# Prosodic Alignment in Conversations Among High-Proficiency Vietnamese Learners of Chinese

**Authors:** Wenbo Yu, Minjie Gong, Ziling Zhang, Jingjing Zhang, Shuang Xu, Shuang Xu

**Date:** 2026-02-15T09:43:40+00:00

## Abstract

Previous studies have investigated the prosodic alignment in different groups, such as adults, typically developing adolescents and children with autism. However, few studies have examined how second language learners align with one another during conversation. The current study addressed this gap by comparing the prosodic alignment performance between Mandarin native speakers and high-proficiency Vietnamese L2 learners. The present study identifies typical patterns of prosodic alignment among Mandarin native speakers and reveals that Vietnamese L2 learners struggle to achieve conversational alignment in F0. We argue that this difficulty is attributable to the cognitive pressure experienced by L2 speakers when conversing in the target language. This study reveals the prosodic alignment patterns of Vietnamese Chinese learners in conversation and highlights the significant role of pitch-related acoustic features.

## Full Text

## Preamble

Short title: Prosodic alignment in Chinese learners Full title: Prosodic Alignment in Conversations Among High-Proficiency Vietnamese Learners of Chinese* Authors:

Yu Wenbo1.3, Gong Minjie1,3, Zhang Ziling1,3, Zhang Jingjing2, Xu Shuang4 Institutes:

### 3. The Joint Interdisciplinary Laboratory of Language and Cognitive Science,

Nanjing Normal University

## 4. City University of Hong Kong

Address for correspondence: Xu Shuang City University of Hong Kong Tat Chee Avenue, Kowloon Hong Kong SAR Email: shuangxu@cityu.edu.hk Prosodic Alignment in Conversations Among High-Proficiency Vietnamese Learners of Chinese Abstract: Previous studies have investigated the prosodic alignment in different groups, such as adults, typically developing adolescents and children with autism. However, few studies have examined how second language learners align with one another during conversation. The current study addressed this gap by comparing the prosodic alignment performance between Mandarin native speakers and high-proficiency Vietnamese L2 learners. The present study identifies typical patterns of prosodic alignment among Mandarin native speakers and reveals that Vietnamese L2 learners struggle to achieve conversational alignment in F0. We argue that this difficulty is attributable to the cognitive pressure experienced by L2 speakers when conversing in the target language. This study reveals the prosodic alignment patterns of Vietnamese Chinese learners in conversation and highlights the significant role of pitch-related acoustic features.

Keywords: Chinese language learners, conversation, prosodic alignment

## 1. Introduction

Conversation, as the primary context for language use, offers critical insights into both the social foundations of linguistic interaction (Manson et al., 2013; Templeton et al., 2022). Speech alignment, an unconscious tendency of interlocutors to converge on lexical, syntactic, and prosodic features during dialogue (Pickering & Garrod, 2004), has been demonstrated as a fundamental mechanism facilitating communicative efficiency. This dynamic process reflects interlocutors' behavioral coordination, whereby they synchronize their verbal behavior rhythmically to achieve mutual attunement during conversation. While research on speech alignment has been extended to diverse populations, including typically developing adults, children, and individuals with various language disorders (Cai et al., 2021; Ostrand & Chodroff, 2021; Stabile & Eigsti, 2022; Chieng et al., 2024), research on second language acquisition learners has disproportionately focused on lexical alignment, defined as interlocutors' mutual adoption of identical or synonymous lexical items, and syntactic alignment, characterized by convergence toward shared grammatical structures (Kim et al., 2020; Suffill et al., 2021; Zhang & Nicol, 2022; Kim & Michel, 2023; Shen & Wang, 2025). This emphasis has left a significant gap in our understanding of prosodic alignment, particularly regarding how L2 learners exhibit prosodic alignment, and how target-language proficiency modulates such alignment patterns. This study investigates prosodic alignment in conversational interactions among advanced Vietnamese learners of Chinese in mainland China. Focusing on this high-proficiency L2 learner group, we systematically analyze their speech alignment patterns with native Chinese speakers, contributing insights to both second language acquisition in Chinese dialogic contexts and pedagogical strate-

gies for Chinese as a second language.

Prosodic alignment has been empirically verified across multiple languages, including English, Mandarin, and Danish (e.g., Xia et al., 2014; Trujillo et al., 2023). The systematic investigation of prosodic alignment was formally established through Levitan and Hirschberg's (2011) landmark study, which proposed a tripartite typology to characterize acoustic coordination patterns in dialogue: proximity, convergence, and synchrony. Wynn and Borrie (2022) reconceptualized prosodic alignment through three analytical dimensions: (1) class of alignment, (2) temporal level of alignment, and (3) dynamicity of alignment. In this revised framework, the authors incorporated convergence under the broader category of proximity while refining two key dimensions, namely the temporal level of alignment and dynamicity of alignment, thereby establishing a more comprehensive model. The present study builds upon this theoretical framework.

Empirical investigations have systematically documented prosodic alignment in native speaker conversations through multiple acoustic features, including intensity, fundamental frequency (F0), speech rate, and voice quality. For instance, Levitan and Hirschberg (2011) analyzed speech data from the Columbia Games Corpus, revealing that speakers' mean and maximum intensity values exhibited proximity alignment at the global level, while these same acoustic measures demonstrated synchrony alignment in local-level coordination. Similarly, an investigation of conversational data from Mandarin native speakers revealed a critical distinction: only amplitude features demonstrated bidirectional alignment, manifesting simultaneously at both global and local levels (Xia et al., 2023).

Regarding speech rate, Weise et al. (2019) demonstrated significant alignment effects for speech rate in both proximity and synchrony dimensions, establishing it as a primary index of prosodic alignment. Complementing this work, Bonin et al. (2013) provided evidence for a hierarchy of prosodic alignment strength, indicating that F0 alignment occurs with greater frequency than temporal alignment in spontaneous dialogue.

Two theoretical frameworks have been proposed to account for the mechanisms underlying speech alignment. The Interactive alignment model (IAM, Pickering & Garrod, 2004) posits that alignment operates as an automatic, implicit process whereby interlocutors spontaneously converge on lexical, syntactic, and phonological representations during dialogue, without conscious monitoring or intentional control. In contrast, Communication accommodation theory (CAT, Giles et al., 2012; 2023) emphasizes the role of speaker agency and cognitive effort, conceptualizing alignment as a strategic manifestation of social motivation. Within this framework, lower-status speakers actively accommodate to higher-status interlocutors to negotiate social approval or minimize interpersonal distance. In the context of second language acquisition, CAT has garnered relatively stronger empirical support in explaining alignment effects. For instance, interlocutors tend to exert a robust influence on phonetic accommodation: when L2 learners believed they were interacting with a native speaker

(as opposed to a non-native interlocutor), they exhibited greater convergence toward their partner' s vowel production patterns (Jiang & Kennison, 2022). Moreover, Kim et al. (2011) found that the alignment effect is lower in dialogue with two interlocutors from different native languages or dialects compared with two interlocutors from the same native language. This attenuated alignment effect between L2 interlocutors may be attributed to the allocation of cognitive resources during conversation. When interlocutors engage in dialogue using a second language, substantial cognitive resources are diverted to the phonological characteristics, including the perception of non-native speech and the production of target language (Costa et al., 2008; Kim et a., 2011), thereby interfering with the automatic processes underlying prosodic alignment. These findings suggest that L2 alignment does not operate as an automatic process akin to IAM but is instead constrained by cognitive resources, a pattern consistent with the theoretical claims of CAT. Some studies further provided similar evidence. For example, L2 interlocutors demonstrate superior performance in lexical alignment, which imposes minimal demands on cognitive resources, yet exhibit weak syntactic alignment where greater cognitive effort is required (Wang et al., 2025). Accordingly, it is reasonable to assume that for any given pair of L2 interlocutors, alignment will be more robust on dimensions imposing minimal cognitive demands, but attenuated on dimensions requiring greater processing resources.

Regarding the experiment design, research on linguistic alignment in L2 learners is rarely conducted in authentic conversational settings. For instance, Shen and Wang (2025) examined lexical alignment using text-based dialogue tasks, where English L2 learners were required to match the favored names (e.g., computer) or disfavored names (laptop) to the pictures given the pre-design experiment program. In addition, Hong et al. (2023) employed a two-stage paradigm to investigate phonetic alignment. The design first captured participants' baseline productions of target words in isolation, then recorded their subsequent realizations of these lexemes during controlled conversational interactions. Next, they quantified prosodic alignment effects by comparing the acoustic features of target words in isolated recordings versus conversational contexts. However, this study fails to examine the alignment relationship between interlocutors, thus deviating from the original purpose of speech alignment. Recently, a new paradiagm, alternating reading task (ART), was employed to systematically investigate the phonetic alignment in a controlled and less-spontaneous setting in researches of L2-L2 conversations (Yuan et al., 2024a, 2024b). While this task indeed demonstrates certain advantages in terms of the naturalness of language production, participants are able to devote greater attention to prosodic production due to the pre-provided contexts. Consequently, ART is susceptible to confounding the experimental results of prosodic entrainment. In sum, research on second language prosodic alignment in natural conversational settings has not yet been reported; it is imperative to explore how L2 learners align with their interlocutors in ecologically valid conversational settings.

Building upon these analyses, the present study investigates the manifestation of prosodic alignment in the conversation of Chinese learners, with specific at-

tention to whether three acoustic features (i.e., intensity, F0, and speech rate) exhibit differential alignment patterns. This inquiry is motivated by the phonological characteristics of Mandarin Chinese. For L2 learners of Chinese, lexical tones constitute one of the most formidable aspects of phonological acquisition (Zhou & Veríssimo, 2025; Huang et al., 2026). Given that F0 not only serves a distinctive lexical function but also plays a role in intonation realization in Mandarin, L2 speakers may be compelled to allocate substantial cognitive resources to the accurate production of tonal contrasts, thereby compromising their capacity to align the utterance F0 with their interlocutors. By contrast, intensity and speech rate, being less burdened by such demands, may remain available for alignment. In addition to examining this hypothesis, we further investigate the relationship between L2 proficiency and prosodic alignment. This is because although the association between language ability and speech alignment has been extensively documented (e.g., Ulbrich, 2024; Shen & Wang, 2025), this link remains underexplored in the phonetic domain. Moreover, given the hypothesized asymmetry of the alignment effect across acoustic features, the proficiency- alignment relationship may prove to be considerably more nuanced than previously assumed.

This study focuses on Vietnamese learners of Chinese. The primary reason for this choice is that the number of Vietnamese students studying in China has gradually increased in recent years, leading to increased demand for Chinese language instruction (https://news.gmw.cn/2024- 01/11/content_{37082346}.htm). This study exclusively recruited female participants, both because they constitute the majority of Vietnamese international students in China and because gender effects on prosodic alignment remain inconsistent in the literature (e.g., Sun & Ding, 2024).

Therefore, a female-only sample thus ensured demographic representativeness while minimizing potential confounds. In addition, the current study adopts Wynn and Borrie' s (2022) analytical framework to investigate prosodic alignment through two principal metrics: static global proximity and dynamic global proximity. The selection of these specific measures is methodologically warranted for two key reasons. First, these indices represent the most robust and widely implemented in contemporary alignment research. Second, preliminary data indicate that Vietnamese L2 learners' limited Mandarin proficiency results in substantially reduced verbal output per conversational turn, thereby precluding reliable local-level analysis due to insufficient data points. Accordingly, this investigation addresses two core research questions: (1) How do prosodic alignment patterns differ between Vietnamese L2 learners and native Mandarin speakers during conversational interaction? (2) How does Mandarin proficiency modulate prosodic alignment patterns in Vietnamese L2 learners?

## 2.1 Participants

The study involved 40 female participants, including 20 native Mandarin-speaking Chinese graduate students (mean age = 23.17 years, range = 21-25)

and 20 Vietnamese international students with HSK Level 6 or higher Mandarin proficiency (mean age = 24.30 years, range = 18- 36). All participants were recruited as pre-existing roommate dyads to ensure natural conversational dynamics while controlling for interlocutor familiarity effects. Rigorous screening confirmed all participants had normal hearing and speech abilities. The study protocol was approved by the institutional ethics review board (××$×$202506030). Notably, despite the relatively small sample size, the present study has the advantage of participant homogeneity in terms of first language background: all participants were native speakers of Vietnamese. This stands in contrast to other studies, such as Song et al. (2025), which involved learners from six different L1 backgrounds. In other words, our design allows for better control of potential L1 influence on Chinese language learning.

## 2.2 Materials and Procedure

The study employed a validated picture-sequencing task (Xia et al., 2023) to elicit naturalistic dialogue between dyads. During the conversation, the two interlocutors were assigned the roles of instruction giver and follower, respectively. Each pair completed the task twice with reversed roles.

Interlocutors were first separated by an opaque partition to prevent non-verbal communication. In each session, the giver could view 15 disyllabic Tone-1 objects in a fixed sequence and describe the order of each object. For the follower, the objects were randomly positioned, and they were required to fill in the serial number of each object in the corresponding position in the table according to the giver' s description. Before the formal task, participants first practiced with four excluded images. The materials and experimental setup are illustrated in Figure 1 [Figure 1: see original paper]. Audio was captured via M-Track DUO interface with directional microphones in Adobe Audition (48kHz/16- bit WAV), with speakers recorded on separate channels. The average trial completion time was 90 seconds for native dyads and 130 seconds for L2 pairs, consistent with expected proficiency difference. Vietnamese participants' Mandarin proficiency was verified using Feng et al.'s (2020) Quick Chinese Proficiency Test, a measure with established psychometric properties in L2 Chinese research.

Figure 1. Schematic diagram of experimental materials and setup

## 2.3 Speech Annotation

The audio files were processed using the Search for IPUs function in SPPAS (Speech Phonetics and Phonology Analysis System) to automatically segment Inter-Pausal Units (IPUs), with a minimum silence threshold of 80ms between units, generating corresponding TEXTGRID files for further analysis. Linguistics undergraduates manually verified the segmentation results in Praat before transcribing the speech content, with any discrepancies resolved through consultation between the annotators, the author, and linguistics faculty. The extracted acoustic measures included speech rate (syllables per second), F0 (min,

mean, and max), and Int (min, mean, and max), all of which were consistent with Xia et al' s study in 2023, ensuring methodological comparability across studies.

## 2.4 Analytical Approach

In the analysis of prosodic alignment, an independent sample t-test was employed to compare two conditions: (1) conversational partner pairs (within-dyad) and (2) non-conversational partner pairs (between-dyad). If A1 and A2 formed one conversational dyad, while B1/B2 and C1/C2 represented two additional dyads, the acoustic distance (e.g., speech rate difference) between A1 and A2 was treated as the within-dyad condition, whereas the distances between A1 and both B2 and C2 were aggregated as the between-dyad condition; thus alignment effect for speech rate was further quantified using the following formula: = ( 1 − 2) − 1/2(( 1 − 2) + ( 1 − 2)) A significant difference between within-dyad and between-dyad in the independent samples t-test was interpreted as evidence of robust prosodic alignment in dialogue. The present study initially investigated the alignment effect within the native and L2 group respectively and subsequently conducted a comparative analysis to discern the disparities in the alignment effect between the two groups. For static global proximity analysis, acoustic measures were duration- weighted by IPU, followed by the calculation of acoustic distance measures between conversational partners and non-partners. Dynamic global proximity analysis involved partitioning the interaction into 15 sections corresponding to the 15 pictures, with section-specific weighting and speaker-wise standardization of acoustic measures before conducting t-tests for each section.

The relationship between L2 proficiency and prosodic alignment was assessed through correlation analyses. Due to the small sample size in the current study, permutation testing framework (Pitman, 1937; Unseld et al., 2025) was employed for all inferential statistics in the current study, providing distribution-free inference particularly advantageous for small-sample scenarios.

Additionally, false discovery rate (FDR) correction was conducted to mitigate Type I error inflation. All analyses were implemented in R (V4.5.2) with complete data and code available on OSF website (https://osf.io/53zxf/?view_{only}=1511999d5c894487aeb687da8a989149).

## 3.1 Analysis of static global proximity

Permutation-based independent samples t-tests indicated that native Mandarin speakers exhibited an alignment effect across all acoustic features except speech rate (as reflected by negative alignment values in Table 1 ), with only the mean F0 demonstrating marginally significant negative alignment. In contrast, Vietnamese learners of Chinese exhibited a distinct prosodic alignment pattern: no significant convergent adaptation across all acoustic features, except a marginal significance on intensity.

Table 1. Alignment values and p-values between conversational and non-conversational groups Acoustic features Speech rate F0 min F0 mean Native Speakers L2 Learners Alignment p value Alignment p value Inter-group Difference p value F0 max Int min Int mean Int max An inter-group comparison of prosodic alignment patterns revealed significant differences, particularly in F0-related features. Permutation tests showed non-significant alignment effects between native speakers and Vietnamese learners on speech rate and three indexes of intensity. In contrast, significant inter-group differences emerged in F0 measures, suggesting that native and L2 speakers' prosodic alignment diverges primarily along pitch dimensions rather than temporal or intensity features (see Table 1 and Figure 2 [Figure 2: see original paper]).

Figure 2. Alignment effect across seven acoustic features between Chinese and Vietnamese

## 3.2 Correlation between Chinese proficiency and prosodic alignment

In static global proximity analysis, Pearson correlations between L2 interlocutors' Mandarin proficiency and alignment effects were all non-significant after FDR correction (see Table 2 ).

Table 2. Correlation results of Chinese proficiency and prosodic alignment Acoustic features Speech rate F0 min F0 mean F0 max Int min Int mean Int max p value

## 3.3 Analysis of dynamic global proximity

A series of permutation-based independent samples t-tests were further conducted to examine dynamic global proximity patterns between native Mandarin speakers and Vietnamese L2 learners within each of 15 sections. Statistical analysis showed that none of the sections yielded significant differences across any acoustic alignment features. To address potential power limitations due to constrained verbal output per section, we subsequently aggregated data into larger units (3-section and 5-section combinations), yet the results remained non-significant across all comparative analyses.

# 4. Discussion

This study presents the first investigation of prosodic alignment in high-proficiency Vietnamese learners of Mandarin during conversational interaction. The results show that both groups exhibited a certain degree of prosodic alignment. However, Vietnamese learners demonstrated significant differences from native speakers in pitch-related acoustic measures, and this difference was confined to the static dimension. Furthermore, L2 learners' prosodic alignment showed no association with their Mandarin proficiency.

## 4.1 Prosodic Alignment in Native Mandarin Speakers and Vietnamese Learners of Mandarin.

The interactive alignment theory (Pickering & Garrod, 2004) posits automatic cross-level alignment in conversation, whereby alignment at one linguistic level induces alignment at others.

However, emerging evidence suggests that alignment manifestations are asynchronous across phonological, syntactic, and lexical levels and are subject to various influencing factors. Among the most direct of these are constraints arising from language-specific characteristics. For instance, Dideriksen et al. (2023) reported that linguistic entrainment was overall higher in Danish than in Norwegian, attributing this to the greater opacity of the Danish speech signal. Additionally, Hungarian interlocutors exhibited a stronger alignment effect than those speaking English, Spanish, or Slovak. This may be attributed to Hungarian' s agglutinative nature, which facilitates efficient lemmatization and consequently greater stem re-utilization between conversational partners (Kejriwal & Beňuš, 2025). In the current study, prosodic alignment of native Mandarin speakers observed was primarily manifested in F0-related measures, which is in line with observations by Levitan et al. (2018), who also noted that Chinese has the strongest and most prevalent negative pitch alignment among four languages examined. In addition, our findings further revealed non-significant speech rate alignment, consistent with the results reported by Xia et al. (2023). As a syllable-timed language, Mandarin syllables are bound together by tonal constraints, resulting in relatively uniform syllable durations across different words or characters (with the exception of shorter fourth-tone syllables), limiting speech rate variability and thus reducing the likelihood of significant prosodic alignment. Therefore, both the pronounced F0 alignment and the absence of significant speech rate alignment are intimately linked to the phonological properties of Mandarin.

Having established a pattern of prosodic alignment among native speakers, this study is eager to investigate which acoustic features distinguish prosodic alignment between Vietnamese learners of Chinese and native speakers. Our analysis of three F0 metrics revealed significant or marginally significant between-group divergence. Given native speakers' pronounced pitch-based alignment, these results suggest that even advanced Chinese learners fail to fully acquire the pitch-driven alignment patterns characteristic of native speech.

Unexpectedly, Vietnamese learners demonstrated a positive F0 alignment effect (Table 1), indicating greater acoustic variability within conversational pairs compared to non-conversational pairs. CAT posits that speakers dynamically adjust their behavior based on interpersonal convergence motives. A negative alignment effect reflects mutual adaptation to sustain dialogue, facilitating shared representations and yielding fewer turn transitions with longer speech units. In contrast, a positive alignment effect implies frequent dyadic misalignment, necessitating iterative repair through turn-taking (Sun & Ding, 2024). In other

words, positive alignment effects in our study entail that interlocutors engage in frequent turn-taking to maintain conversation and accomplish the task. Supporting this, we indeed found that native speakers produced an average of

**28.73 IPUs, with each IPU containing a mean of 4.96 syllables. The former number was**

significantly lower than the 43.80 IPUs produced by Vietnamese participants, while the latter was substantially higher than the 3.02 syllables per IPU observed in Vietnamese participants. Thus, the positive alignment effects, combined with more yet shorter IPUs, suggest that L2 speakers experience heightened cognitive load during conversation, resulting in fragmented output, a phenomenon consistent with stress-induced prosodic decoupling (Abel & Babel, 2017).

Recent research has proposed that although interlocutors actively engage in linguistic alignment during conversation, the magnitude of alignment effects is also constrained by cognitive resources. Alignment occurs only when interlocutors possess sufficient cognitive capacity to implement it (Cai et al., 2021; Shen & Wang, 2025). Therefore, the reverse alignment pattern observed in our study may reflect a trade-off in cognitive resource allocation during speech processing. Our study employed a goal-directed conversational task in which interlocutors are required to collaborate to sequence pictures, rendering accurate production of each picture's name and order critical. Thus, to complete the task, L2 learners must ensure accurate utterance production, which consumes substantial cognitive resources. Although both Mandarin and Vietnamese are tonal languages, they differ substantially in terms of tonal inventory size, pitch register, and contour typology (Tu, 2023). Moreover, the phonetic differences between the native language and the target language inevitably exert a substantial negative influence on L2 communication, as demonstrated in Olmstead et al.' s (2021) study. Specifically, although Chinese English L2 learners began to utilize the spectral dimension, which is typically employed by native English interlocutors, to distinguish vowel pairs, they persisted in producing distinct vowel length durations for /I/ and /i /, a strategy characteristic of Mandarin. Consequently, Vietnamese L2 learners in conversational contexts must first inhibit negative L1 transfer and construct phonological categories in Mandarin at the mental lexicon level (e.g., the four Mandarin tones), before mapping these abstract representations onto naturalistic speech signals that can be dynamically coordinated with interlocutors in real-time interaction. In other words, L2 interlocutors must allocate greater cognitive resources to lexical tone at the expense of intonation, leaving insufficient capacity to adjust prosodic features for alignment with their conversational partners.

To our knowledge, this study represents the first investigation of dynamic prosodic alignment in L2 learners. Moreover, dynamic prosodic alignment remains exceedingly rare in other populations as well (except Sun & Ding, 2024). The results of the present study indicate that the prosodic alignment effects between L2 speakers and native speakers did not emerge gradually as

the conversation progressed; rather, they remained relatively stable throughout. This pattern may be attributable to the fact that the participants were friends, and that the conversational task was goal-oriented rather than informal, thereby precluding marked changes in the adaptation process between interlocutors. This finding further indicates the importance of both task type (Wynn et al., 2024) and interlocutor familiarity (Wang et al., 2025) in shaping speech alignment

## 4.2 Relationship Between L2 Proficiency and Prosodic Alignment in Vietnamese Learners

After correction, there were no significant effects between L2 learners' proficiency in Chinese with any acoustic features of prosodic alignment. This result contrasts with prior research on lexical and phonetic alignment (Zhang & Nicol, 2022; Christiane Ulbrich, 2024; Shen & Wang, 2025). We attribute this discrepancy to methodological differences. For example, Chinese L2 learners of English in Shen and Wang' s (2025) study showed increased lexical alignment by producing more disfavored names after their less proficient L2 interlocutor produced such names.

Notably, their picture-naming task was without time limit and thus induced a relatively weak processing load compared with picture-sequencing task relying on real-timing interaction in the current study. As a result, even advanced Vietnamese learners in our study, when engaged in such a cognitively demanding task, demonstrated insufficient capacity to produce utterances accurately while simultaneously aligning with their partner. In another speech alignment study, a significant alignment effect on speech rate was observed between high-proficiency learners with German as a native language and low-proficiency learners (Christiane Ulbrich, 2024). Although their adopted task is similar to ours, the authors restricted their analyses to the group level and did not include correlation analyses comparable to those employed in this study, thereby limiting the possibility of more fine-grained interpretations.

## 4.3 Future Directions and Second Language Pedagogy

Two methodological limitations of the present study should be acknowledged and addressed in future studies. First, this study primarily adopted the methodological framework of Xia et al. (2023), focusing on three prosodic dimensions (intensity, duration, and pitch) while excluding other acoustic features such as formant patterns and harmonic-to-noise ratio. Given the inherent correlations among acoustic measures, future research could employ principal component analysis (PCA) with a more comprehensive set of vocal features to better capture speech alignment patterns (e.g., Patel et al., 2022). Additionally, since the current sample comprised exclusively female participants (reflecting the gender distribution of Vietnamese students in China) and given the established influence of gender on prosodic alignment, subsequent studies should include more

diverse participant populations to provide a fuller understanding of L2 speakers' prosodic and phonetic alignment phenomena.

This study further offers practical insights for second language teaching. Current second- language instruction primarily focuses on classroom instruction and learners' mastery of specific linguistic structures (e.g., phonological features, syntactic patterns). However, for learners immersed in an L2 environment, there are abundant opportunities for natural communication with both native and non-native speakers (Fernández Dobao, 2012). Unlike formal classroom settings, these daily interactions represent the most frequent context for language acquisition. Given that prosodic alignment significantly enhances conversational success (Lamekina & Meyer, 2023; Lamekina et al., 2024), examining Chinese learners' prosodic alignment patterns can promote language acquisition. Our findings reveal significant differences in pitch-related alignment between native and L2 speakers, suggesting that future Chinese pedagogy should expand beyond character-level tone instruction to incorporate discourse-level intonation training.

## 5. Conclusion

This investigation examined prosodic alignment patterns among Vietnamese learners of Chinese during conversational interactions. The results revealed significant divergences between L2 learners and native speakers, particularly in pitch-related features, underscoring the crucial role of phonological features in Chinese conversational dynamics. These findings advance our understanding of the complex interplay among linguistic proficiency, prosodic alignment, and communicative effectiveness in L2 contexts, and further underscore the need for more refined models of alignment mechanisms, both theoretically and in L2 acquisition practice.

Reference Abel, J., & Babel, M. (2017). Cognitive load reduces perceived linguistic convergence between dyads. Language and Speech, 60(3), 479-502.

Bonin, F., De Looze, C., Ghosh, S., Gilmartin, E., Vogel, C., Campbell, N., ···Vinciarelli, A. (2013). Investigating fine temporal dynamics of prosodic and lexical accommodation. In Proceedings of the Annual Conference of the International Speech Communication Association (pp. 539–543). INTERSPEECH.

Cai, Z. G., Sun, Z., & Zhao, N. (2021). Interlocutor modelling in lexical alignment: The role of linguistic competence. Journal of Memory and Language, 121, 104278.

Chieng, A. C. J., Wynn, C. J., Wong, T. P., Barrett, T. S., & Borrie, S. A. (2024). Lexical Alignment is Pervasive Across Contexts in Non-WEIRD Adult–Child Interactions. Cognitive science, 48(3), e13417.

Costa, A., Pickering, M. J., & Sorace, A. (2008). Alignment in second language dialogue. Language and cognitive processes, 23(4), 528-556.

Dideriksen, C., Christiansen, M. H., Dingemanse, M., Højmark-Bertelsen, M., Johansson, C., Tylén, K., & Fusaroli, R. (2023). Language-specific constraints on conversation: Evidence from Danish and Norwegian. Cognitive Science, 47(11), e13387.

Feng, L., Feng, H., Bai, S., & Wu, J. (2020). Development and analysis of a rapid Chinese L2 proficiency test: A study based on fixed-ratio cloze tests. Applied Linguistics, (03), 69-79.

Fernández Dobao, A. (2012). Collaborative dialogue in learner–learner and learner–native speaker interaction. Applied linguistics, 33(3), 229-256.

Giles, H., Edwards, A. L., & Walther, J. B. (2023). Communication accommodation theory: Past accomplishments, current trends, and future prospects. Language Sciences, 99, 101571.

Giles, H., Mulac, A., Bradac, J. J., & Johnson, P. (2012). Speech accommodation theory: The first decade and beyond. In Communication yearbook 10 (pp. 13-48). Routledge.

Hong, Y., Chen, S., Zhou, F., Chan, A., & Tang, T. (2023). Phonetic entrainment in L2 human- robot interaction: an investigation of children with and without autism spectrum disorder. Frontiers in psychology, 14, 1128976.

Huang, Y., Xu, Z., Bei, X., & Huang, H. (2026). Perception–Production of Second-Language Mandarin Tones Based Interpretable Computational Methods:

Review. Mathematics, 14(1), 145. https://doi.org/10.3390/math14010145 Jiang, F., & Kennison, S. (2022). The impact of L2 English learners' belief about an interlocutor' s English proficiency on L2 phonetic accommodation. Journal of Psycholinguistic Research, 51(1), 217-234.

Kejriwal, J., & Beňuš, Š. (2025). Lexical, syntactic, semantic and acoustic entrainment in Slovak, Spanish, English, and Hungarian: A cross-linguistic comparison. Speech Communication, 103240/ Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. Laboratory Phonology, 2(1), Kim, Y., & Michel, M. (2023). Linguistic alignment in second language acquisition: A methodological review. System, 115, 103007.

Kim, Y., Skalicky, S., & Jung, Y. (2020). The role of linguistic alignment on question development in face-to-face and synchronous computer-mediated communication contexts: A conceptual replication study. Language Learning, 70(3), 643-684.

Lamekina, Y., & Meyer, L. (2023). Entrainment to speech prosody influences subsequent sentence comprehension. Language, Cognition and Neuroscience, 38(3), 263-276.

Lamekina, Y., Titone, L., Maess, B., & Meyer, L. (2024). Speech Prosody

Serves Temporal Prediction of Language via Contextual Entrainment. Journal of Neuroscience, 44(28).

Levitan, R., & Hirschberg, J. B. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In Proceedings of INTERSPEECH 2011.

Levitan, S. I., Xiang, J., & Hirschberg, J. (2018). Acoustic-prosodic and lexical entrainment in deceptive dialogue. In Proc. 9th International Conference on Speech Prosody.

Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. Evolution and Human Behavior, 34(6), 419-426.

Olmstead, A. J., Viswanathan, N., Cowan, T., & Yang, K. (2021). Phonetic adaptation in interlocutors with mismatched language backgrounds: A case for a phonetic synergy account. Journal of Phonetics, 87, 101054.

Ostrand, R., & Chodroff, E. (2021). It' s alignment all the way down, but not all the way up:

Speakers align on some features but not others within a dialogue. Journal of phonetics, 88, Patel, S. P., J. Cole, J. C. Y. Lau, G. (2022). Fragnito & M. Losh. Verbal entrainment in autism spectrum disorder and first-degree relatives. Scientific Reports, 12(1): 11496.

Pickering, M. J. & S. Garrod. (2004). Toward a mechanistic psychology of dialogue. Behavioral and brain sciences, 27(2):169-190.

Pitman, E. J. (1937). Significance tests which may be applied to samples from any populations. Supplement to the Journal of the Royal Statistical Society, 4(1), 119-130.

Shen, H., & Wang, M. (2025). Effects of interlocutors' linguistic competence on L2 speakers' lexical alignment. Bilingualism: Language and Cognition, 1-13.

Song B, Jiang X, Maurer U, & Li S. (2025). Neural tuning for Chinese characters in adult Chinese L2 learners: evidence from an ERP study. Bilingualism: Language and Cognition. 28(2):286- 299. doi:10.1017/S1366728924000403 Stabile, M., & Eigsti, I. M. (2022). Lexical alignment and communicative success in autism spectrum disorder. Journal of speech, language, and hearing research, 65(11), 4300-4305.

Suffill, E., Kutasi, T., Pickering, M. J., & Branigan, H. P. (2021). Lexical alignment is affected by addressee but not speaker nativeness. Bilingualism: Language and cognition, 24(4), 746-757.

Sun, Y., & Ding, H. (2024). Unpacking the gender-role interaction of prosodic entrainment in Chinese long-and-short turn-taking: evidence perceptual acoustic similarities. Humanities and Social Sciences Communications, 11(1), 1-15.

Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2022).

Fast response times signal social connection in conversation. Proceedings of the National Academy of Sciences, 119(4), e2116915119.

Trujillo, J. P., Dideriksen, C., Tylén, K., Christiansen, M. H., & Fusaroli, R. (2023). The dynamic interplay of kinetic linguistic coordination in Danish and Norwegian conversation. Cognitive Science, 47(6), e13298.

Tu, J. Y. (2023). Production of Mandarin disyllabic tones by Vietnamese speakers. In Proceedings of the 20th International Congress of Phonetic Sciences-ICPhS (pp. 2388-2392).

Ulbrich, C. (2024). Phonetic accommodation on the segmental and the suprasegmental level of speech in native–non-native collaborative tasks. Language and Speech, 67(2), 346-372.

Unseld, T., Ruckerbauer, L., & Mayer, B. (2025). Permutation tests are a useful alternative approach for statistical hypothesis testing in small sample sizes. Alternatives to Laboratory Animals, 53(3), 130-137.

Wang, F., Wang, M., & Boland, J. (2025). Effects of language status and interlocutor familiarity on linguistic alignment in L1 and L2 dialogues. TESOL Quarterly, 59(4), 2206-2233.

Weise, A., Levitan, S. I., Hirschberg, J., & Levitan, R. (2019). Individual differences in acoustic- prosodic entrainment in spoken dialogue. Speech Communication, 115, 78-87.

Wynn, C. J., & Borrie, S. A. (2022). Classifying conversational entrainment of speech behavior:

An expanded framework and review. Journal of Phonetics, 94, 101173.

Wynn, C. J., Barrett, T. S., & Borrie, S. A. (2024). Conversational speech behaviors are context dependent. Journal of Speech, Language, and Hearing Research, 67(5), 1360-1369.

Xia, Z., Hirschberg, J., & Levitan, R. (2023). Investigating prosodic entrainment from global conversations local turns tones in Mandarin conversations. Speech Communication, 153, 102961.

Xia, Z.H., Levitan, R., Hirschberg, J. (2014). Prosodic entrainment in Mandarin and English: a cross-linguistic. comparison. In: Proceedings of Speech Prosody, Dublin, Ireland, pp. 65–69.

Yuan, Z., Beňuš, Š., & D' Ausilio, A. (2024a). Language proficiency and f0 entrainment: a study of l2 english imitation in italian, french, and slovak speakers. Speech Prosody 2024, 1265- Yuan, Z., de Jong, D., Beňuš, Š., Nguyen, N., Feng, R., Sabo, R., Fadiga, L., & D' Ausilio, A. (2024b). ART: The Alternating Reading Task Corpus for Speech Entrainment and Imitation.

In Proceedings of LREC-COLING 2024 (pp. 1548–1562). ELRA Language Resource Association.

Zhang, D., & Nicol, J. (2022). Lexical alignment in second language communication: evidence from a picture-naming task. Language, Cognition and Neuroscience, 37(6), 732-749.

Zhou, C., & Veríssimo, J. (2025). L2 difficulties in the perception of Mandarin tones:

Phonological universals or domain-general aptitude?. Bilingualism: Language and Cognition,

*Note: Figure translations are in progress. See original paper for figures.*