

Differentiated Difficulties Instead of Hierarchical Gaps: A Rebuttal to the Bias About Dialect Speakers' English Pronunciation Learning

Hanwen Zhang*

Beijing Institute of Technology, Beijing, China

**Corresponding author: Hanwen Zhang.*

Abstract

This paper aims to rebut a social cognitive prejudice that dialect-speaking learners face greater difficulties in acquiring English pronunciation than Mandarin-speaking learners. To correct such misconceptions, the paper introduces several theories, which collectively prove that pronunciation acquisition only generates differentiated difficulties rather than hierarchical gaps, including the Speech Learning Model (SL) and the Revised Speech Learning Model (SLM-r), positive and negative transfer, phonological sieve and contrastive analysis. This study chooses Cantonese-speaking learners and Wu Dialect-speaking learners as samples and adopts an interdisciplinary perspective combining sociolinguistics and second language acquisition to challenge that bias. To avoid misunderstanding, key concepts such as regional dialect speakers, standard Mandarin speakers and accurate English pronunciation, are redefined. After analyzing both positive and negative transfer of Mandarin, Cantonese and Wu Dialect speakers, the study confirms its core proposition, there are only differences in learning problems rather than hierarchical distinctions. The study contributes to distinguishing the relationship between Chinese dialects and L2 pronunciation acquisition, challenges the Mandarin-centered evaluative framework, strives to provide support for forming comprehensive English pronunciation instruction, and hopes to eradicate the social stereotype.

Keywords

English pronunciation acquisition, regional dialect, mandarin Chinese, language transfer, pedagogical implication

1. Introduction

Nowadays, a prevailing misconception has increasingly gained popularity, sparking a heated debate, that people who master regional dialects face more difficulties in acquiring accurate pronunciation of English as a second language (L2) than speakers who are proficient in Standard Mandarin. It is well known that an array of reasons might influence second language learning, especially pronunciation acquisition. According to the Speech Learning Model (SL), Hypothesis 2, phonetic categories are cognitive units that store speech characteristics specific to a language in long-term memory [1]. Speaking specifically, within a single language, distinct dialectal variations provide multiple phonetic category systems that are stored in learners' long-term memory. In contrast to systemic gaps in overall acquisition difficulty, mismatches between the phonetic categories of different first-language dialects and the target second-language result in uneven matching

relations across speech segments. Therefore, the learning gaps between Mandarin and dialect speakers contribute to the inconsistent matching degrees with English phonemes. Furthermore, from Flege's perspective, Hypothesis 3 of the Speech Learning Model (SL) states that phonetic categories formed in early childhood are not fixed [1], which means speakers' original categories will be reshaped and upgraded via continuous English input and actual pronunciation practice; hence, no learners would be trapped permanently by the phonological system of their first language (L1). This theoretical finding lays a basis for challenging the stereotype mentioned at the beginning. Overall, it is a stereotype that society regards regional dialect speakers as struggling more in English (L2) pronunciation acquisition than Mandarin speakers. This bias brings potential disadvantages to English phonetic instruction and language acquisition of dialect speakers, embodying an inferiority complex, confidence loss or even feeling resistant when learning pronunciation. Consequently, with the help of previous theories and studies, this article aims to break the stereotype and analyze the differential impacts of both dialect and Mandarin phonological systems on English pronunciation acquisition, emphasizing that there are only differences in learning problems rather than hierarchical distinctions. Meanwhile, the passage will spare no effort to correct the cognitive bias in terms of the relationship between dialects and L2 pronunciation acquisition, while it will provide references for English phonetic instruction, reminding teachers to formulate targeted strategies for pronunciation correction and avoiding teaching discrimination.

2. Theoretical Foundations and Core Concepts

2.1 Theoretical Foundations

2.1.1 The Speech Learning Model (SL) and the Revised Speech Learning Model (SLM-r)

Based on the four essential hypotheses of the *Speech Learning Model* (SL), Flege demonstrates that the learners might easier identify the distinctions between the sounds of L2 and the closest L1 via the notable acoustic difference of them; furthermore, with the ability to detect these differences, they are able to establish new phonetic categories for the L2 sounds, rather than classifying the L2 sounds into an existing L1 category, namely equivalence classification, for failure of discernment [1]. Consequently, three groups of speakers will encounter distinct types of pronunciation difficulties distributed across various phonological dimensions, with both advantages and disadvantages involved. Parts of English phonetics are markedly different from Mandarin phonetics, leading to a new phonetic category; meanwhile, that might be highly similar to Cantonese or Wu Dialect phonetics, causing the equivalence classification. Since there is no universal criterion to judge the level of difficulty, none of them has an overall acquisition difficulty. Complemented by the Revised *Speech Learning Model* (SLM-r), multiple-level interactions, including segmental, suprasegmental and phonotactic, made by L1 and L2 phonetic systems bring ultimately different influence to learners, individuals might show several strengths in some sections while show weaknesses in others, due to the compatibility between their L1 and L2 phonetic systems, proving that no evident reasons to assume that one L1 background will lead to overall greater difficulty in L2 pronunciation than another [2]. Therefore, though the L1 backgrounds of Mandarin, Wu Dialect and Cantonese speakers are diverse, learners with various strengths and different weaknesses do not face problems of hierarchical distinctions.

2.1.2 Positive Transfer and Negative Transfer

Briefly, each L1 has its specific positive transfer and the learners have to face numerous challenges due to negative transfer. It is Weinreich who first defined interference phenomena as a result of language contact, including positive interference (the structures of two languages are similar and it facilitates the acquisition of L2), negative interference (structures are different and lead to mistakes) and zero interference (L1 has no corresponding structure). Later, researchers developed the concepts people are more familiar with, positive transfer, negative transfer and zero transfer. According to Weinreich, interference is a matter of selective influence in particular linguistic areas rather than ubiquitous transfer; while the type of interference of each linguistic level is dominated by the structural relations between the systems in contact [3]. Hence, Mandarin, Wu Dialect and Cantonese with various structures and systems might bring different phonological interference both positive and negative. Even within the same linguistic level, every distinguished interference is independent of each other; as a result, among different dimensions, languages just exhibit their respective complementary strengths and weaknesses rather than an overall hierarchical distinction of difficulty.

2.1.3 Phonological Sieve

L1 with different features of the phonological system leads to differences in learning problems rather than hierarchical distinctions. Nikolai Sergeyevich Trubetzkoy, the linguist who first compared the phonological system of a language to a sieve for it determines what is phonologically relevant, regarded that speakers might focus on certain distinctive features (referring to those properties of sounds related to distinguish meaning) and deal with all other languages via the grid of their own language [4]. With specific sieves, in English pronunciation acquisition, Mandarin, Cantonese and Wu Dialect speakers take various advantages of their own grid of languages and pay attention to distinctive features respectively. Several distinctive features in one phonological system might be the non-distinctive features, being irrelevant for phonological purposes and filtered out of the process, in another. Therefore, dialect and Mandarin speakers just face different types of difficulties rather than distinct degrees.

2.1.4 Contrastive Analysis

Robert Lado, who made a landmark contribution to the field of linguistics instruction, put forward the idea of *Contrastive Analysis* (CA) in his book *Linguistics across Cultures*, arguing that the structural differences between the target language system and the native language system are the most significant cause of learners' challenges when learning a foreign language [5]. In phonological acquisition, according to Lado, difficulties distribute across different phonetic and prosodic dimensions rather than appearing as an overall higher difficulty level. The prevailing stereotype that dialect speakers struggle more in English pronunciation acquisition fails to adhere to the fundamental tenet of CA, perplexing the social standardization of language with phonological learning difficulty.

2.2 Core Concepts

2.2.1 Regional Dialect Speakers

A regional dialect speaker is defined as an individual whose L1 is a geographically specific variety of Chinese, such as Cantonese or Wu Dialect, who habitually uses that dialect as the primary medium of daily communication, and who may also be proficient in Mandarin.

2.2.2 Standard Mandarin Speaker

Standard Mandarin Speaker mentioned in this article does not refer to the idealized group of speakers without any accent pronunciation excelling in the standards of *PUTONGHUA SHUIPING CESHI* (PSC, a test for proficiency in Mandarin pronunciation). But it broadly refers to native Chinese speakers who use Mandarin as their primary language in daily communication. In terms of daily usage, they might have slight regional dialect accents and encounter confusions including the confusions between nasal sounds and between alveolar and retroflex sounds, but are still fundamentally distinct from regional dialect speakers.

2.2.3 Accurate English Pronunciation

Accurate English Pronunciation reflects the pronunciation that follows the standard pronunciation norms of native English speakers (such as Received Pronunciation or General American). The speakers or learners can clearly distinguish the core phonemes and master the basic prosodic rules (linking, weak form and stress).

3. The Impact of Mandarin on English Pronunciation Acquisition

3.1 Phonological Features of Mandarin

The major feature of Mandarin phonology is fundamentally syllable-based rather than segment-based, like English. According to Robert L. Cheng, the early generative study shows that syllables serve as the main unit of phonological representation and generation while syllable grammar is the best way to describe Mandarin phonological structure [6]. In terms of this structural perspective, contemporary views reflect that the syllable structure of Mandarin with high constraint significantly conforms to CV patterns (a consonant + a vowel, like “妈” <*mā*> refers to mother or mam, consisting of consonant <*m*> and vowel) and lacks consonant clusters. It is also worth noting that each syllable is compulsively associated with a lexical tone. Moreover, the number

of codas is limited, consisting primarily of /n/ and /ŋ/; hence, Mandarin, as a tonal language, has a relatively small syllable inventory compared to English, a non-tonal language [7].

3.2 Negative Transfer

According to Weinreich, negative transfer refers to L1 phonological structures being really different from L2's, which leads to systematic deviations in L2 pronunciation [3]. The Mandarin phonological system, as a syllable-based organization, lacks consonant clusters. Such phonotactic constraints imply that pronunciations like /pleɪ/ (play), /stri:t/ (street) or /sprɪŋ/ (spring), the initial cluster is /pl-/ , /str-/ or /spr-/; while learners (Mandarin speakers) might possibly pronounce [pəleɪ], [səpəɪŋ] or [sətəɪ:t], because Mandarin does not have the phonological structure /pl-/ , /str-/ or /spr-/ [6]. Therefore, learners resort to syllable repair strategies, a kind of phonological modification employed by speakers to use wrong syllable structures to conform to the phonotactic constraints of their native language, to help their pronunciation acquisition. The occurrence of vowel epenthesis reflects that Mandarin speakers implement native syllable templates to deal with the unfamiliar English consonant clusters rather than an inability to pronounce the targeted consonant. With the limited number of codas, herein lies a tendency that Mandarin speakers might delete the final consonants of English words; for example, learners might pronounce [kæ] instead of /kæt/ (cat) or [mæ] instead of /mæp/ (map). This deletion derives from the prohibition of stop consonants such as /p/, /t/ or /k/ in the final position of syllables in Mandarin, causing the English codas to be simplified to match their familiar structures.

Flege's *Speech Learning Model* (SL) mentioned above could clearly explain the segmental transfer in Mandarin-speaking learners' English pronunciation, illustrating that L2 sounds are created and perceived in connection with the existing L1 phonological categories [1]. Mandarin lacks the phoneme /θ/ which is frequently used in English syllables; as a result, Mandarin speakers have difficulty uttering /θɪŋk/ (think) and transform /θ/ to /s/ or /t/. Also, they pronounce /d/ to replace /ð/ (/ðɪs/ this) and /w/ to replace /v/ (/vəri/ very) due to the absence of consonants. Therefore, learners are apt to assimilate the unfamiliar English phoneme into the closest available Mandarin categories, causing a consonant substitution.

Another challenge Mandarin speakers have to face in English pronunciation acquisition is lexical stress. Mandarin as a tonal language has four tones for differentiating words' meanings, which means the same consonants and vowels lead to various sounds due to the tone; for example, <mā> and <mǎ> refer to “妈 (mother or mam)” and “马 (horse)”. However, compared to English the intonation language, Mandarin does not have lexical stress. Researchers find that although Mandarin speakers employ acoustic cues such as duration, intensity, fundamental frequency and vowel quality, they still produce less native-like stress patterns [8]. For instance, “object” with two classes, both noun (OBject) and verb (obJECT), its stress position changes the acoustic prominence of the syllables. Mandarin speakers produce such contrast differently from native speakers or even neglect this detail.

3.3 Positive Transfer

Although Mandarin and English are significantly different from each other, they share several consonantal articulatory features. As a number of shared consonants with similar articulatory places and manners, Mandarin-speaking learners meet few troubles acquiring proficiency in them, even at their initial stage of learning. Take /m/, /n/, /f/ and /l/ for instance, all of them occur in both Mandarin and English with highly analogous articulatory gestures. When the sounds of L2 are considerably similar to L1 categories, L2 sounds are assimilated to existing L1 phonological categories rather than being regarded as an entirely new one, which is in line with Flege's study. Some English monophthongs, a speech sound that consists of only one vowel sound, are akin to Mandarin vowels in acoustical and articulatory terms; hence this allows Mandarin-speaking learners to assimilate L2 sounds to the existing L1 vowel categories during their acquisition. For example, /i/ in Mandarin and /i:/ in English share strikingly similar tongue position and lip configuration, truly bringing benefits to the first-stage learners.

4. The Impact of Regional Dialects on English Pronunciation Acquisition

4.1 Cantonese and English Pronunciation Acquisition

4.1.1 Phonological Features of Cantonese

With a rich and highly organized phonological inventory at the segmental level, researchers propose that the Cantonese consonantal system is essentially structured on an aspiration contrast instead of the voicing contrast [9]. In addition to maintaining a well-developed inventory of nasals, fricatives, glides, and liquids, stops occur in aspirated and unaspirated pairs, such as /p/–/p^h/, /t/–/t^h/, and /k/–/k^h/ . The vowel system includes front-rounded vowels like /y/ and /œ/, as well as differences in vowel height, backness, and lip rounding. The segmental system of Cantonese is distinguished by its internal regularity and stable phonology rather than just having a large number of phonemes. With strongly constrained syllable structure, one of the most distinctive features of Cantonese phonology, researchers demonstrate that the Cantonese syllables conform to a regular phonotactic pattern to a high degree [9]. Meanwhile, some also highlight that it is vital to attach great importance to syllable structure as a central organizing rule of Cantonese, whose syllables consist of an optional onset, a vocalic nucleus and a highly restricted coda [10]. There are a limited number of syllables that could serve as a coda, involving nasal /m, n, ŋ/ and the unreleased stop /p, t, k/; therefore, Cantonese phonology reflects a high degree of structural regularity with syllables serving as the principal domain of phonological organization. Cantonese, as a tone-dominant language, is different from many other Chinese varieties in the following aspects [11]. To begin with, there is a severely complex tonal specification throughout the language for almost every syllable is associated with a lexical tone. Secondly, there is no phonological counterpart of a neutral-tone category similar to that found in Mandarin for Cantonese lacks a systematic distinction between stressed and reduced syllables. Thirdly, the language shows a close relationship between syllables and lexical units and a strong trend toward monosyllabicity.

4.1.2 Negative Transfer

One of the distinct challenges in plosive acquisition is clearly recognized as final voiced stop devoicing. Study shows that because of the lacking voiced plosives in Cantonese phonological inventory and only voiceless plosives given the access to the final position of the syllables, Cantonese-speaking learners persistently substitute the final voiced plosives /b, d, g/ in English with their voiceless counterparts /p, t, k/, which directly neutralizes the voicing contrast in minimal pairs like rope/robe and mate/maid [12]. However, syllable-initial voiced plosives also undergo devoicing to a lower extent. Researchers also point out that at the phonological level, all final plosives in Cantonese are unreleased, which means no airflow is produced following the formation of the articulatory closure [12]. Cantonese learners negatively transfer this L1 articulatory habit to English, generating English terminal plosives with a “swallowed” unreleased quality, with /step/ (step) pronounced as [step̚] and /nɒt/ (not) as [nɒt̚] as typical examples, as further explained in the section on learning challenges. This transfer further worsens the semantic confusion caused by devoicing and weakens the voicing contrast of final plosives. The paper explicitly documents inter-group variation in Cantonese phonology: especially among younger Cantonese speakers, the contrast between syllable-initial nasal /n/ and lateral /l/ is neutralised, and the two sounds function as free variants. Researchers demonstrate explicitly the inter-group variation in Cantonese phonology, especially among the younger speakers, neutralization of the distinction between syllable-initial nasal /n/ and lateral /l/, functioning as free variants [12]. Such negative transfer of L1 leads learners to commonly substitute word-initial /n/ with /l/ in English, confusing word pairs like nine/line and knife/life. Meanwhile, word-initial /l/ may also take on a nasal quality and surface as a nasalized lateral.

4.1.3 Positive Transfer

As it has been mentioned above, the Cantonese plosive system, whose core distinct feature is aspiration, only involves voiceless plosives, forming a phonemic opposition between aspirated /p^h, t^h, k^h/ and unaspirated /p, t, k/ [9]. Such a phonological feature enables Cantonese-speaking learners to map onto the two sets of allophones of English voiceless plosives. English voiceless plosives /p, t, k/ appear as unaspirated allophones after /s/ (e.g., speak, stay, sky), yet must be aspirated in stressed syllable-initial position (e.g., peak, team, key). As native Cantonese speakers, who do not need to learn new articulatory gestures, they have already acquired the phonemic-level skill, the articulatory regulation of aspiration. Therefore, learners could naturally grasp the

allophonic patterns of English voiceless plosives by mapping L1 aspirated plosives onto English word-initial voiceless plosives, while mapping L1 unaspirated plosives onto post-/s/ plosives. Compared to L1 backgrounds, a phonemic aspiration confers an exceptional positive transfer advantage. As a prototypical tone language, Cantonese with six distinctive lexical tones, involving pitch height, direction and contour, straightforwardly determines lexical meaning [9]. Native speakers grasp pitch perception and production control maturely due to their daily and enduring L1 acquisition; hence, they obtain relatively higher sensitivity to fundamental frequency variation. Such property fosters positive transfer, so that Cantonese-speaking learners are more likely to detect differences in English intonational contours. Furthermore, intonational patterns, like declarative falling tone, interrogative rising tone and fall-rise tones, could be more accurately distinguished by learners whose performance in intonation perception and imitation is much more perfect.

4.2 Wu Dialect (Jiangsu-Zhejiang-Shanghai Area) and English Pronunciation Acquisition

4.2.1 Phonological Features of Wu Dialect

The preservation of a contrast between voiced and voiceless obstruents is one of Wu Dialect's most notable phonological characteristics and phonemic contrasts like /p-b/, /t-d/, and /k-g/ exist in many Wu varieties. Acoustic analyses confirm that the so-called "voiced" consonants are consistently differentiated from voiceless unaspirated consonants, proving that the voiced obstruent contrast is still a fundamental element of Wu phonology [13]. With contrasts such as /pa/ versus /ba/ or /ta/ versus /da/ in some Wu varieties, Wu speakers routinely identify consonants based on voicing. The phonological function of phonation type is another significant aspect of the Wu Dialect. Researchers contend that breathy phonation and other laryngeal changes are often linked to the so-called voiced consonants in Wu, rather than only vocal-fold vibration [13]. Spectral analyses find significant acoustic variations at the onset of vowels following voiced and voiceless consonants, revealing that the contrast incorporates both phonation and voicing differences. For instance, a syllable can be realized as a modal [pa] or as a breathy [ba]. As Wu speakers, therefore, the difference includes variations in voice quality as well as whether or not the vocal folds vibrate. The register system of Wu Dialect is a further essential feature. Regarding the investigation, the seven citation tones, which can be classified into upper register and lower register (the lower ones are frequently characterized by depressed pitch onsets and non-modal phonation, such as breathy voice), are distinguished by pitch contours [14]. For instance, even when the two syllables exhibit falling pitch contours, one might be in the upper register and the other in the lower. Typically produced with a lower pitch onset, the latter one may also show breathy phonation.

4.2.2 Negative Transfer

One of the major challenges Wu-speaking learners have to overcome is over-reliance on phonation type during speech categorization. The distinction between voiced and voiceless obstruents in Wu Dialect is accompanied by voicing as well as by variations in voice quality, like breathy phonation [13]. Furthermore, researchers discover that Wu-speaking learners actively use voice-quality cues in speech classification, and that phonation constitutes a perceptually significant dimension across languages [15]. Hence, instead of solely relying on two crucial cues in English stop perception, aspiration and voice onset time (VOT), learners might fully comprehend English consonantal differences through phonation. However, this might cause under-aspiration of voiceless stops, non-native breathy realizations of voiced stops, and over-reliance on voice quality for identifying English obstruents. For example, learners produce [bæt] rather than /bæt/ (bat) due to over-reliance on phonation. As Wu Dialect preserves a phonemic voicing contrast, the speakers acquire a stable category of voiced consonants from an early age. According to Flege's *Speech Learning Model* (SL), learners often assimilate similar L2 sounds into an existing L1 category rather than establishing a new phonetic category [1]. Consequently, learners might combine several English voiced consonants into a single L1 category based primarily on voicing, ignoring variations in articulatory manner. For example, learners overlook the fricatives in English and concentrate their focus on voiced only, producing [w] instead of /v/, [d] instead of /ð/, or [dz] instead of /z/.

4.2.3 Positive Transfer

The systematic description of Wu Dialect (Lili Wu Dialect, spoken in a town named Lili) states that an essential phonetic characteristic of Northern Wu Dialect is the three-way laryngeal contrast of obstruents (including stops and affricates), voiceless unaspirated, voiceless aspirated and voiced [16]. With the stable phonation patterns and perceptual categories of voiced obstruents, native Wu speakers can detect the phonatory

distinctions between voiceless and voiced obstruents. Such L1 phonetic experience is beneficial to L2 phonetic acquisition, accelerating learners' proficiency in the voicing contrast system of English stops and affricates. For example, the minimal stop contrast in Wu, voiceless unaspirated /p/ and voiced /b/, can be transferred to the distinction between English /p/ (pen /pen/) and /b/ (ben /ben/). The Wu Dialect's fricatives exhibit a two-way laryngeal contrast between voiceless and voiced, according to the acoustic study. In word-initial position, voiceless fricatives are produced with modal voice, whereas voiced fricatives are accompanied by a slight breathiness. Meanwhile, acoustic measurements reveal that the proportion of frication duration in total syllable duration is significantly higher for voiceless fricatives than for voiced ones, forming a stable acoustic contrast [16]. The articulatory control and perceptual markers of fricative voicing contrast have been mastered by native Wu speakers. Consequently, learners may appropriately create and comprehend paired voiceless and voiced fricatives in English with the aid of this L1 experience.

4.3 Brief Discussion

Mandarin-speaking learners and dialect-speaking (Cantonese and Wu Dialect) learners encounter significantly different types of challenges in terms of English pronunciation acquisition, yet each group develops its distinct advantages. According to the existing experimental data and findings, people could not explicitly conclude that dialect-speaking learners face greater difficulties than Mandarin-speaking learners in acquiring the pronunciation of English as an L2. Though dialect speakers truly exhibit negative transfer due to their L1 phonological categories, they simultaneously possess positive transfer which Mandarin speakers do not demonstrate. Hence, herein solely lie different types rather than different degrees.

5. Pedagogical Inspirations

In terms of pedagogy, people should critically reconsider the existing educational structure of English pronunciation acquisition, or the implicit Mandarin-centered structure. According to Rosina Lipi-Green, a standard language is an ideological construct influenced by social power dynamics rather than a structurally superior form of language [17]. Mandarin is unintentionally regarded as the standard for measuring other dialects, because Mandarin is gradually viewed as the sole phonological starting point for Chinese learners. However, such equivalence without linguistic justification is an extension of Standard Language Ideology into pedagogical practice. Therefore, instruction should acknowledge dialects like Cantonese and Wu Dialect as legitimate linguistic systems rather than as departures from Mandarin. It is important to avoid the imposition of a single norm that might force the learners under pressure to perform well. Secondly, to prevent misunderstandings of L1 transfer processes, English pronunciation instruction should move from holistic assessments of learner ability to a componential study of phonological competence. Terence Odlin proposes that language transfer, as a selective phenomenon, takes place in subsystems like phonology, lexicon, and syntax, yet not as a single process that affects the whole linguistic system [18]. Hence, it is a simplification and misunderstanding of second language acquisition processes to regard a specific dialect as naturally making English harder to learn. In terms of reality, various Chinese varieties make different impacts on English pronunciation, which are typically restricted and influence syllable structure or prosodic features rather than an overall linguistic ability. Pronunciation instruction should break down competence into multiple levels and pinpoint dialect-specific transfer patterns, which helps to foster much more targeted pedagogical intervention. Finally, the bias that connects accentedness with linguistic deficiency needs to be revised in English pronunciation instruction. There is no notable relation between accentedness and intelligibility in L2 speech, which means that a more noticeable accent does not definitely lead to less effective communication [19]. However, in reality, listeners evaluate language ability based on the prominence of phonetic distinctions, leading to systematic cognitive bias. With certain dialect backgrounds, learners might display perceptually apparent phonetic elements in English pronunciation, which are often misinterpreted as having lower overall proficiency. Accordingly, assessment systems in pronunciation instruction should accurately reflect learners' language competency via shifting focus from native-likeness to communicative intelligibility.

6. Conclusion

Throughout the whole study, the paper aims to confirm again that there are only differences in learning problems rather than hierarchical distinctions, and the widespread point that dialect-speaking learners face

greater challenges than Mandarin-speaking learners is just a stereotype or prejudice. Mandarin as the standard language in China, still exhibits negative and positive transfer in English pronunciation acquisition as well as other Chinese varieties, such as Cantonese and Wu Dialect. Challenges learners have to overcome are determined by their L1 phonological categories rather than the L1 itself. In conclusion, this study tries hard to address cognitive bias and provide a relatively objective analysis of the impact of Mandarin, Cantonese and Wu Dialect on English pronunciation learning, in terms of phonological features and positive/negative transfer of L1. Furthermore, the findings also try to offer practical, differentiated, and just guidance for English phonics instruction. The limitations of the paper are explicit, as it only involves two kinds of Chinese varieties, Cantonese and Wu Dialect, resulting in an insufficient sample size. However, it is worth noting that the numerous dialects in China help to enrich Chinese cultural diversity. It is a pity to fail to integrate specific linguistic analysis, such as comparing pronunciation recordings between different learning groups, and solely rely on theoretical analysis. Moreover, this study does overlook individual elements including personal learning motivation, personal sensitivity to pronunciation and so on. And the analyses of negative transfer and positive transfer of Mandarin, Cantonese and Wu Dialect are not comprehensive; while this study just selects and lists the most distinctive aspects to facilitate the understanding of readers. Meanwhile, the analyses of phonological features display a similar shortcoming. Regarding the future, it is prudent to expand the coverage of dialects, exploring their impact on English pronunciation acquisition in comparison with Mandarin, which could contribute to a much more accurate and scientific outcome. Focus should also be put on the combination with corpora such as recordings and practical tests so as to conduct quantitative analysis of pronunciation errors. Furthermore, the interactions between individual differences and L1 background are worth investigating as well, which possibly influence L2 pronunciation acquisition.

References

- [1] Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (Vol. 92, No. 1, pp. 233–277). York Press.
- [2] Flege, J. E., Aoyama, K., & Bohn, O. S. (2021). The revised speech learning model (SLM-r) applied. In O. S. Bohn & J. E. Flege (Eds.), *Second language speech learning: Theoretical and empirical progress* (pp. 84–118). Cambridge University Press.
- [3] Weinreich, U. (2010). *Languages in contact: Findings and problems* (No. 1). Walter de Gruyter.
- [4] Trubetzkoy, N. S. (1969). *Principles of phonology*. University of California Press.
- [5] Lado, R. (1957). *Linguistics across cultures; applied linguistics for language teachers*. University of Michigan Press.
- [6] Cheng, R. L. (1966). Mandarin phonological structure. *Journal of Linguistics*, 2(2), 135–158. <https://doi.org/10.1017/S0022226700001444>
- [7] Neergaard, K. D., & Huang, C.-R. (2022). Mandarin Chinese syllable structure and phonological similarity: Perception and production studies. In C.-R. Huang, Y.-H. Lin, I.-H. Chen, & Y.-Y. Hsu (Eds.), *The Cambridge handbook of Chinese linguistics* (pp. 245–274). Cambridge University Press.
- [8] Zhang, Y., Nissen, S. L., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *The Journal of the Acoustical Society of America*, 123(6), 4498–4513. <https://doi.org/10.1121/1.2902165>
- [9] Matthews, S., & Yip, V. (2011). *Cantonese: A comprehensive grammar* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203835012>
- [10] Cheung, W. H. Y., Wee, L.-H., & Zhao, K. (2024). The phonology of Cantonese. In *Reference Module in Social Sciences*. Elsevier. <https://doi.org/10.1016/B978-0-323-95504-1.00209-X>
- [11] Beckman, M. E., Wong, W. Y. P., & Chan, M. K. M. (2005). An autosegmental-metrical analysis and prosodic annotation conventions for Cantonese. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 271–300). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199249633.003.0010>

- [12] 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. <https://doi.org/10.1080/07908310008666590>
- [13] Cao, J., & Maddieson, I. (1992). An exploration of phonation types in Wu dialects of Chinese. *Journal of Phonetics*, 20(1), 77–92. [https://doi.org/10.1016/S0095-4470\(19\)30255-4](https://doi.org/10.1016/S0095-4470(19)30255-4)
- [14] Rose, P., & Yang, T. (2024). Diagnosing dominance: Problematic sandhi types in the Chinese Wu dialect of Jinshan. *Proceedings of the Linguistic Society of America*, 9(1), 5667. <https://doi.org/10.3765/plsa.v9i1.5667>
- [15] Esposito, C. M., & Khan, S. D. (2020). The cross-linguistic patterns of phonation types. *Language and Linguistics Compass*, 14, e12392. <https://doi.org/10.1111/lnc3.12392>
- [16] Shi, M., & Chen, Y. (2022). Lili Wu Chinese. *Journal of the International Phonetic Association*, 52(1), 157–179. <https://doi.org/10.1017/S0025100320000092>
- [17] 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>
- [18] Odlin, T. (1989). *Language transfer*. Cambridge University Press.
- [19] Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45, 73–97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Copyrights

Copyright for this article is retained by the author (s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).