitutions, prosodic deviations, and restrictions in phonotactic patterns. The adoption of corpus linguistics, acoustic analysis, and psycholinguistics has made the study of pronunciation

errors more systematic, thus revealing the roles of speech perception, working memory, and motor control in speech production. Nevertheless, existing studies largely remain descriptive, with limited connection to teaching practice. Traditional training methods often overlook learners' individual differences and specific needs. Although visual feedback and digital tools have attracted attention in pronunciation instruction, their effective and personalized use needs further study. This review aims to examine the research progress on ESL pronunciation errors, exploring common issues among learners from different L1 backgrounds, influencing factors, and existing training methods and technologies. It also discusses their limitations and proposes directions for future research, with the goal of providing practical guidance for ESL pronunciation instruction.

2. Pronunciation errors in second language acquisition

2.1. Definition and classification of pronunciation errors

In second language acquisition, pronunciation errors refer to learners' speech productions that consistently deviate from native-like norms and exhibit systematic and developmental patterns [3]. These errors mainly arise from multiple factors, including L1 transfer, overgeneralization of target language rules, difficulties in speech perception, and insufficient training [4]. L1 transfer is the most common factor; for example, Mandarin Chinese lacks the interdental /θ/ (as in think), so learners often substitute it with /s/ (as in sink). Rule overgeneralization occurs when learners apply a learned pronunciation pattern in inappropriate contexts—for instance, some learners may lengthen /ɪ/ (as in bit) to /iː/ (as in beat) where it is not required, often accompanied by changes in mouth opening. Speech perception difficulties stem from limitations in the L1 phonemic system, making it hard for learners to distinguish minimal pairs such as /iː/ (sheep) and /ɪ/ (ship), which affects both perception and production. Without systematic training and timely correction, certain sounds, such as the flap /ɾ/ (in butter), which requires a quick tongue tap against the alveolar ridge, are often substituted with /d/ (/ˈbʌdər/). When learners lack structured practice and effective feedback outside the classroom, such errors are more likely to persist over time.

2.2. Relationship between educational background and pronunciation errors

Educational background plays a significant role in second language pronunciation acquisition, as factors such as teaching methods, teachers' phonetic competence, learners' cognitive development, and the availability of instructional resources directly influence pronunciation performance [5]. Systematic pronunciation instruction can enhance both phoneme perception and production skills; however, overly uniform teaching that neglects authentic communicative contexts often fails to balance accuracy with naturalness and fluency [6,7]. Equally important is the phonetic proficiency of teachers, as a lack of systematic training may lead them to provide inaccurate models, thereby unintentionally reinforcing errors [8]. This issue is closely linked to the insufficient emphasis on pronunciation in teacher training programs in certain regions, which can constrain the time and strategies devoted to pronunciation correction in the classroom [9]. Moreover, Learners' age and cognitive development impact pronunciation plasticity. For example, children demonstrate greater phonetic malleability, and early exposure strongly affects later pronunciation accuracy [10]. In contrast, adult learners are more prone to L1 influence and perceptual constraints, often producing errors in prosody or connected speech. The availability of teaching resources directly influences learners' access to timely and effective feedback and practice, reducing errors [11]. Conversely, resource shortages often lead to insufficient practice, allowing pronunciation errors to persist.

2.3. Effects of pronunciation errors on communication and comprehension

The intelligibility of speech and the effectiveness of communication are affected by pronunciation errors in second language learners. These errors have been mainly examined in terms of segmental and suprasegmental aspects, social communication, as well as technological interventions. In terms of segments, semantic confusion usually arises from phoneme substitution, vowel reduction, and stress misplacement. For example, Mandarin speakers who substitute the English interdental /θ/ with /s/ may confuse think with sink, thereby compromising accurate information transfer. In terms of stress patterns, Chinese learners often favor fixed initial-syllable stress, whereas the dynamic stress shifts in English can lead to lexical ambiguity, as seen in the difference between the noun ˈrecord and the verb reˈcord. These errors result in semantic confusion, disrupt speech fluency, and increase the need for repairs, thereby affecting conversational efficiency. Suprasegmental errors are often more subtle; inappropriate use of rhythm and intonation can reduce the naturalness of speech flow and lead listeners to misinterpret sentence meaning. From a social interaction perspective, long-term pronunciation errors not only weaken children's confidence in expressing themselves in daily language interactions but also affect adult learners' professional image in academic settings or workplace communication. Moreover, non-standard pronunciation can also reinforce stereotypes in cross-cultural communication, posing a potential barrier to interpersonal interaction and cultural integration. In recent years, new approaches to pronunciation correction have been made possible by recent technological developments. For instance, targeted feedback to improve segmental and prosodic accuracy can be delivered by intelligent training systems employing automatic speech recognition (ASR) and speech visualization, though further refinement is required to account for individual differences and feedback precision.

3. Evaluation and identification methods of pronunciation errors

3.1. Types and characteristics of pronunciation errors

In second language learning, pronunciation errors are commonly observed and can be categorized according to linguistic competence, underlying mechanisms, and surface manifestations. Based on Chomsky's distinction between competence and performance, Corder was the first to differentiate language deviations into cognitive errors and performance mistakes. In particular, Errors represent systematic deviations from target language norms, thus indicating learners' stage of development in language competence. They exhibit regularity, are not amenable to self-correction, and constitute a defining feature of interlanguage. In contrast, mistakes are incidental errors, often caused by stress or inattention, and can usually be self-corrected. In second language learning, developmental errors are categorized into intralingual errors, which arise from the complexity of target language rules, and interlingual errors, which result from negative transfer from the learner's native language. From a methodological perspective, Dulay et al. proposed a fourfold framework for classifying errors based on language subsystems, learning strategies, communicative effects, and language contrasts. In addition, using a hierarchical approach, James classified errors into three types: intrinsic errors, which encompass spelling and pronunciation; textual errors, covering vocabulary and syntax; and discourse errors, relating to pragmatic coherence deviations [12]. Moreover, based on the stages of language acquisition, errors are pre-systematic, systematic, and post-systematic, corresponding to stages where learners' rules are not yet established, not fully developed, or already acquired but not yet applied automatically.

3.2. Diagnostic techniques and methods for pronunciation errors

With the advancement of speech recognition technology, intelligent speech diagnosis based on constructivist principles has emerged as a viable approach for detecting pronunciation errors. This model emphasizes learner-centered approach and dynamic feedback through technology-supported immersive interaction and self-assessment, encouraging learners to actively build knowledge and reflect. In particular, speech recognition-based platforms can create authentic language interaction scenarios, thus enabling frequent self-directed practice with immediate feedback and reinforcing a learning model that blends group collaboration with individual assessment. The real-time generation of pronunciation diagnostic reports helps learners promptly identify pronunciation errors and make targeted corrections, fostering a spiral of continuous improvement. Besides, the adaptive training module of the SRT system can dynamically adjust training content and difficulty based on learners' language proficiency, effectively addressing the issue of insufficient personalized support in oral training during high school. This differentiated training approach reflects constructivist principles of gradual learning and attention to individual differences, enhancing the precision and effectiveness of oral practice. Related research indicates that pronunciation diagnosis and feedback mechanisms based on speech recognition significantly promote the development of students' language abilities and the reconstruction of their knowledge systems. The system-generated traceable learning reports support learners in conducting metacognitive reflection, while teacher-side data analysis further enables tailored instructional interventions, facilitating the effective operation of the "diagnosis - feedback - refinement" cycle.

3.3. Intervention and feedback mechanisms for pronunciation errors

By combining structured practice with targeted feedback, pronunciation error intervention helps learners to gradually master difficult sounds and improve clarity and naturalness in speech. Once a precise diagnosis is made, the subsequent interventions should adhere to the principles of language acquisition and cognitive traits, and create exercises that are personalized, visually engaging, and incrementally organized. At the basic stage, learners can observe complex sounds through moving mouth models or acoustic waveforms, clearly showing tongue placement and airflow control. For example, native Korean speakers frequently confuse English short and long vowels, which can be corrected using minimal pair practice like ship versus sheep, along with tongue mobility exercises to enhance articulatory coordination. Besides, native Arabic speakers tend to pronounce English /p/ as /b/, which can be corrected by combining aspiration detection with mouth-shape demonstrations to reinforce perception and production of aspirated consonants. For fluency training, shadowing combined with phoneme dictation tasks can be used to strengthen weak syllables, linking, and stress patterns within sentences, enhancing the naturalness of speech flow. Intervention should follow a step-by-step progression based on proficiency levels, with beginners focusing on individual sound accuracy, intermediate learners practicing flow and rhythm, and and advanced learners developing varied intonation and expression through activities such as speech imitation or film dubbing. The feedback system emphasizes a combination of explicit and implicit approaches, where critical errors that affect comprehension, such as pronouncing /p/ as /b/, can be corrected through direct teacher demonstration or guided self-correction using repeated correct forms. Indirect feedback can use situational Q&A and metalinguistic cues to boost learners' phonetic awareness and self-monitoring. Combined with AI assessment, virtual reality simulations, and multimodal input, pronunciation training is becoming more personalized, visual, and efficient, offering learners flexible correction paths and continuous progress tracking.

4. Pronunciation correction strategies and teaching methods

4.1. Comparison of traditional and modern pronunciation correction methods

The integration of linguistic theory and educational technology has driven advances in English pronunciation correction. Traditional approaches, based on behaviorist and structuralist principles, rely on imitation and repetition to establish correct pronunciation. For example, the "shadowing" technique requires learners to imitate speech either simultaneously or with a slight delay after listening, hence reinforcing both phoneme perception and output consistency. Moreover, systematic instruction in phonemes combined with demonstrations of tongue placement helps learners master commonly mispronounced sounds, such as the interdental /θ/ and the open vowel /æ/. In addition, teachers often provide feedback through a "demonstration-correction-reinforcement" cycle. Such face-to-face guidance emphasizes individual attention, supporting beginners in developing phonetic awareness and confidence. However, traditional methods heavily rely on teachers' demonstrations and experience, often providing delayed and subjective feedback, and offer relatively limited coverage in training connected speech features such as linking, weak forms, and stress patterns.

Notably, modern pronunciation correction methods utilize technologies like speech recognition, acoustic analysis, and virtual reality, providing improved diagnosis and instant feedback. Intelligent scoring systems powered by AI analyze learners' pronunciation across multiple acoustic dimensions, compare it with standard speech databases, and provide visual reports highlighting deviations and recommended improvements. By visualizing tongue and lip movements through VR or 3D motion capture, advanced systems enable learners to understand and refine pronunciation. Though acoustic analyses of parameters such as formants and VOT (voice onset time) can now be automated, these technologies require advanced equipment, stable networks, and operator proficiency. Compared with traditional approaches, modern methods excel in personalization and traceability, yet they fall short of teachers' social and emotional support, emphasizing the value of integrating both methods.

4.2. Personalized and differentiated pronunciation correction strategies

With the integration of intelligent technology, English pronunciation correction is gradually shifting from traditional standardized teaching to a more personalized and differentiated approach. Through layered diagnosis and dynamic adjustment, learners' specific error patterns can be targeted for more efficient pronunciation intervention. Personalized correction emphasizes identifying and correcting individual pronunciation features. Employing acoustic analysis and visual demonstrations, learners learners can clearly see the gap between their own pronunciation and the standard, while reinforcing perception and production through minimal pair training. For example, with often-mixed sounds like /θ/ and /s/ in think versus sink, or /l/ and /r/ in light versus right, dynamic tongue-position demonstrations and staged pronunciation guidance can enhance the precision and consistency of sound production. Furthermore, differentiated correction targets learner groups with similar native language backgrounds, designing specific interventions for their systematic errors. For example, common errors among Chinese learners, such as missing interdental sounds and confusion between short and long vowels, can be addressed through staged practice and visual comparison, gradually strengthening tongue placement control and vowel duration perception. Such group-focused designs help improve training efficiency and prevent the reinforcement of repetitive errors. To ensure the effectiveness of personalized and differentiated strategies, dynamic assessment is especially crucial. By regularly comparing recordings, reviewing visual reports, and receiving teacher feedback, new errors can be quickly spotted, and training priorities adjusted flexibly, creating an ongoing iterative

correction process. Technologies like intelligent speech recognition, virtual reality (VR) interaction, and multimodal input offer strong support for this process, yet balancing reliance on technology with human guidance remains essential to prevent overly mechanical training.

4.3. Technology-driven pronunciation correction methods and applications

The integration of acoustic analysis, virtual reality, biofeedback, and artificial intelligence provides high-precision, visualized, and personalized support for diagnosing and correcting pronunciation errors. Using acoustic analysis tools such as Praat, learners' speech can be measured in multiple dimensions. This includes precise tracking of vowel formants, where the F1 value of /æ/ generally falls between 700 and 800 Hz, and consonant VOT. This helps identify differences in aspiration intensity and articulation points, thus providing a basis for monitoring training outcomes. Besides, VR combined with 3D tongue-position tracking, such as Speech Mirror, captures the real-time 3D coordinates of articulators like the tongue and lips, compares them against a native-speaker model, and calculates detailed parameters for corrections with sub-millimeter precision. In an immersive environment, learners can repeatedly imitate and fine-tune their articulation, enhancing tongue precision for sounds like the interdental /θ/. Biofeedback technology combined with EMG sensors monitors the activation of relevant throat and oral muscles in real time, assisting in correcting errors like voicing, insufficient aspiration, and other deviations. EMG signals can be integrated into the VR view, giving learners a clear visual of muscle coordination during speech. And the affective computing module utilizes cameras and speech emotion analysis to monitor learners' focus and emotional fluctuations during training, automatically adjusting the pace and feedback methods to minimize errors caused by tension or fatigue, thereby enhancing the adaptability and continuity of practice. AI algorithms like Bayesian optimization and neural networks can dynamically allocate training emphasis based on learners' historical pronunciation data, targeting and strengthening weak areas. For instance, to address the common omission of /θ/ among native Chinese speakers, a three-step task of tongue-tip contact with teeth friction with aspiration and vowel linkage can be implemented with EMG monitoring used to enhance articulation control. For confusion between the long and short vowels /iː/ and /ɪ/, acoustic spectrometers and interactive acoustic games can display energy distribution differences in real time, improving sensitivity to timbre and duration and helping learners achieve targeted improvements.

5. Practical applications and challenges in pronunciation teaching

To ensure effective and sustainable teaching, English pronunciation instruction across educational stages and cross-cultural contexts must balance technological tools with cultural adaptation while addressing learners' actual needs. In primary education, teachers should prioritize fostering interest and building basic phonemic skills. Typical strategies include incorporating difficult phonemes into playful or contextual tasks, supported by immediate feedback and visual demonstrations, such as exaggerated mouth movements and gestures, to foster intuitive understanding. In addition, younger learners are highly influenced by peers, making peer evaluation and recorded scoring effective tools to promote self-correction of pronunciation in a relaxed setting. In higher education and workplace contexts, pronunciation training emphasizes natural speech flow and context-specific professional communication. In higher education, shadowing paired with intonation curve analysis helps learners align their speech with native patterns, progressively acquiring natural flow features like linking and weakening. In workplace settings, English training prioritizes precise articulation of field-specific terms, supported by 3D visualizations to enhance learners' perception and accuracy. Furthermore, in

cross-cultural settings, teachers encounter two main challenges in managing learners' linguistic diversity and addressing cultural sensitivity. To mitigate systematic biases and L1 transfer, teachers should integrate acoustic analysis and biofeedback techniques into daily instruction to continuously monitor learners' pronunciation. Moreover, learners from different cultural backgrounds vary in their receptiveness to error correction and public feedback. Pronunciation correction plans should therefore emphasize cultural responsiveness, employing inclusive and supportive strategies that reduce anxiety, enhance motivation, and foster cross-cultural communication skills.

6. Conclusion

This study shows that learners' L1 backgrounds create specific obstacles in the process of acquiring accurate English pronunciation, primarily due to negative transfer from the native language (e.g., Chinese speakers pronouncing /θ/ as /s/), overgeneralization of target language phonological rules, and limitations in phonetic perception compounded by insufficient training. Traditional methods based on repetitive practice are insufficient to address the personalized needs of modern learners. Though emerging technologies such as biofeedback and speech analysis hold great promise, their effective integration into classroom instruction remains underexplored. Given the unique challenges of cross-cultural speech instruction, an urgent need exists for a multi-dimensional pronunciation correction system that integrates linguistic theory, technological tools, as well as cultural insight. By examining phonological differences, such as Chinese tones and English intonation, learners can reduce the influence of their native language. The use of VR and related technologies to simulate multi-accent environments can enhance learners' ability to adapt to diverse cultural and linguistic contexts. To enhance pronunciation instruction, teachers should integrate linguistic, cognitive, and cultural considerations into error analysis, and technological tools should be developed to support practical classroom use like lightweight solutions for fragmented practice. Moreover, cross-cultural teaching should use dynamic assessments to track learners' pronunciation in different contexts. In the future, further research can explore technology-driven adaptive learning models, integrating emerging techniques such as voice cloning to enable precise pronunciation modeling and advance ESL instruction toward intelligent, personalized development.

References

[1] Bao, L.N. (2025) The Role of English as a Global Language in Shaping Students' Worldviews. GBP Proceedings Series, Scientific Open Access Publishing, 6, 9-18.

[2] Rehman, I., Silpachai, A., Levis, J., et al. (2020). The English pronunciation of Arabic speakers: A data-driven approach to segmental error identification. Language Teaching Research, 26(6), 1362-1388.

[3] Younghwan, B., & Lan, C. J. (2019). A study on the pronunciation errors of Chinese learners through narratives. The Journal of Language & Literature, 8, 387-415.

[4] Shi, L. (2020). An Analysis of Pronunciation Errors and Teaching Approaches for International Students at the Elementary Level in Mixed Classes. Jiangsu University.

[5] Alharbi, M.J. (2024). Acquired versus learned systems in second language acquisition: A review of studies based on Krashen's hypothesis. Theory and Practice in Language Studies, 14(1), 177-185.

[6] Pereira, B.V.L., de Carvalho, M.B.F., Alves, P., et al. (2024). Automatic phoneme recognition by deep neural networks. The Journal of Supercomputing, 80(11), 16654-16678.

[7] Xue, L. (2022). Application of discriminative training algorithm based on intelligent computing in English translation evaluation. Applied Mathematics and Nonlinear Sciences, 8(2), 193-202.

[8] Jiang, Y.L., & Peng, J.E. (2024). A comparative study of discussion forum comments in foreign language MOOCs from a big data perspective: Based on sentiment and content analysis. Journal of Zhejiang International Studies University, (03), 26-34.

[9]   Li, Y. (2020). Problems and teaching strategies in Chinese pronunciation instruction at Gongju Girls' High School in Korea. Zhengzhou University.

[10] Cychosz, M., Scarpelli, C., Stephans, J., Sola, A.M., Kolhede, K., Ramirez, R., Christianson, E., Chan, V., & Chan, D.K. (2025). Rapid increases in children's spontaneous and responsive speech vocalizations following cochlear implantation: Implications for spoken language development. Ear & Hearing, 46(4), 1029-1043.

[11] Mahdi, H. S., Alkhammash, R., & Al-Athwary, A.A.H. (2023). Using high variability phonetic training as a contextualized tool in the development of English consonant clusters pronunciation among Saudi EFL learners. Education and Information Technologies, 29(6), 6821-6840.

[12] Chen, X. (2024) Corpus-based Study on Language Errors in English Writing. International Journal of Education and Humanities 15(2), 144-153.