Proceedings of ICILLP 2025 Symposium: Psychological Perspectives on Teacher-Student Relationships in Educational Contexts
DOI: 10.54254/2753-7048/2025.NS27536

A Review of Pronunciation Challenges Faced by ESL Learners from Different Varieties of Chinese

Yuning Du

School of Foreign Languages, Sichuan Normal University, Chengdu, China
yuning.du.kelly@gmail.com

Abstract. Accurate pronunciation plays a vital role in second language (L2) acquisition, yet many Chinese English learners continue to face persistent challenges due to negative transfer from their first language (L1). While prior studies have largely focused on Standard Mandarin, the impact of other Chinese varieties such as Cantonese and Sichuan dialect has received less systematic attention. This review aims to synthesize existing empirical findings on the pronunciation difficulties experienced by ESL learners from these three representative Chinese language varieties. Drawing on over thirty peer-reviewed studies, this review categorizes learners' challenges into segmental and suprasegmental domains, identifies common and dialect-specific error patterns, and analyzes the phonological features driving these patterns. The paper further highlights research gaps, particularly the lack of cross-dialect comparison, the limited exploration of suprasegmental transfer, and the narrow participant demographics. By adopting a comparative approach, this review intends to enhance theoretical insights into L2 pronunciation acquisition and provide evidence-based guidance for dialect-sensitive pronunciation instruction, ultimately supporting more effective communication for diverse Chinese ESL learners.

Keywords: Pronunciation, ESL, Chinese Varieties, Phonological Transfer

1. Introduction

Pronunciation is a critical component of second language (L2) acquisition. In the field of English as a Second Language (ESL), research has shown that poor pronunciation can impede intelligibility and fluency more severely than grammatical or lexical errors [1]. While recent years have seen a revival of interest in L2 pronunciation teaching and learning, many studies continue to focus on segmental features or learners from relatively homogeneous linguistic backgrounds [2].

Among Chinese learners of English, a wide range of pronunciation difficulties has been documented, particularly due to negative transfer from first language (L1) phonological systems [3]. This review focuses on synthesizing research findings related to three representative Chinese varieties: Standard Mandarin, Cantonese, and the Sichuan dialect. The study aims to identify major pronunciation challenges faced by learners from these dialect groups, examine how phonological features of each variety contribute to L2 English pronunciation difficulties, and highlight areas where further empirical research is needed. Drawing from over 30 peer-reviewed articles, the review categorizes issues into segmental and suprasegmental domains and compares patterns across

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2025.NS27536

dialects. Ultimately, this paper endeavors to furnish a more lucid understanding of how L1 phonology influences ESL pronunciation performance—with the aim of providing valuable references for the formulation of more dialect-responsive teaching strategies and yielding insights for future inquiries in the field of cross-linguistic phonology.

2. Theoretical background

2.1. Second language pronunciation acquisition

Compared to grammar, learners often have low metalinguistic awareness of phonology, making it difficult for them to detect or self-correct errors [1]. Pronunciation acquisition varies significantly across individuals. Factors such as age of onset, exposure environment, phonological awareness, imitation ability, and motivation have all been proven to affect outcomes [4]. Moreover, strong interference from the native phonological system may lead to "phonological fossilization" early in the learning process, causing persistent error patterns that are difficult to overcome.

2.2. Phonological transfer and cross-linguistic influence

Language transfer is a central concept in behaviorist theories of language learning. It was systematically discussed by Lado, who argued that learners inevitably carry over forms, meanings, and even cultural patterns from their first language into their second [5]. According to definitions by Ellis and Odlin, transfer refers to the influence that similarities and differences between the target language and a previously acquired language exert on learning [2]. When structural similarities exist, learning is facilitated through positive transfer. When differences are substantial, negative transfer is likely to occur, resulting in systematic errors.

Phonological transfer is among the most common forms of negative transfer. Numerous studies have shown that learners often substitute unfamiliar target language phonemes with similar sounds from their native language, leading to issues such as voicing confusion, vowel merging, or incorrect intonation patterns [6]. For Chinese learners, phonemes such as /θ/, /ð/, and /v/ are absent in most Chinese varieties, and are often substituted with /s/, /z/, or /w/ respectively, reflecting clear patterns of negative transfer [7]. Moreover, differences between the prosodic features of Chinese and English result in unnatural pacing and misplaced stress in spoken English [8]. The term cross-linguistic influence offers a broader framework, encompassing phenomena such as transfer, avoidance, borrowing, and attrition [9]. In addition, cognitive factors such as language proficiency, processing load, and metalinguistic awareness are also known to mediate the extent and type of transfer observed in L2 learners [10]. Exploring the relationship between L2 pronunciation acquisition and L1 is key to understanding the variation in learners' phonological performance and to developing more effective pronunciation instruction.

3. Pronunciation challenges in different Chinese varieties

3.1. Standard Mandarin speakers' challenges

Standard Mandarin, as the most widely spoken variety of Chinese, has a unique phonological system that significantly affects ESL learners' pronunciation through negative transfer. Existing research consistently indicate that speakers face major difficulties in both segmental and suprasegmental aspects.

DOI: 10.54254/2753-7048/2025.NS27536

In terms of consonants, Standard Mandarin lacks some of the phonemes found in English, which leads learners to produce systematic substitution errors. A study by Chen and Zhang used Praat software to analyze pronunciation and found that Standard Mandarin speakers have significantly more difficulty pronouncing the voiced plosive /d/ at the end of a word than the voiceless /t/ [11]. Common error types include devoicing, such as pronouncing the /d/ in "bed" as /t/, and vowel insertion, such as pronouncing "bad" as /ˈbædə/. This is closely related to the dominance of voiceless consonants in Standard Mandarin's phonological system (17 voiceless vs. 5 voiced). In addition, interdental fricatives /θ/ and /ð/ do not exist in Standard Mandarin. A study by Hamzah et al. compared learners of Chinese in Malaysia and China and found that Chinese learners were more likely to substitute /θ/ with /s/, such as pronouncing "three" as /sriː/ [12]. Liang further pointed out that this phenomenon is particularly prominent among northern speakers [13].

Regarding vowel production, the Standard Mandarin vowel system is primarily composed of monophthongs, with relatively few diphthongs, which leads learners to frequently simplify English diphthongs. A systematic review by Wang et al. indicated that Standard Mandarin speakers often pronounce /eɪ/ as the monophthong /e/, and /aɪ/ as /a/ [14]. Meanwhile, confusion between /iː/ and /ɪ/ as in "sheep" and "ship" is extremely common. A study by Flege et al. showed that even learners with some experience demonstrated significantly lower acoustic contrast between these pairs compared to native English speakers [6]. A study by Liu further found that learners lacked precision in controlling vowel height and backness, causing /iː/ to be produced too low and back, and /ɪ/ too high and front [15].

As for suprasegmental features, standard Mandarin is a tonal language, where pitch variation distinguishes lexical meaning, while pitch is primarily used to indicate stress and intonation in English. A study by the British Accent Academy revealed that Standard Mandarin speakers frequently transfer Chinese tonal patterns into English pronunciation, resulting in incorrect stress placement such as incorrectly placing the stress on the second syllable of "photograph" (noun), or failing to differentiate the stress between "record" as a noun and as a verb [16]. Moreover, Standard Mandarin's syllable-timed rhythm (in which each syllable has roughly equal duration and intensity) conflicts with English's stress-timed rhythm (where stressed syllables are longer and louder). A study by Liu found that this results in Standard Mandarin learners' English pronunciation sounding "choppy" or "monotonous," severely affecting the naturalness of speech flow [15].

In summary, the challenges that Standard Mandarin speakers face in English pronunciation are mainly divided into consonant substitution, vowel simplification, and misalignment of suprasegmental features. These findings highlight the importance of developing targeted teaching strategies, such as specialized training in voiced plosives and interdental fricatives, and using acoustic contrast exercises to enhance accuracy in vowel production and stress placement.

3.2. Cantonese speakers' challenges

As a Chinese dialect with a unique phonological system, Cantonese poses a distinct dialect-specific influence on English pronunciation.

At the segmental level, phonemic differences between Cantonese and English result in systematic substitution errors. The interdental fricatives /θ/ and /ð/ do not exist in Cantonese; speakers often replace them with /f/ or /d/ (e.g., "think" pronounced as /fɪŋk/, "this" as /dɪs/) [17,18]. A more distinctive issue is the absence of the labiodental fricative /v/ in Cantonese. Chen and Li, based on phonetic analyses of 82 English-major students who were native Cantonese speakers, along with Stibbard's study on Hong Kong Cantonese speakers, jointly found that approximately 53% of learners substituted /v/ with /w/ or /f/, pronouncing "vest" as /fest/ or /west/ [18,19]. Moreover,

DOI: 10.54254/2753-7048/2025.NS27536

Cantonese lacks the rhotic /ɹ/, and in some regions, /n/ and /l/ are not distinguished. Through pronunciation tasks with 12 Cantonese ESL learners, Chan found that "red" was often pronounced as /led/, and "night" and "light" were confused [20]. Lan conducted a production experiment involving three age groups of Cantonese speakers (20s, 30s, 60s), judged by three phoneticians, and found that younger speakers (20s) systematically transferred the alveolar affricate /ts/ to the English post-alveolar affricate /tʃ/, especially in back vowel contexts (/ɔ/, /u/) [21]. Furthermore, in Cantonese, checked syllables end in unreleased stop codas /p/, /t/, /k/. Through a comparative recording analysis of 30 Cantonese-speaking students and 20 Northern Mandarin-speaking students, Jia found that Cantonese learners had a high error rate with English final voiceless stops (-p, -t, -k) [22]. Combined with Chen and Li's findings, around 60% of learners either omitted or failed to release such codas, pronouncing "help" as /hel/, and omitting the /d/ in "need" [18]. Additionally, the Cantonese vowel system is simplified and lacks the tense-lax contrast found in English (e.g., /iː/ vs. /ɪ/). Moreover, the diphthong system in Cantonese is limited, often leading to simplification of English diphthongs such as /eɪ/ and /aɪ/ into monophthongs (e.g., "face" pronounced as /fes/) [23].

At the suprasegmental level, as a syllable-timed language, Cantonese speakers tend to assign equal weight to all syllables, leading to features such as failure to de-stress function words or misplaced initial sentence stress [24]. Furthermore, the nine-tone system inherent to Cantonese renders it prone for learners to transfer its tonal patterns to English, thereby resulting in abrupt final intonation [23]. In a language-switching experiment with 60 Cantonese-English bilinguals, Tsui et al. found that phonological features from the dominant language could shift into the non-dominant one, reflecting the influence on English phonology [25].

In summary, the pronunciation challenges faced by Cantonese learners of English are rooted in dialect-specific phonemic inventories, checked syllable-final habits, and prosodic features.

3.3. Sichuan dialect speakers' challenges

As a representative dialect of Southwestern Mandarin, the phonological system of the Sichuan dialect presents both common challenges similar to those faced by Standard Mandarin speakers and unique difficulties due to its own distinctive features.

Sichuan dialect speakers are also affected by similar negative transfer when pronouncing English consonants and vowels as Standard Mandarin speakers. For example, in terms of consonants, both groups struggle with the English interdental fricatives /θ/ and /ð/. Sichuan dialect speakers tend to substitute these sounds with /s/ or /z/, just as Standard Mandarin speakers commonly replace /θ/ with /s/ [26]. Regarding vowels, both groups have difficulty distinguishing between long and short vowel pairs and over 65% of Sichuan dialect speakers fail to differentiate them [27].

However, Sichuan dialect speakers also face unique challenges. In consonants, the most prominent issue is the lack of distinction between /n/ and /l/ as Cantonese speakers. In the Sichuan dialect, /n/ and /l/ are variants of the same phoneme [28], and approximately 55% of speakers substitute /n/ for /l/, while 30% are unable to distinguish the two [27]. This results in English words like "night" being pronounced as "light [30]. Furthermore, there is confusion between /h/ and /f/. In some regions, speakers pronounce English /h/ as /f/, such as "house" pronounced as /faus/, or /f/ as /h/, as in "fish" pronounced as /hɪʃ/ [31]. Another study found that some voiced initials in the Sichuan dialect, such as /b/ and /d/, becoming voiceless /p/ and /t/ respectively, which may lead to English words like "boy" being pronounced as /pɔɪ/ and "dog" as /tɒg/ [32]. In terms of vowels, a distinct feature is the merger of front and back nasal finals, where "ing" is often pronounced as "in" and "eng" as "en", resulting in English words like "sing" being pronounced as /sɪn/ [31]. Additionally, the Sichuan dialect lacks the monophthong "e," and over half of its speakers exhibit

DOI: 10.54254/2753-7048/2025.NS27536

mispronunciations of the English vowels /e/ and /ɔɪ/, whereas roughly 25% produce /ɑː/ and /eɪ/ with inaccuracy [33].

In terms of suprasegmental features, the syllable-timed rhythm of the Sichuan dialect conflicts with the stress-timed rhythm of English, resulting in a "monotone effect" as Standard Mandarin [28]. However, the intonation patterns of the Sichuan dialect are more variable, and this manifests as abrupt sentence-final intonation when transferred to English. For instance, declarative sentences such as "The cat is black." may be pronounced with a rising intonation resembling a question [30]. In addition, as observed by Gong, Sichuan dialect exhibits progressive nasal assimilation in negative constructions, where the bilabial nasal /m/ assimilates to the place of articulation of following consonants [34]. This pattern transfers to English, causing overgeneralization in negative prefixes like mispronouncing "indefinite" as /ɪmˈdefɪnət/ (over-applying bilabial assimilation), violating English's regressive nasal assimilation rules.

In summary, the pronunciation challenges faced by Sichuan dialect speakers include both common issues shared with Standard Mandarin and Cantonese speakers as well as specific difficulties. Instruction should adopt targeted strategies, such as focused training on distinguishing /n/ and /l/, /h/ and /f/.

4. Future outlook

Although recent years have witnessed a growing number of studies on the impact of L1 transfer on English pronunciation acquisition among Chinese learners, several important research gaps remain to be addressed:

First, most existing studies focus on individual segmental substitutions, lacking comprehensive modeling of systematic phonological transfer patterns. For example, while the challenges Standard Mandarin speakers face in producing English interdental fricatives /θ/ and /ð/ have been relatively well-documented, few studies explore the underlying transfer mechanisms from a phonological system perspective.

Second, research on suprasegmental features remains far less developed than that on segmental issues. Although some studies have identified prosodic transfer issues, these investigations tend to be descriptive. There is a lack of quantitative studies using acoustic parameters.

Third, there is a lack of systematic comparison of transfer phenomena across different Chinese dialects. The current literature predominantly focuses on isolated dialect groups, with few cross-dialect studies that reveal shared and distinct mechanisms of phonological transfer. For instance, speakers from three backgrounds all exhibit confusion between long and short vowels, but whether this stems from similar phonological characteristics or from different perceptual strategies remains unclear.

Fourth, research participants are mainly university-level English majors, resulting in a relatively narrow learning stage and language environment. Most studies target learners with intermediate or advanced English proficiency, while younger learners or non-English majors are underrepresented.

5. Conclusion

This review has synthesized existing research on the pronunciation challenges faced by Chinese ESL learners from three representative regional varieties: Standard Mandarin, Cantonese, and the Sichuan dialect. Across these groups, several common difficulties emerge-such as substitution of unfamiliar English consonants (e.g., /θ/, /ð/, /v/), confusion between long and short vowels, and rhythm-related prosodic issues. At the same time, each variety contributes distinct phonological

DOI: 10.54254/2753-7048/2025.NS27536

features that shape specific error patterns, such as checked syllable codas in Cantonese or the /n/-/l/ merger in Sichuan dialect.

A central theme that emerges across the literature is the pervasive influence of phonological transfer, where learners' L1 sound systems interfere with the acquisition of English pronunciation. This transfer operates at both segmental and suprasegmental levels, with varying degrees of persistence and perceptual impact. While the existing studies have contributed valuable insights, many remain focused on isolated features or learner groups. There is a pressing need to expand the empirical scope, especially through cross-dialectal comparison, quantitative analysis of suprasegmental features, and inclusion of younger or non-major learners. Given the continued importance of intelligible pronunciation for language learners, the findings reviewed here underscore the need for dialect-sensitive pedagogical strategies. Teachers should be equipped with knowledge of regional transfer patterns and trained to use contrastive phonological techniques. Equally, learners stand to benefit from instructional interventions that foster phonological awareness and deliver explicit training in sounds and rhythm patterns that pose challenges.

In summary, by consolidating existing research, this review contributes to a more comprehensive understanding of L1 transfer in pronunciation among Chinese ESL learners. It also lays the groundwork for more inclusive, evidence-based teaching and future studies that better reflect the linguistic diversity of China's English learning population.

References

[1] Ellis, R. (1994). The Study of Second Language Acquisition. Oxford University Press, Oxford.
[2] Derwing, T. M., Diepenbroek, L. G., & Foote, J. A. (2013). How Well do General-Skills ESL Textbooks Address Pronunciation?. TESL Canada Journal, 30(1): 22.
[3] Odlin, T. (1989). Language transfer: Cross–linguistic influence in language learning. Cambridge University Press, Cambridge.
[4] Best, Catherine & McRoberts, Gerald & Goodell, Elizabeth. (2001) Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. The Journal of the Acoustical Society of America, 109: 775–94.
[5] Lado, R. (1957) Linguistics across Cultures: Applied Linguistics and Language Teachers. University of Michigan Press, Ann Arbor.
[6] Flege, J. E., Bohn, O.–S., & Jang, S. (1997). Effects of experience on non–native speakers' production and perception of English vowels. Journal of Phonetics, 25(4): 437–470.
[7] Chen, Y. (1987). A Comparison of Chinese and English Pitch. Foreign Language Education, 1: 20–29.
[8] Wang, G. Z. (1990). The characteristics of Chinese and English sound fluctuations and frequency mode and its influence on English Pronunciation Teaching. Modern Foreign Languages, 1: 89–100.
[9] Sharwood Smith, M., & Kellerman, E. (1986). Crosslinguistic Influence in Second Language Acquisition. Pergamon Press, New York.
[10] Gass, S., & Selinker, L. (2001). Second Language Acquisition: An Introductory Course (2nd ed.). Lawrence Erlbaum, Mahwah, NJ.
[11] Cheng, Y., & Zhang, W. (2024). Final plosive production by Mandarin EFL learners in oral English teaching. Overseas English, (24): 52–54.
[12] Hamzah, M. H., Ahmad, A., & Yusuf, M. H. (2017). A Comparative Study of Pronunciation among Chinese Learners of English from Malaysia and China: The Case of Voiceless Dental Fricatives /θ/ and Alveolar Liquids /r/. Sains Humanika, 9 (4–2): 1–9.
[13] Liang, E. L. (2014). Pronunciation of English Consonants, Vowels and Diphthongs of Standard Mandarin – Chinese Speakers. Studies in Literature and Language, 8 (1): 62–65.
[14] Wang, H., Knight, R. A., Dipper, L., et al. (2025). Systematic review: The identification of segmental Mandarin–accented English features. Speech Communication, 167: 103–168.
[15] Liu, Y. (2022). The influence of L1 prosody on L2 English rhythm production. Journal of Second Language Pronunciation, 8(2): 187–205.

DOI: 10.54254/2753-7048/2025.NS27536

[16] British Accent Academy. (2023) Most Common English Pronunciation Errors Made by Mandarin Speakers. https://www.britishaccentacademy.com/most–common–english–pronunciation–errors–made–by–mandarin–speakers
[17] Chan, A. Y. W., & Li, D. C. S. (2000) English and Cantonese phonology in contrast: Explaining Cantonese ESL learners' English pronunciation problems. Language, Culture and Curriculum, 13(1): 67–85.
[18] Chen, J., & Li, C. (2021). A study on the negative transfer in English consonant acquisition by Cantonese L1 speakers based on ELF theory: A case study of Nanguo Business College of Guangdong University of Foreign Studies. Overseas English, (2): 3–6.
[19] Stibbard, R. (2004) The spoken English of Hong Kong: A study of co–occurring segmental errors. Language, Culture and Curriculum, 17(2): 127–142.
[20] Chan, A. Y. W. (2006) Cantonese ESL learners' pronunciation of English final consonants. Language, Culture and Curriculum, 19(3): 296–313.
[21] Lan, Y. (2017) Foreign Influence and Sound Change: A Case Study of Cantonese Alveolar Affricates. PACLIC, 31: 155–160.
[22] Jia, S. (2011). An experimental study on the negative transfer of Cantonese checked syllable codas to English voiceless stop finals. Journal of Guangdong Second Normal University, 31(4): 74–79.
[23] Wong, S. W. L., Dealey, J., Leung, V. W. H., & Mok, P. K. (2019) Production of English connected speech processes: an assessment of Cantonese ESL learners' difficulties obtaining native–like speech. The Language Learning Journal, 47(4): 453–470.
[24] Zhang, J. (2017). The influence of Cantonese on English phonetics and intonation from the perspective of transfer theory and corresponding strategies. Journal of Changsha University, 31(3): 155–158.
[25] Tsui, R. K. Y., Tong, X., & Chan, C. S. K. (2019) Impact of language dominance on phonetic transfer in Cantonese–English bilingual language switching. Applied Psycholinguistics, 40(1): 29–58.
[26] Huang, Jing & Shibata, Kye & Hsieh, Feng–fan & Chang, Yueh–Chin & Tiede, Mark. (2023). The L~N merger in Southwestern Mandarin: An articulatory study.
[27] Qin, Q. (2014) A Probe into the Negative Impact of Sichuan Dialect on English Phonetics Learning and Countermeasures. Studies in Literature and Language, 9: 16.
[28] Ma, C. D., Tan, L. H. (2013) Analysis on Sichuan College Students' Segmental Phoneme Errors. Creative Education, 4: 415–417.
[29] Ma, C., & Tan, L. (2015). Comparison on the word stress patterns and functions between English and Sichuan dialect. Open Journal of Social Sciences, 3(10), 33–39.
[30] Zeng, J. (2023) International Intelligibility of English Spoken by Sichuan English Learners. International Journal of Arts Humanities and Social Sciences Studies, 8: 190–195.
[31] Huang, J., Wang, L., & Li, X. (2018). Phonological features of Sichuan–Chongqing dialects and their influence on Korean pronunciation acquisition. Journal of Yanbian University (Social Sciences Edition), 51(4): 78–85.
[32] Qu, T. (2019) A comparison of /b/ and /d/ initials between Mandarin and Sichuan dialect. Journal of Youth Studies, (5): 79–82.
[33] Jin, J., & Liu, D. (2021). Negative transfer of Sichuan dialect in English vowel acquisition. In 2021 International Conference on Asian Language Processing (IALP) (pp. 247–251). IEEE.
[34] Gong, Y. P. (2022) Brief exploration of nasal assimilation phenomenon in negative prefixes: Comparison between Chinese dialect & English. Journal of Literature and Art Studies, 12(3): 289–291.