LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Roles of Domain-General Auditory Processing in Second Language Speech Learning Revisited: What Degree of Precision Makes a Difference?

Kazuya Saito[a,b] and Adam Tierney[c]

[a]University College London [b]Tohoku University [c]Birkbeck, University of London

Abstract: This study expands on the practical application of the critical role of auditory processing in the rate of naturalistic L2 speech acquisition. In Study 1, the prosodic production of English by 46 Chinese college students was tracked over a five-month study abroad program in the UK. Learners with extensive L2 input opportunities demonstrated improvements in prosodic accuracy; however, those with pitch acuity below a certain threshold showed regression, potentially reinforcing L1 interference. To determine what percentage of participants fell below the auditory processing threshold

CRediT author statement – **Kazuya Saito**: conceptualization (lead); data curation (lead); formal analysis (lead); funding acquisition (equal); investigation (lead); methodology (lead); project administration (lead); resources (lead); visualization (lead); writing – original draft preparation (lead); writing – review and editing (lead). **Adam Tierney**: conceptualization (support); data curation (support); formal analysis (support); funding acquisition (equal); investigation (support); methodology (support); project administration (support); resources (support); visualization (support); writing – original draft preparation (support); writing – review and editing (support).

A one-page Accessible Summary of this article in nontechnical language is freely available in the Supporting Information online and at https://oasis-database.org

We gratefully acknowledge Hui Sun for data collection and management, as well as Siying Chen, Jiali Guo, Changli Li, Ruiyang Li, Xian Li, Xinyue Luo, Yaoyao Ruan, and Yujie Shao for collecting and sharing their data sets. This project was supported by the Leverhulme Trust Research Fund (RPG-2024-391) and the UK-ISPF Grant (1185702223) awarded to KS, and the ESRC Standard Research Grant awarded to AT (ES/V007955/1).

Correspondence concerning this article should be addressed to Kazuya Saito, University College London, 20 Bedford Way, London WC1H 0AL. Email: k.saito@ucl.ac.uk

The handling editor for this article was Shelley Staples.

DOI: 10.1111/lang.12722

determined in Study 1, Study 2 administered pitch processing tests to 400 Chinese college students learning English, all with normal hearing, and developed a provisional corpus to assess pitch acuity variation within this cohort. The comparison of findings from Studies 1 and 2 suggests that insufficient auditory precision hampers naturalistic L2 learning. Approximately the bottom 1.5 quartiles of the population (35%) may fall below this threshold. These learners could benefit from remedial strategies (e.g., explicit phonetic instruction, auditory training) to fully capitalize on their naturalistic L2 learning opportunities.

Keywords Second language speech; individual differences; auditory processing; prosody; aptitude

Introduction

Over the past 30 years, scholars have examined the mechanisms underlying the rate and ultimate attainment of second language (L2) speech proficiency in relation to a range of learner-extrinsic variables. These include the timing of immersion in L2-speaking environments (age of arrival), the length of immersion (quantity of input), daily frequency of L2 use with fluent versus nonfluent talkers (quality of input), and educational background (e.g., length and onset of L2 education, the presence of pronunciation and musical training; for a comprehensive summary, see Flege & Bohn, 2021). While these variables show medium-sized predictive power for various L2 speech outcomes (e.g., $r = .30–.50$; Saito, 2015), they alone cannot explain all of the variance. This suggests that even if learners engage in the same amount and type of practice opportunities, there will be a great deal of individual variability, with some successfully developing advanced L2 speech proficiency and others struggling despite ample input and output.

Drawing on theories of the auditory foundations of first language (L1) acquisition (Goswami, 2015; Tallal & Gaab, 2006; Tierney & Kraus 2014), scholars have linked individual variation in L2 learning outcomes to individual differences in the ability to encode the acoustic characteristics of sounds. Under the perceptual account of language acquisition, auditory processing is considered a bottleneck for successful language learning; it serves as the initial ability that learners rely on for phonetic and phonological analysis when receiving aural input, impacting subsequent processes at lexical, morphosyntactic, and discoursal levels. In the context of L1 acquisition, extensive research suggests that children with lower auditory precision typically demonstrate slower language development (Kalashnikova et al., 2019) and a higher likelihood of language disorders such as dyslexia (Boets et al., 2011) and specific language impairment (McArthur & Bishop, 2005). According to

McWeeny and Norton's (2024) meta-analysis, the disparity in various types of auditory processing between normal and dyslexic groups may manifest as medium-to-large effects ($g = 0.70$–0.80).

Some scholars have posited that the impact of auditory processing might extend to L2 learning contexts (Mueller et al., 2012). More recent discussion has suggested that these effects could be even more clearly observed in L2 than in L1 acquisition due to fundamental differences in the quantity and quality of learning experiences (Saito et al., 2020). In L1 acquisition, individuals with auditory processing deficits may compensate through consistent and rich linguistic input provided by caregivers (e.g., Rosen, 2003). During such ample exposure, individuals with relatively low auditory precision may adopt remedial strategies to perceive speech by capitalizing on redundancies in speech processing. For example, individuals with amusia can demonstrate normal speech and music processing while relying on durational rather than pitch cues (Jasmin et al., 2020). Native listeners can also make use of semantic cues when phonetic cues are not clearly accessible (Bradlow & Alexander, 2007). In L2 learning, however, learners have access to a limited amount of input and output. Any deficit in auditory precision may therefore hinder their ability to fully utilize the already limited opportunities for input and output.

Additionally, auditory processing in L2 learning could be considered a dual task and thus more complex and demanding than in L1 learning. Unlike L1 acquisition, which takes place free of any prior language learning experiences, L2 acquisition occurs in a phonetic and phonological space where the L1 system is already established (fully or partially). When exposed to L2 sounds, learners need to encode their acoustic properties while suppressing L1 interference. Then, they need to compare them against L1 counterparts, deciding whether to assimilate or dissimilate these sounds (Flege & Bohn, 2021). Any disadvantages that learners have in terms of their reliance on auditory processing on domain-general levels could immediately impact their abilities to maintain or separate two phonetic and phonological systems effectively and efficiently.

Recent research has demonstrated both positive and negative effects of auditory processing in L2 speech learning. On the one hand, L2 learners who have achieved advanced L2 speech proficiency tend to possess relatively high (or at least normative) levels of auditory precision (Saito, Sun et al., 2022). The predictive power of auditory processing for successful L2 speech learning is medium to large even after accounting for experience-related and cognitive individual differences (e.g., $R^2 > .50$; Saito et al., 2024).

On the other hand, individuals with lower auditory precision tend to struggle to analyze new sounds due to L1 interference, which ultimately hinders successful L2 speech learning. Emerging evidence indicates that this auditory disadvantage can be observed under naturalistic, communicatively authentic, and consequently challenging conditions for L2 learners. These contexts include not only naturalistic immersion (e.g., study abroad) but also meaning-oriented instruction in which learners are engaged in processing language through real-life conversational tasks (e.g., task-based language teaching; Ruan & Saito, 2023; Xu et al., in press) and/or incidental and implicit training paradigms (e.g., Correia et al., in press) with natural input (e.g., intensive exposure to multiple speakers; Chandrasekaran et al., 2010; Perrachione et al., 2011; W. Zhang et al., 2024).

One critical message for practitioners (language teachers and learners) is that the auditory processing account of L2 speech learning can lay the groundwork for individualized approaches for L2 learners with diverse auditory processing profiles. To help learners develop and automatize their L2 knowledge and become functional L2 users, communicatively authentic L2 learning experiences and training are vital (Suzuki, 2023). However, while such an approach could benefit L2 learners with at least normative auditory precision, those with lower auditory precision may show limited progress (Perrachione et al., 2011). These auditorily disadvantaged learners may benefit from a combination of language and auditory processing training (Saito, Petrova et al., 2022) and/or detailed, explicit phonetic instruction (Suzukida & Saito, 2023) prior to engaging in meaning-oriented discourse. These individuals are not necessarily deficient in some way; rather, learners with different auditory profiles may require different forms of instruction (McMurray et al., 2023).

It is important to recognize that auditory processing functions as a continuous variable, and the aptitude–acquisition relationship identified in prior literature has remained highly abstract about the degree of auditory precision needed to promote L2 speech learning. Typically, research shows that learners with higher aptitude demonstrate more advanced L2 proficiency, whereas those with lower aptitude show limited L2 learning. However, these studies often avoid identifying a specific threshold for auditory precision, treating the impact of auditory acuity as a gradient.

In practice, however, decisions made by practitioners may be more dichotomous. For practical and pedagogical purposes, it is essential for practitioners to not only diagnose learners' auditory processing profiles but also interpret their significance. Identifying learners with relatively low auditory precision is crucial for providing tailored support, ensuring that all learners can maximize their input opportunities to promote optimized L2 speech learning, regardless of their perceptual differences.

One notable and exceptional attempt to provide categorical interpretations of auditory processing is the seminal work by Wong and Perrachione (Perrachione et al., 2011; Wong & Perrachione, 2007). Their research investigated how individual differences in pitch processing interact with training paradigms in the acquisition of a novel phonological contrast involving pitch. Participants first completed the team's in-house pitch discrimination tests and then a series of training sessions. The results showed that those who scored above 70% on the pitch tests were more likely to successfully acquire novel pitch-based phonological contrasts (Wong & Perrachione, 2007). Additionally, learners above the 70% threshold benefited from high-variability training, whereas those below the threshold performed better in low-variability training environments. These findings offer valuable insights for both theoretical development and practical application, suggesting that practitioners can assess learners' pitch processing abilities in advance and recommend the most appropriate phonetic training approach (high-variability training for those with high pitch acuity vs. low-variability training for those with low pitch acuity). Recently, our team has made a set of auditory processing assessment tools and user manuals publicly available to researchers and practitioners, enabling accurate evaluation of various dimensions of auditory processing (Saito & Tierney, 2024). Given this, we argue that it is both essential and timely to shed light on how adults with normal hearing typically perform on these tests; what distinguishes relatively low, normative, and high auditory precision; and how learners with varying levels of auditory precision acquire L2 speech over time.

The Present Study

To advance the concept of aptitude–treatment interaction in practical settings, it is essential to provide concrete guidance for both researchers and practitioners regarding typical adult performance on auditory processing tests. To contextualize this topic, the current project focuses on a specific instance of L2 speech learning: learning of English prosody by university-level Chinese learners during study abroad. The project consists of two studies: In Study 1, we examined how 46 Chinese college students with varying pitch acuity profiles naturally acquired English prosody over 5 months of study abroad. Statistical analysis revealed two distinct groups: those who showed continuous improvement and those who made limited progress. To better understand this threshold, we expanded the research: In Study 2, we collected data from a larger group of L2 learners (400 college-level Chinese learners of English) and used a battery of tests to assess pitch processing. By comparing the threshold identified in Study 1 with the corpus results from Study 2, we aim to predict:

- how auditory processing profiles vary among normal-hearing adults (bottom 25%, normative, top 25%),
- the extent to which learners can continue to improve their L2 abilities with practice (top 65%), and
- the likelihood that some learners may show limited progress despite receiving the same amount of practice (bottom 35%).

We distinguish our approach to aptitude–treatment interaction from the dominant, traditional, naturalistic psychology view, which assumes the existence of model language users and regards deviations from this norm as deficits. In L1 acquisition and hearing literature, there is ample precedent for diagnosing participants with standardized tests, particularly in assessing domain-general auditory impairments (see Neijenhuis et al., 2019, for a critical review). While binary distinctions (presence or absence of impairment) are often established based in part on descriptive analyses of corpus data (e.g., 1.5 standard deviations from the population mean; the American Speech-Language-Hearing Association's guidelines; https://www.asha.org/practice-portal/clinical-topics/articulation-and-phonology/#collapse_5), these thresholds do not necessarily reflect the impact of these variables on language learning behaviors.

McMurray et al. (2023) argued that individual differences should be viewed not in terms of deviations from the norm, but rather as key to understanding language function, offering more nuanced insights. In their review of the word recognition literature, they highlighted that populations often labeled as "deviations"—such as cochlear implant users, prelingually deaf children, and adolescents with developmental language disorder—use distinct strategies for word recognition, shaped by their specific sensory, developmental, or cognitive challenges (e.g., cochlear implant users likely maintain multiple possible candidates and interpretations of the word for longer). The study suggests that these differences should not be framed as deficits but rather as alternative solutions to the same challenge of real-time processing of language. By examining these variations, researchers can gain a deeper understanding of the core mechanisms involved in language processing across diverse populations.

Another example that challenges the binary distinction between aptitude and its complex underlying mechanisms is congenital amusia. Amusia, often defined by performance two standard deviations below the mean on a pitch-sequence memory test (Peretz et al., 2003), is characterized by difficulties in detecting small pitch changes across musical and nonmusical stimuli (Hyde & Peretz, 2004; Vuvan et al., 2015). Despite this, only 7% of English-speaking amusic people report problems with speech perception in everyday life (Liu et al., 2010). This highlights how listeners with perceptual difficulties can compensate by using other acoustic dimensions, such as relying on durational cues for English prosody perception (Jasmin et al., 2020).

This demonstrates that using predetermined binary thresholds, such as those used in naturalist frameworks, can overlook the complexity of language learning mechanisms. In contrast, our approach is reactive. We first observe how language learning unfolds in a real-life context (e.g., 5 months of prosody acquisition by 46 Chinese learners of English). After identifying differences in language development from a longitudinal perspective, we assess the auditory processing levels that correspond to limited progress ("relatively low auditory precision"). Finally, we examine corpus data from a larger group (400 learners) to provide descriptive interpretations of how these auditory processing thresholds account for population-wide variations.

Our primary motivation for providing binary analyses is to ensure the practical application of auditory processing research, especially in guiding interventions. We follow McMurray et al.'s (2023) notion of mechanistic functionalism, which posits that language and cognition emerge from multiple underlying processes that vary between individuals: These variations are not inherently better or worse, but simply different ways of solving the same problems. Echoing McMurray et al., we advocate for an improved understanding of auditory individual differences so that learners with relatively low auditory precision can receive targeted strategies (e.g., phonetic instruction, auditory training) to enhance their naturalistic L2 speech learning experiences, as the combination of naturalistic exposure with follow-up training may be particularly beneficial for these learners.

Study 1: Longitudinal Data

To explore the dimension-specific relationship between audition and acquisition, Study 1 focused on the role of pitch discrimination in Chinese learners' acquisition of English prosody—such as word stress and intonation—over the course of 5 months of study abroad in the UK. Our prediction was that pitch processing would significantly predict L2 prosodic production development. Here, we aimed to explore not only whether auditory processing relates to L2 speech learning but also what threshold of auditory precision is needed to promote L2 speech learning.

Prosodic proficiency is widely recognized as one of the most crucial aspects of developing comprehensible and intelligible L2 speech (Kang et al., 2010) and is considered particularly challenging for Chinese learners of English (Chrabaszcz et al., 2014; J. Zhang et al., 2018). Although Mandarin uses both pitch height and contour to mark every syllable, Chinese learners may struggle with crosslinguistic differences in English prosody, where pitch is used in a distinct manner to mark stressed syllables and convey meaning at the sentence level.

To test our hypothesis, we recruited *moderately* experienced Chinese learners of English (length of residence [LOR] > 5 months). These participants were presumed to have adjusted to their new L2 learning environment after several months in the UK. According to the extant literature, substantial learning typically occurs within the first 3 to 4 months of immersion in an English-speaking environment, followed by slower, more gradual learning trajectories, which sometimes plateau (e.g., Munro & Derwing, 2008, for vowels; Saito & Munro, 2014, for consonants; Trofimovich & Baker, 2006, for prosody).

Based on the recent longitudinal investigation of L2 speech learning trajectories beyond this initial phase (Munro, Derwing, & Saito, 2024), we expect three characteristics among the moderately experienced L2 participants:

1. Substantial, significant improvement is unlikely to be observed at the group level, as the participants had already benefited from an initial phase of relatively rapid improvement (Saito & Munro, 2014).
2. Considerable variation in their linguistic performance is likely, with some experiencing continuous development while others may show more static levels or even regression due to ongoing L1 interference (Munro et al., 2024).
3. Such varied L2 performance may be linked to a variety of individual difference variables, including experience (e.g., how much learners regularly use a L2; Flege & Bohn, 2021) and aptitude (e.g., Sun et al., 2021).

Focusing on L2 learners who have passed the initial phase of rapid improvement (LOR = 5–10 months), we examined the role of auditory processing in L2 speech development. Specifically, we anticipate that learners with relatively high pitch precision will continue to improve their L2 prosodic proficiency, particularly when they frequently use the L2, whereas those with relatively low pitch precision may show limited development or even regression. To establish the threshold for distinguishing high from low auditory precision, we applied a post hoc statistical approach (namely, Johnson–Neyman techniques; see the Results section below).

Participants

A total of 46 Chinese college students (3 males and 43 females) studying at universities in London participated in the study. Their ages ranged from 21 to

29 years (M_{age} = 23.60 years, SD = 2.27). None of the participants reported having any prior study-abroad experience. Initially, at the onset of the project, Time 1 (T1), their duration of stay in the UK was 5 months. Interested participants were first interviewed about their previous L2 learning experiences prior to studying abroad and their current L2 learning experiences. Participants self-reported a single percentage of L2 English usage for each context. The latter referred specifically to the percentage of L2 English used both inside and outside of school. In the present study, consistently with previous findings (Flege & Bohn, 2021), we focused on the extent to which participants used L2 English outside of school, as this measure reflects their voluntary efforts to engage with the target language and the potential for improvement in L2 speech proficiency.

Next, participants completed a range of auditory processing and L2 tests. Five months later, at Time 2 (T2), participants returned to take the same auditory and L2 tests and also reported on their L2 English usage during the project period. While parts of the data (comprehension data) were reported in our preliminary report (Sun et al., 2021), the current paper focuses on the production data.

Prior to the project, participants had reported extensive foreign language education ($M_{\text{length of learning}}$ = 13.40 years, SD = 2.02, range = 10–19). According to their self-reported IELTS scores, their general English proficiency could be classified as intermediate-to-advanced ($x > 6.5$). A total of 27 out 46 participants reported musical training (defined and operationalized as more than 6 years of musical training; J. D. Zhang et al., 2020). To survey participants' current L2 learning experience, individual interviews were conducted based on the Language Contact Profile (Freed et al., 2004). As summarized in Table 3 and the Results section, their L1 and L2 use widely varied at both T1 and T2.

Auditory Processing Measures

In the current study, two types of auditory processing—pitch and formant acuity—were assessed to examine the dimension-specific relationship between auditory processing and L2 speech learning. We hypothesized that participants' L2 prosody learning would be linked to their pitch processing but not to their formant processing because previous research (Jasmin et al., 2021) suggests that Chinese learners primarily rely on their L1 strategies (prioritizing pitch information) when learning L2 English prosody.[1] To follow up on the dimension-specific relationship between formant discrimination and L2 vowel acquisition in task-based L2 training (simulating naturalistic L2 speech learning), we conducted a separate experiment with a total of 70

Chinese students of English as a foreign language; the results are reported elsewhere (Xu et al., in press).

Both pitch and formant acuity were measured through two separate discrimination tasks using nonspeech stimuli, a format widely employed in L1 acquisition and hearing research (Surprenant & Watson, 2001). The test–retest reliability of these tasks was considered "fair" (ICC = .50–.60; Saito & Tierney, 2024). Additionally, the outcomes of these discrimination tasks have been shown to predict various L2 phenomena under both naturalistic (Saito, Sun et al., 2022) and training (Lengeris & Hazan, 2010) conditions.

Materials

A range of synthesized tokens were prepared, characterized by simple acoustic properties (e.g., flat fundamental frequency contours and unchanging spectral shape), making them unlikely to be perceived as human speech. For each of the two subtests (formants, pitch), the stimuli were identical except for the target acoustic dimension. For the formant subtest, we created 101 complex tones, consisting of one standard stimulus (Level 0) and 100 comparison stimuli (Levels 1 to 100). Each tone lasted 500 ms, with two 5-ms linear amplitude ramps at the start and end. The fundamental frequency was set at 100 Hz, with harmonics up to 3000 Hz. Three formants were included at 500 Hz, 1500 Hz, and 2500 Hz, using a parallel formant filter bank (Smith, 2007). The second formant (F2) for the standard was fixed at 1500 Hz, while for the comparison tones it ranged from 1502 to 1700 Hz in increments of 2 Hz. For the pitch subtest, 101 four-harmonic complex tones were prepared. These also featured a 5-ms linear ramp at both the beginning and end. The standard stimulus's fundamental frequency was set at 330 Hz, with comparison tones ranging from 330.3 Hz to 360 Hz in 0.3 Hz steps.

Procedure

During each trial, participants listened to three synthesized stimuli. The second stimulus was always constant, while either the first or last stimulus varied. Participants had to identify the differing sound by pressing "1" or "3" on a computer screen, corresponding to whether the first or third stimulus was the odd one out. For time efficiency, Levitt's (1971) adaptive threshold procedure was used to adjust the difference between stimuli based on participant performance. Testing began at the midpoint of the comparison stimuli (Level 50 out of 100), with an initial step size of 10. If a participant responded incorrectly, the task difficulty was decreased by increasing the stimulus difference by 10 steps. Conversely, if a participant gave three consecutive correct responses, the difficulty was increased by reducing the stimulus difference by 10 steps. The size of the change in difficulty across trials decreased after every "reversal," which was defined as either a correct followed by an incorrect response or an incorrect followed by a correct response. Following the first reversal, the step size was reduced from 10 to 5, and after the second reversal from 5 to 1. The test concluded after either 70 trials or eight reversals. Participants' auditory processing ability was gauged by averaging the final seven reversal points, with lower scores indicating greater perceptual acuity, meaning participants could discern smaller differences between the standard and comparison stimuli.

All the auditory processing tests are available in L2 Speech Tools (Mora-Plaza et al., 2022).For test manuals and reliability results, see Saito and Tierney (2024).

L2 Speech Production Measures

Although prosodic aspects of L2 speech proficiency have been extensively examined, findings have largely been based on controlled production tasks (e.g., delayed repetition tasks; Trofimovich & Baker, 2006). It has been reported that adult L2 learners can monitor and correct their production when they can fully concentrate on the task, yet such performance may not accurately reflect their L2 proficiency in real-life situations, where they must use L2 prosody accurately and fluently while simultaneously attending to various aspects of language (e.g., vocabulary and grammar; Major, 2001). To elicit participants' spontaneous L2 speech, two different versions of picture narratives from the EIKEN English Exam (Versions A and B; see Appendix S1 in the Supporting Information online) were used. They were counterbalanced to avoid test–retest effects, with the first half of participants engaging in Versions A → B and the other half in B → A. A total of 92 picture narratives (46 participants × T1/T2) were submitted to the rater analyses.

Following the procedure established by Saito et al. (2017), the first 30-s stretch of each participant's speech sample—after removing false starts and silences—was normalized by matching peak amplitude across samples and eliminating initial disfluencies (e.g., false starts, pauses). To rate the speech samples, we recruited five native speakers of British English who had extensive experience (> 10 years) of teaching a variety of L2 learners and who held Master of Arts degrees in Teaching English to Speakers of Other Languages (TESOL), qualifying them as "expert raters." The rating session was individually conducted face to face (in tandem with a trained research assistant).

The raters first received a detailed explanation of English word stress, intonation, and typical L2 errors (e.g., misplacement and absence of

Table 1 Training scripts for English prosody (word stress, intonation)

Word stress	When an English word has more than one syllable, one of the syllables will be a little bit louder and longer than the others. For example, if you say the word "computer", you may notice that the second syllable has more stress (comPUter). If you hear stress being placed on the wrong syllable, or you hear equal stress on all of the syllables in a word, then there are word stress errors.
Intonation	Intonation can be thought of as the melody of English. It is the natural pitch changes that occur when we speak. For example, you may notice that when you ask a question with a yes/no answer, your pitch goes up at the end of the question. If someone sounds "flat" when they speak, it is likely because their intonation is not following English intonation patterns.

stress/intonation; see Table 1). They then listened to five practice L2 speech samples (not included in the main rating session) to familiarize themselves with the rating procedure. They rated each of these sound samples on the quality of English word stress and intonation using a moving slider on a computer screen. For each rating, raters also justified their ratings aloud. On this basis, the trained assistant attempted to identify whether the raters had any misunderstanding about the rubrics (e.g., confusion between word stress vs. intonation or any other dimensions of L2 pronunciation). It was confirmed that all the raters clearly understood and applied the two rubrics: word stress and intonation. The pronunciation quality of each sample was recorded on a 1,000-point scale, with each end of the continuum marked by icons representing *excellent* (a smiling face) and *poor* (a frowning face). For on-screen labels, see Figure 1.

Once comfortable, the raters proceeded to assess the main data set in a session lasting approximately 1.5 hr, with the speech samples played in randomized order. After evaluating the first half of the samples ($n = 46$), raters took a 5-min break. Throughout the sessions, raters could consult with the assistant if they had any questions or confusion, though no such issues arose during the training sessions. Similarly to the raters in a previous study (Saito et al., 2017), the five raters from the current study exhibited a relatively high level of agreement (Cronbach's alpha = .93 for word stress and .90 for intonation). For each speaker, the raters' scores were averaged across dimensions (word stress, intonation) at T1 and T2, respectively. These scores (summarized in Table 3) were subsequently subjected to mixed-effects regression analyses (for details, see the Results section).

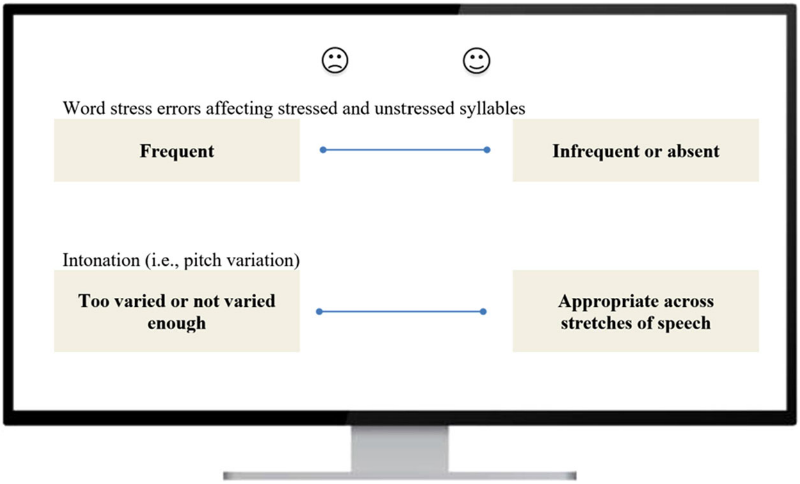

Figure 1 On-screen labels for rating English prosody. Raters assessed the qualities of stress and intonation separately, using a moving slider upon hearing each sample. Ratings ranged from 0 (*poor*) to 1,000 (*excellent*).

Table 2 Descriptive summary of auditory processing test results for Study 1

Test	*M*	*SD*	IQR	Q1	Q2	Q3	Q4
Pitch discrimination	13.76	9.30	7.59	3.00–8.00	8.00–10.83	11.00–15.50	15.71–45.00
Formant discrimination	24.67	18.63	30.26	3.00–8.00	9.00–17.29	20.00–37.71	39.75–67.33

Note. Lower scores represent more precise auditory processing. IQR = interquartile range; Q = quartile.

Results

All the relevant data sets are available via the Open Science Framework (http://doi.org/10.17605/OSF.IO/9UAN8). Participants' auditory processing scores for pitch and formant discrimination are summarized in Table 2 and visually presented in Figure 2. The results of the Kolmogorov–Smirnov test revealed that both discrimination scores were positively skewed (pitch discrimination: $D = .237$, $p = .009$; formant discrimination: $D = .200$, $p = .043$). Many participants demonstrated particularly precise auditory processing in pitch discrimination ($M = 13.76$) compared to formant discrimination ($M = 24.67$). These results are unsurprising, given that the participants were tonal language users (L1 Mandarin Chinese). Interestingly, Spearman correlation analyses showed that the correlation between the pitch and formant discrimination test scores did not reach statistical significance ($r = .210$, $p = .160$), suggesting that the two discrimination tests assessed different dimensions of auditory processing (pitch vs. formant processing).

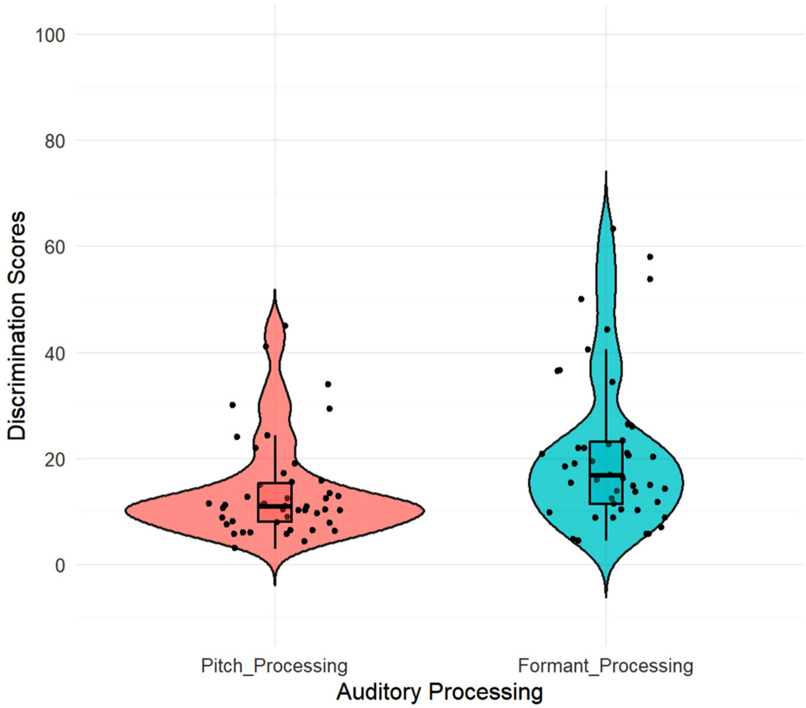

Figure 2 Distribution of two types of auditory processing scores among 46 Chinese learners of English in the UK. Participants' pitch processing scores were generally lower, indicating greater precision, compared to their formant processing scores.

As shown in Table 3, the descriptive statistics of participants' L2 prosodic proficiency indicated little change in their performance between T1 and T2. The relatively stable results may be attributed not only to the brief duration of the project (5 months of study abroad) but also to the participants' significant individual differences in L2 usage both inside and outside of school settings, as well as in auditory processing abilities.

The first objective of the inferential statistical analyses was to examine how participants' performance changed during the 5 months of study abroad in the UK in relation to their different profiles of auditory processing (pitch discrimination) even when their L2 use was accounted for. To this end, a series of mixed-effects regression analyses were performed using the R statistical environment (Version 4.3.1; R Core Team, 2023). To construct models (Models 1–4; see below), the lme4 package was adopted (Bates et al., 2023).

Table 3 Descriptive statistics of participants' L2 prosodic proficiency, L2 use, and auditory processing

	T1						T2					
Characteristic	*M*	*SD*	95% CI	Range	D^a	*p*	*M*	*SD*	95% CI	Range	D^a	*p*
L2 prosodic proficiency												
Word stress[b]	595	125	[557, 632]	328–854	.183	.090	587	134	[547, 627]	365–867	.181	.097
Intonation[b]	476	148	[432, 520]	275–895	.179	.104	478	151	[433, 523]	227–849	.176	.115
L2 use												
L2 outside school (%)	29.9	17.7	[24.6, 35.1]	2.50–80.00	.174	.121	28.1	20.1	[22.1, 34.1]	2.50–95.00	.173	.127
Auditory processing												
Pitch discrimination	13.4	9.07	[11.0, 16.5]	3.00–45.00	.152	.237						
Formant discrimination	21.5	14.0	[16.8, 25.3]	4.43–63.25	.158	.200						

Note. T1 = Time 1; T2 = Time 2. [a]Values for Kolmogorov–Smirnov tests of normality. [b]1,000 points.

Table 4 Summary of Model 1

Fixed effects	F	p	η_p^2	$R^2_{conditional}$	$R^2_{marginal}$
Dimension	46.982	< .001	.26	.779	.225
Time	5.631	.019	.04		
Pitch processing	5.363	.025	.11		
Dimension × Time	0.399	.528	< .01		
Dimension × Pitch Processing	1.130	.289	< .01		
Time × Pitch Processing	9.405	.002	.07		
Dimension × Time × Pitch Processing	1.091	.298	< .001		

Note. Model formula: DV ∼ time∗dimension∗pitch_processing + (1|ID).

In these models, the participants' L2 prosody scores were used as dependent variables. The predictor variables included time (T1, T2), prosodic dimension (word stress, intonation), pitch processing (pitch discrimination scores), L2 inside (T1, T2), and L2 outside (T1, T2). ("L2 inside" and "L2 outside" refer to L2 usage inside and outside of school.)

In Model 1 (DV ∼ time*dimension*pitch_processing + (1|ID)), we aimed to examine how participants' prosodic performance changed over time in relation to their continuous auditory processing scores. As shown in Table 4, main effects of time and auditory processing reached statistical significance (p = .019, .025, respectively), suggesting that the 5 months of study abroad did impact participants' performance and that those with more precise pitch processing demonstrated more targetlike L2 English prosody. Importantly, interaction effects of time and auditory processing were significant (p = .002). First, using emtrends in the emmeans package (Version 1.10.7; Lenth, 2025), we carried out a simple slopes analysis to explore how the effect of auditory processing on L2 prosodic proficiency changed over time. For T1, the slope of pitch threshold was −0.00266 (SE = 0.002), with a confidence interval ranging from −0.00667 to 0.00134, suggesting a nonsignificant effect. At T2, the slope was −0.00619 (SE = 0.002), with a confidence interval from −0.01020 to −0.00219, indicating a significant negative effect of pitch threshold on the dependent variable (DV). This demonstrates a stronger and statistically significant influence of pitch threshold at T2, showing that participants with more precise auditory processing attained more advanced L2 prosodic proficiency during this period (for a visual summary, see Figure 3).

To further examine the interaction effects of time and auditory processing, a post hoc analysis was conducted to determine the level of auditory precision required for participants to show significant improvement over time. For this

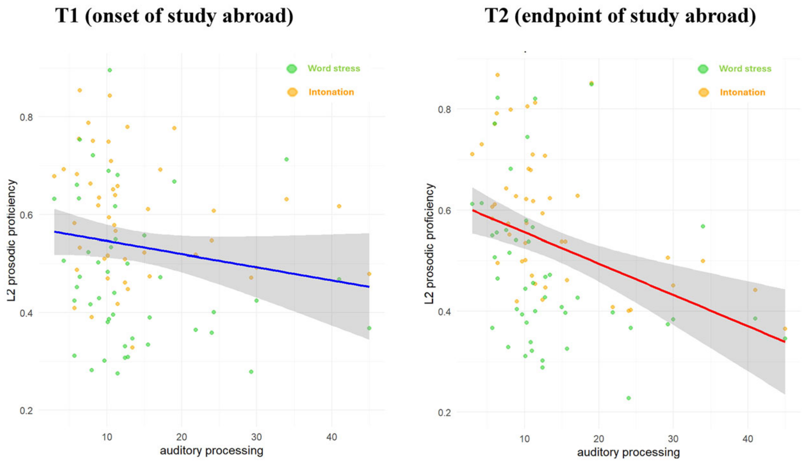

Figure 3 Visual plots depicting the relationship between L2 prosodic proficiency and auditory processing at the beginning (T1, left panel) and end (T2, right panel) of the project. Word stress production is represented by green dots and intonation production by yellow dots. The lines represent linear trendlines, with shaded regions indicating the 95% confidence intervals. Higher prosodic proficiency scores indicate more advanced proficiency, while lower auditory processing scores signify more precise pitch discrimination. At T1, the role of auditory processing in L2 prosodic proficiency remained nonsignificant. At T2, participants with more precise auditory processing demonstrated significantly more advanced L2 prosodic proficiency.

purpose, we applied the Johnson–Neyman technique using the interaction package. This technique identifies regions where the expected values of a criterion variable for two groups differ significantly, based on one or more predictor variables (Potthoff, 1964). This approach allows the statistical analyses to determine the auditory processing scores at which participants can be categorized as either improving or regressing, rather than relying on predetermined indices.

As shown in Figure 4, a pitch threshold of 14 was identified as the cutoff point: Simple effects of time were significant ($p < .05$) when participants' pitch processing scores were above 14 (out of 100). However, the time effects became nonsignificant when their pitch processing scores were below 14. Given that higher auditory processing scores indicate lower auditory precision, these findings suggest that when participants' auditory precision was less accurate than the threshold (14 out of 100), auditory processing began to negatively impact their L2 prosody development.[2]

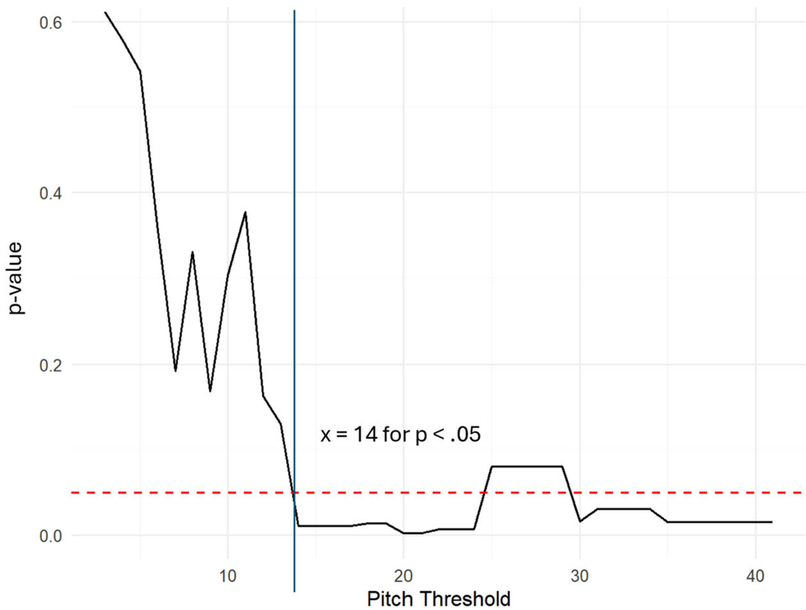

Figure 4 Visual plot showing that simple effects of time reached statistical significance ($p < .05$) after the pitch threshold of 14. This indicates that the impact of auditory precision on L2 speech learning during the period of study abroad was significant among those beyond the threshold ($x = 14$), whereas those below the threshold did not show statistically significant change over time.

To further follow up the nature of the interaction effects (significant effects among those beyond the pitch threshold of 14 out of 100 points), we divided the 46 participants into post hoc group categories based on their threshold: $x < 14$ for the normative group ($n = 33$) and $x < 14$ for the low-precision group ($n = 13$). Model 2 (DV $\sim$ time*dimension*group + (1|ID)) was constructed to determine how each group (normative vs. low precision) changed their L2 prosodic proficiency over time.

As shown in Table 5, the interaction effects between time and group reached statistical significance ($p = .017$). We performed a set of post hoc multiple comparison analyses using the emmeans package. As visually summarized in Figure 5, pairwise comparisons adjusted for multiple testing using Tukey's method revealed no significant difference between T1 and T2 for the normative auditory precision group ($b = -0.012$, $SE = 0.012$, $t = -1.012$, $p = .313$). However, in the lower auditory precision group, a significant decrease in proficiency was observed from T1 to T2 ($b = 0.044$,

Table 5 Summary of Model 2

Fixed effects	F	p	η_p^2	$R^2_{conditional}$	$R^2_{marginal}$
Dimension	99.3055	< .001	.43	.773	.192
Time	1.7593	.187	.01		
Group	3.0576	.087	.06		
Dimension × Time	0.7544	.386	< .01		
Dimension × Group	0.9142	.340	< .01		
Time × Group	5.7688	.017	.04		
Dimension × Time × Group	1.2691	.261	< .01		

Note. Model formula: DV ∼ time∗dimension∗group + (1|ID).

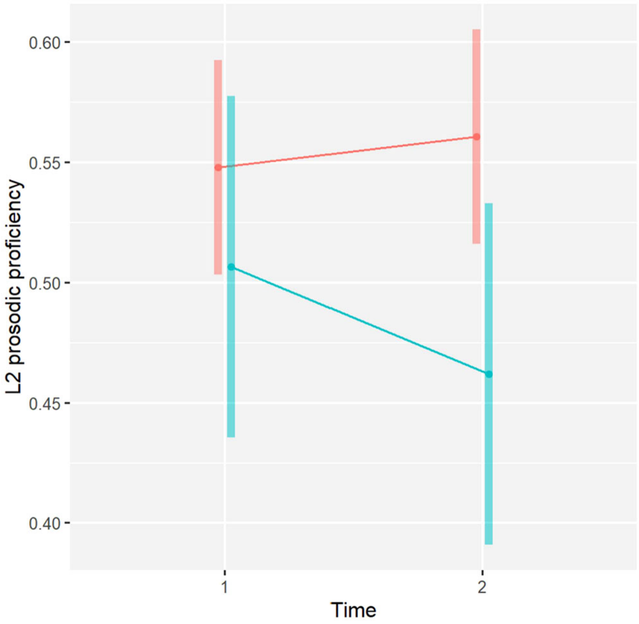

Figure 5 Changes in group means and 95% confidence intervals for L2 prosodic proficiency over time (T1 vs. T2), comparing the normative auditory precision group (red) and the lower auditory precision group (blue). The prosodic proficiency in the normative precision group remained consistent over time; however, there was a significant reduction in proficiency among those with lower auditory precision, resulting in significantly lower scores than those of the normative group at the project's end (T2).

$SE = 0.012$, $t = 2.201$, $p = .0029$). These findings suggest that while there was no clear improvement in the participants' L2 prosodic proficiency over time, the proficiency of those with lower auditory precision became less targetlike.

In terms of the group differences at T1 and T2, another set of post hoc analyses showed that the difference in L2 prosodic proficiency between the normative and lower auditory precision groups was small and not statistically significant at T1 ($b = 0.041$, $SE = 0.041$, $t = 0.990$, $p = .326$). At T2, however, a statistically significant difference was observed between the same groups, where the group with lower auditory precision showed less advanced L2 prosodic proficiency relative to the normative group ($b = 0.098$, $SE = 0.041$, $t = 2.362$, $p = .022$). The results suggest that the effects of auditory processing in L2 speech learning could be mainly attributable to those with relatively low auditory precision, as their learning patterns appear limited and may be characterized as regression despite their intensive input opportunities during study abroad. See Figure 5 for the post hoc group summary (normative vs. low) and Figure 6 for individual learners' trajectories within each post hoc group category.

While the majority of the participants with relatively low auditory precision ($x > 14$) showed regression, the normative group's L2 prosodic proficiency ($x < 14$) did not show significant changes over time. A follow-up analysis was performed to examine whether certain variables are linked to improvement over time within individuals in the normative precision group ($n = 33$). Given the relatively small sample size, the subsequent findings should be considered tentative.

Model 3 ($DV_{\text{normative group}} \sim$ time*dimension*L2_use + (1|ID)) was constructed to examine the complex interaction effects of time, L2 use, and auditory processing across the two dimensions of L2 prosodic proficiency (word stress, intonation). L2 use was calculated based on the percentage of time in which participants used L2 English (relative to L1 Mandarin) outside schools at the beginning and endpoint of the project ($M = 28.3\%$, $SD = 18.8\%$, range = 2.5–95%). As summarized in Table 6, the interaction effects of L2 use and time reached statistical significance, $F = 5.613$, $p = .019$. This suggests that the predictive power of auditory processing for L2 prosodic development among those with normative auditory processing profiles may depend on their experience profiles.[3]

For the post hoc analysis, participants were divided into two subcategories. First, we categorized them into regular users versus nonregular users, using a median split for L2 use (23.75%). Calculating and analyzing the estimated

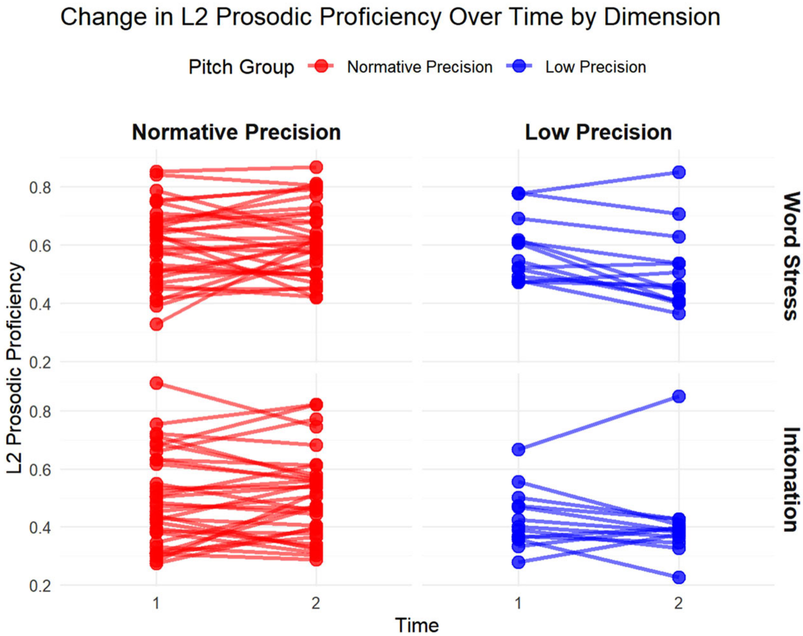

Figure 6 Changes over time (T1 vs. T2) in individual participants' L2 prosody performance, plotted by prosodic features (word stress, intonation) and group (normative precision, low precision). Significant individual variation was observed within the normative precision group, whereas a general declining trend in performance characterized the low precision group.

marginal means (via the emmeans package), multiple comparison analyses demonstrated that regular users with normative auditory precision significantly improved their proficiency over time, from $M = 0.525$ to $M = 0.584$; $b = -0.059$, $SE = 0.023$, $t = -2.498$, $p = .014$ (Bonferroni adjusted). However, such significant learning was not observed in the other group context. The results suggest that a subset of participants with good auditory processing and high L2 use demonstrated improvement. In essence, normative auditory precision alone may not guarantee continuous development; however, it can positively influence L2 speech proficiency, provided that learners regularly use the target language.

While the analyses presented so far have supported a dimension-specific relationship between auditory processing and L2 pronunciation learning (i.e., pitch discrimination scores significantly predicting participants' L2 English prosody development), our final objective was to examine the presumably

Table 6 Summary of Model 3

Fixed effects	F	p	η_p^2	$R^2_{conditional}$	$R^2_{marginal}$
L2 Use	5.328	.022	.04	.770	.104
Time	6.310	.013	.06		
Dimension	35.618	< .001	.28		
L2 Use × Time	5.613	.019	.05		
L2 Use × Dimension	3.278	.073	.03		
Time × Dimension	0.606	.438	< .01		
L2 Use × Dimension	1.202	.275	.01		

Note. Model formula: $DV_{normative\ group} \sim time*dimension*L2\_use + (1|ID)$.

Table 7 Summary of Model 4

Fixed effects	F	p	η_p^2	$R^2_{conditional}$	$R^2_{marginal}$
Dimension	30.549	< .001	.19	.764	.167
Time	1.046	.308	< .01		
Formant processing	1.746	.193	.04		
Dimension × Time	0.080	.776	< .01		
Dimension × Formant Processing	0.113	.736	< .01		
Time × Formant Processing	2.075	.152	.02		
Dimension × Time × Formant Processing	0.004	.948	< .001		

Note. Model summary: $DV \sim time*dimension*formant\_processing + (1|ID)$.

nonsignificant role of participants' auditory processing in another dimension (formant discrimination scores) in L2 prosodic learning. To this end, Model 4 (DV ~ time*dimension*formant_processing + (1|ID)) was constructed. As summarized in Table 7, the results from mixed-effects modeling analyses revealed no significant main or interaction effects of time and auditory processing ($p > .05$) when formant discrimination scores were entered as group variables. These findings confirmed our prediction that auditory processing impacts L2 speech learning at dimension-specific levels, with pitch but not formant discrimination scores being predictive of L2 prosodic learning.

Study 2: Corpus Data

In the context of Chinese learners' acquisition of English prosody during 5 months of study abroad, Study 1 confirmed emerging evidence suggesting a relationship between auditory precision and L2 speech learning (Mueller et al., 2012). Specifically, our results align with previous findings that the effects of auditory processing are particularly pronounced among individuals with lower auditory precision, especially when these learners are exposed to real-life, naturalistic, and challenging learning environments (Chandrasekaran et al., 2010; Correia et al., in press; Perrachione et al., 2011; Ruan & Saito, 2023; Xu et al., in press; W. Zhang et al., 2024). The findings indicate a threshold of 14 out of 100 on the pitch discrimination task: Those unable to detect differences smaller than 4.2 Hz in fundamental frequency exhibited very limited learning progress. Based on the data, we tentatively conclude that approximately one quarter of the population (28.2%; 13 out of 46 participants) may fall into the category of relatively low auditory precision, which could negatively impact relevant areas of L2 speech acquisition, particularly English prosody.

In order to get a precise idea of how many individuals fall below the threshold measured in Study 1, we designed Study 2 to recruit a larger sample of participants with similar backgrounds (400 university-level Chinese learners of English). These participants were asked to take the same pitch discrimination tests (along with other auditory processing tests). Specifically, Study 2 aimed to determine what proportion of learners would fall above or below the 14 out of 100 score (i.e., 4.2 Hz differences in fundamental frequency) in the broader corpus of auditory processing data among Mandarin Chinese learners of English.

Participants

To represent adult individuals with varied proficiency levels and account for the influence of their L1 background, we initially recruited 400 Mandarin Chinese learners of English in both China and the UK. To represent the potential age range of university-level students and match the range of the participants in Study 1 ($M = 23.60$, $SD = 2.27$, range = 21–29), we focused on learners within a narrow age range: 18 to 32 years old. To maximize the sample size, we utilized the online psychology experiment builder, Gorilla (Anwyl-Irvine et al., 2020). We implemented a procedure, established in our previous research (Saito & Tierney, 2024), to ensure the homogeneous quality of data collected via online platforms, comparable to face-to-face conditions. Of the total participants, 95 were derived from our team's previous published study (Saito et al., 2024). The remaining participants formed a combined data set from seven unpublished master's dissertation projects, which investigated the auditory processing profiles of Chinese learners of English and their relation to various types of naturalistic and instructed L2 speech proficiency.

While participants in each project underwent slightly different combinations of auditory processing tests (e.g., discrimination of pitch, formant, duration, and amplitude rise time; reproduction of melody and rhythm), all 400 participants completed both pitch and formant discrimination tests. In this paper, we report the results of these pitch and formant discrimination tests, meanwhile continuing to collect and develop sufficiently large data sets for the other auditory processing test results to inform future analyses.

The participants' ages ranged from 18 to 32 years (*M* = 24.8, *SD* = 2.8). They had received an extensive period of English-as-a-Foreign-Language education in China (*M* = 13.5 years, *SD* = 3.8, range = 5–24). Of these participants, 167 had no experience abroad, while 333 reported varying lengths of immersion experience in the UK (*M* = 1.6 years, *SD* = 2.9, range = 0.4–11), with different timings of arrival (*M* = 21.3 years, *SD* = 3.6, range = 15–32). Additionally, 154 participants reported having musical training, with durations varying widely (*M* = 4.8 years, *SD* = 4.6, range = 0.5–23). All participants reported normal hearing.

Results

All the relevant data sets are available via the Open Science Framework (http://doi.org/10.17605/OSF.IO/9UAN8). In the current investigation, our primary objective was to define various levels of auditory processing abilities (first, second, third, and fourth quartiles) among a total of 400 college-level Chinese learners of English with normal hearing. The descriptive statistics of participants' auditory processing scores are summarized in Table 8 and visually represented in Figure 7. The results of the Kolmogorov–Smirnov normality test indicated that both pitch and formant discrimination scores significantly deviated from a normal distribution (pitch discrimination: $D = .208$, $p < .001$; formant discrimination: $D = .076$, $p < .001$). Similarly to Study 1, participants demonstrated notably lower scores, indicating more precise auditory processing, for pitch information (*M* = 14.70, interquartile range [IQR] = 11.37) compared to formant information (*M* = 31.64, IQR = 27.54). However, Spearman correlation analyses revealed a statistically significant correlation between pitch and formant discrimination scores ($r = .378$, $p < .001$).

An interquartile analysis was conducted to determine the relative position of the threshold identified in Study 1 (pitch processing score = 14 out of 100) within the larger data set ($n = 400$ college-level Chinese learners of English). The results indicated that the pitch discrimination threshold from Study 1 ($x = 14.00$) falls into the third quartile ($11.16 < x < 17.91$). Based on the raw data, 259 participants scored below 14 (indicating more precise pitch processing; range = 2.33–13.83), while 141 participants scored above

Table 8 Descriptive summary of auditory processing test results for Study 2

Test	*M*	*SD*	IQR	Q1	Q2	Q3	Q4
Pitch discrimination	14.70	11.25	11.37	2.33–6.66	6.80–11.00	11.16–17.91	18.16–97.75
Formant discrimination	31.64	17.15	27.54	2.33–16.33	16.66–30.00	30.50–43.83	44.00–85.83

Note. Lower scores represent more precise auditory processing. IQR = interquartile range; Q = quartile.

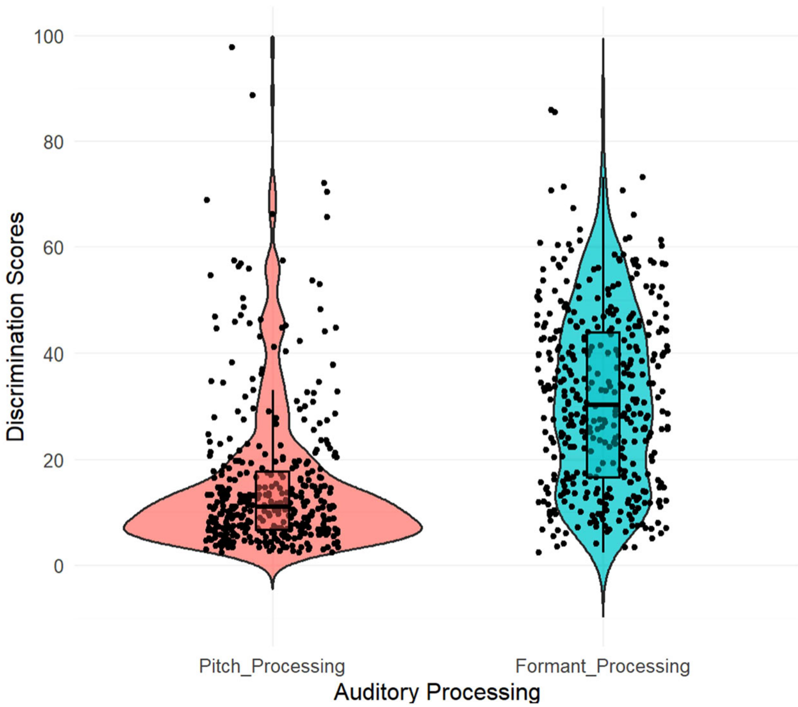

Figure 7 Distribution of two types of auditory processing scores among a total of 400 college-level Chinese learners of English in Study 2. In a similar pattern to that of Study 1, participants' pitch processing scores were generally lower, indicating greater precision, compared to their formant processing scores.

14 (indicating less precise pitch processing; range = 14.00–97.75), suggesting that approximately 35.2% of the population has lower auditory precision. Interestingly, the suggested ratio in Study 2 (35.2%) was slightly higher than the ratio in Study 1 (28.2%).

Discussion

Emerging evidence suggests that domain-general auditory precision significantly impacts the rate of naturalistic L2 speech learning. Notably, individuals exhibiting lower auditory precision seem to make limited progress and regress, even when receiving similar input opportunities to those with normative auditory processing abilities (e.g., Perrachione et al., 2011). These findings are crucial, as learners with relatively low auditory precision may require remedial strategies to fully benefit from each learning opportunity, regardless of their aptitude profiles. Recently, auditory processing tests have been made publicly available for researchers and practitioners to evaluate individuals' aptitude profiles, particularly in real-life, naturalistic, and challenging L2 speech learning environments (e.g., study abroad; Saito & Tierney, 2024).

The current investigation aimed to define what constitutes relatively low auditory precision in the context of English prosody acquisition among college-level Chinese learners of English with varying pitch processing abilities. In Study 1, we tracked the English prosodic proficiency of 46 Chinese learners over a 5-month period of study abroad in the UK, seeking to identify the pitch discrimination threshold that determined who progressed or regressed during the study. To estimate the percentage of the population that may fall into the categories of relatively high versus low auditory precision, we carried out Study 2, developing a larger corpus data set of 400 college-level Chinese learners of English.

First and foremost, individual differences in auditory processing were significantly correlated with the rate of L2 learning. As predicted, a negative aptitude effect was observed among learners with relatively low pitch processing abilities. Thirteen out of 46 learners who did not meet the pitch processing threshold ($x > 14$ out of 100 points; unable to detect differences smaller than 4.2 Hz in fundamental frequencies) exhibited a significant decline in L2 prosodic proficiency after an extended period of study abroad.

Our findings align with those of existing research on L2 speech training, which show that learners with normative auditory processing benefit from real-life speech stimuli (e.g., multiple talkers; Chandrasekaran et al., 2010), communicatively authentic tasks (e.g., task-based language learning; Ruan & Saito, 2023; Xu et al., in press), and incidental L2 speech training without provision of explicit phonetic information (Correia et al., in press). In contrast, learners with lower auditory precision may encounter obstacles and could benefit more from tailored speech stimuli, such as those produced by a single talker (Perrachione et al., 2011) or stimuli where key acoustic features are enhanced (Iverson et al., 2005).

In our data set, participants with low pitch processing abilities demonstrated less L2-like and more L1-like prosody after 5 months of immersion. Those with low pitch precision may struggle to notice crosslinguistic differences in pitch usage between Mandarin (a tonal language) and English (a stress-timed language). Mandarin uses pitch variations within a syllable to convey meaning, whereas English uses pitch variations across syllables to emphasize meaning at the sentence level. Learners with imprecise pitch processing may reinforce L1 prosodic structures, such as varying pitch at every syllable, without fully adopting L2-like sentence-level prosody, thereby producing L2 prosody that is less L1-like over time (Chrabaszcz et al., 2014; J. Zhang et al., 2018).

Given the findings of recent longitudinal research, regression in L2 speech learning is not unexpected. Some learners may continue to rely on L1 strategies even when exposed to L2 input. For instance, Munro et al. (2024) demonstrated that while certain Slavic learners' production of English stops became more intelligible in the initial months of immersion (voice onset time = 30 ms), their production became more L1-like as they received more L2 input (voice onset time = 10 ms). Similarly, Japanese learners of English struggle with the distinction between [r] and [l], often relying on F2 and duration variations rather than the more reliable F3 variations, reflecting patterns from their L1 (e.g., the Japanese tap). Research on adult Japanese learners indicates that their perceptual strategy for distinguishing English [r] and [l] tends to be resistant to change, even after extensive focused training (Ingvalson et al., 2012) and years of immersion experience (Saito, 2013).

In Study 2, using the larger corpus data set of 400 learners, we found that approximately 35.2% of the population of L1 Chinese speakers fell below the pitch discrimination threshold identified in Study 1 ($x > 14$ out of 100 points). Without remedial strategies, these learners may face challenges when engaging in naturalistic L2 speech learning (as in study abroad). This figure (35.2%) is relatively high and needs to be discussed with caution, especially in light of research on neurodiversity and language learning challenges. In both the UK and USA, for example, around 10–20% of children are considered neurodiverse, with many experiencing language-related difficulties (e.g., dyslexia, affecting 5–10% of children; Snowling, 2013). Similarly, approximately 10% of UK children struggle with language disorders severe enough to impact academic performance (Norbury et al., 2016).

While the 35.2% figure represents a relatively large proportion, the findings do not suggest that the negative impacts of auditory imprecision are inevitable. Even individuals diagnosed with amusia can improve their pitch perception and surpass the amusia threshold (Whiteford & Oxenham, 2018). Instead, our findings align with the concept of aptitude–treatment interaction (DeKeyser, 2012), which advocates for profile-matched treatment based on individual neurocognitive backgrounds. Learners with adequate auditory precision (approximately the top 2.5 quartiles; 65%) are encouraged to engage in real-life L2 speech learning environments, such as study abroad, task-based language learning, and incidental L2 speech learning. Conversely, learners with low auditory precision (bottom 1.5 quartiles) may benefit from additional directed training, such as explicit phonetic instruction (Suzukida & Saito, 2023) or auditory processing training (Saito, Petrova et al., 2022).

Here, it is crucial to stress our conceptual stance once again (i.e., mechanistic functionalism: McMurray et al., 2023), which posits that learners with varying perceptual-cognitive profiles require different optimal strategies for any learning experiences. We do not view learners with low auditory precision in terms of deviations or deficits (e.g., regressing during L2 immersion compared to normative auditory learners who benefit from such immersion). Instead, we argue that a combination of naturalistic immersion and follow-up training (e.g., phonetic and/or auditory training) represents an optimal strategy for learners with low auditory precision to fully capitalize on the naturalistic immersion experience. It is important to stress that this follow-up training is intended to complement, rather than replace, naturalistic immersion.

Another key methodological consideration is that the analysis was limited to a very specific time window of L2 speech learning (LOR = 5–10 months). Although learners with relatively low auditory precision were found to be at a disadvantage for L2 English prosody acquisition during this period, we still lack insight into the rest of their L2 learning journey. Future studies that track the prosodic development of learners with low auditory precision beyond the mid-phase of immersion could shed light on whether these learners cease progressing, with their L1-like performance fossilized, or whether they eventually notice the acoustic characteristics of L2 speech and resume finetuning. In the latter case, significant individual variation may still exist in terms of how much and to what extent these "readjusted" learners halt regression (L1-like performance) and ultimately attain L2-like performance over time.

Indeed, Munro et al.'s (2024) 10-year longitudinal data set of Slavic L2 learners of English revealed that even if some L2 learners regressed, particularly around 6–12 months of immersion (following an initial spike in L2 speech learning), they often regained L2-like performance after extensive immersion (LOR = 7 years). This developmental trajectory reflects a U-shaped learning pattern, where initial success is followed by regression and then further refinement. This phenomenon has been observed in many different L2 grammar learning contexts, where learners first display success, then regression due to overgeneralization, and finally refinement as they gain more exposure (McLaughlin, 1990). According to the usage-based and emergentist account of L2 learning (Ellis, 2015), learners initially restructure their mental representations of language based on the frequency and distribution of forms encountered in their L2 input. They often overgeneralize linguistic rules during this process, leading to errors. However, through exposure to more input and corrective feedback in socially interactive environments, learners refine their L2 system.

In the context of L1 acquisition, it is important to note that, while auditory imprecision may slow down the speed of learning, auditorily disadvantaged individuals still demonstrate normal language development in the long run. This could be attributed to ample exposure to caregivers' linguistic input (Rosen, 2003), their reliance on executive functions (e.g., greater attentional control; Snowling et al., 2018), and their optimal use of compensatory perceptual strategies (e.g., amusic individuals using duration rather than pitch cues for speech processing; Jasmin et al., 2020).

Taken together, our current investigations provide some of the first longitudinal evidence suggesting that normative auditory processing is necessary, but not sufficient, for successful L2 speech learning. On the one hand, normative auditory precision facilitates and accelerates L2 learning, particularly when learners actively seek, receive, and use more communicatively authentic L2 input (Flege & Bohn, 2021). On the other hand, a lack of adequate auditory precision may disadvantage learners in such environments. These learners may struggle in naturalistic L2 speech settings at least within a short period of time (LOR $<$ 1 year), and thus may benefit more from explicit phonetic instruction and/or auditory processing training, particularly before engaging in naturalistic L2 learning. This approach would help them to fully optimize their L2 learning opportunities.

Given the results presented so far, it is finally pertinent to discuss why we observed significant correlations between auditory processing and L2 prosodic proficiency at the endpoint (T2) but not at the outset of the project (T1). These results align with emerging cross-sectional evidence indicating that auditory processing can explain significant variances, particularly among L2 learners with adequate amounts of L2 immersion experience but not among those with little or no immersion experience (e.g., Saito et al., 2021). These studies suggest that the effects of auditory processing on L2 speech learning become more apparent as learners gain more naturalistic immersion experience (the link between pitch processing and L2 prosody performance being significant at T2 but not T1).

At the onset of naturalistic immersion, learners can quickly improve their L2 speech proficiency, regardless of their auditory processing profiles. After the initial phase of L2 speech learning, those with greater auditory processing capabilities can maximize each practice opportunity, provided they actively seek such opportunities. However, those who do not reach normative auditory precision may become confused if merely exposed to real-life, complex L2 input. This exposure could potentially fossilize L1 speech processing patterns and hinder L2 speech processing (Perrachione et al., 2011).

Limitations

Given the exploratory nature of this project, several limitations should be acknowledged to guide future replication studies. These are discussed in the sections below.

Scope of Auditory Processing and Diversity of Participants

While insightful, the benchmarks suggested in this study are exclusively limited to pitch and formant discrimination and their relation to Chinese learners' English prosody development. Future studies should extend the scope to include (a) a larger and more diverse sample of L2 learners with varied L1 backgrounds and (b) different types of auditory processing tests (see, e.g., Holt et al., 2018, for dimension-selective attention; Saito et al., 2021, for audio-motor integration). Validation studies are also needed to determine how L2 learners' auditory processing performance can uniquely predict dimension-specific aspects of L2 speech learning (e.g., discrimination of second and third formants vs. English [r] and [l] acquisition among Japanese learners; Saito, Hanzawa, et al., 2022).

Normative Auditory Precision

The study hinted at the possibility that a lack of normative auditory precision may hinder successful L2 speech learning. In the L1 acquisition literature, while many studies have shown such negative impacts, these often compare groups that are substantially contrastive (those with vs. without language disorders). The relationship between auditory processing and its impact on language learning/delay could be highly mixed within normal hearing individuals (Suprenant & Watson, 2001). Thus, it is premature to draw definitive conclusions about the negative effects of auditory imprecision in L2 speech learning. Further studies are needed to precisely disentangle why less precise auditory processing can slow down or hinder L2 speech learning processes.

Cognitive Abilities and Auditory Processing

Although we demonstrated that participants' pitch discrimination scores were independent of formant discrimination scores, these test scores could be linked to other cognitive abilities such as attentional control and working memory, given that auditory processing was measured via behavioral tasks (Rosen, 2003). Future studies should control for the confounding effects of other cognitive abilities and explore the relative strength of predictive power among auditory processing and other well-studied cognitive variables (see, e.g., Huensch, 2024, for working memory; Darcy, Mora, & Daidone, 2016, for inhibitory control). It would be particularly interesting to investigate whether those with low auditory precision can compensate for the negative impact of auditory processing, especially when they possess relatively great cognitive abilities (Snowling et al., 2018). Another way to investigate the predictive power of auditory processing for language learning without the confounding influence of cognitive variables would be to make use of neural measures such as the frequency-following response, an assessment of auditory neural encoding that is relatively unaffected by attention and cognitive state (Varghese et al., 2015).

Long-Term L2 Immersion and Cognitive Abilities

Our analyses focused on moderately experienced L2 learners (length of residence < 1 year), whereas much attention has been given to the role of aptitude in ultimate attainment among highly experienced L2 learners (length of residence > 10 years; Abrahamsson, 2012). The incidence of highly advanced L2 speech proficiency has been linked to a range of cognitive abilities, such as associative, working, and procedural memory (Granena & Long, 2013; Linck et al., 2013; Silbert et al., 2015). Future studies could explore how those with varied perceptual and cognitive abilities engage in long-term L2 immersion experiences and the extent to which they can improve their L2 speech proficiency.

Tonal Versus Nontonal L1 Backgrounds in L2 Prosody Acquisition

While the current study highlighted the acquisition of L2 prosody, there is ongoing discussion regarding how tonal versus nontonal language speakers approach this feature differently. Research suggests that tonal language users may possess an advantage in L2 prosody acquisition, particularly when pitch-related cues, such as in English word stress and intonation, are involved, owing to their enhanced sensitivity to various pitch features, including pitch height and contour. In contrast, nontonal L1 speakers tend to be more influenced by the spectral properties of stress, such as vowel duration and intensity, even when these are not the primary cues in L2 prosody acquisition (for a comprehensive summary, see Jongman & Tremblay, 2020). The present study represents an initial step toward examining the complex interaction between pitch processing and L2 prosody acquisition, focusing on Chinese speakers as tonal language users. This emphasis opens up several avenues for future research. First, it would be informative to explore whether the findings differ if the L2 learners are nontonal language speakers (e.g., French speakers), who may rely more on spectral and intensity cues rather than pitch for L2 prosody acquisition. Second, the findings could be replicated in contexts where English prosody is driven by pitch as a primary cue (e.g., word stress and intonation), as well as in contexts where duration, rather than pitch, plays a critical role (e.g., phrase boundaries).

Temporal Cues in L2 Prosody Acquisition

In the current investigation, we posited and provided supporting evidence that more precise pitch processing could facilitate L2 English prosody acquisition. In the case of word stress, existing research suggests that native English listeners utilize not only pitch as a primary cue but also duration and intensity as secondary cues (Fry, 1958; Lieberman, 1960). Interestingly, Mandarin Chinese listeners tend to rely excessively on pitch (Y. Zhang et al., 2008). Future research should investigate how Chinese listeners with heightened sensitivities to duration and intensity may enhance their English prosody proficiency, enabling them to continue using pitch as a cue while achieving near-native performance.

Conclusion

Drawing on the sensory theory of L1 acquisition, a growing body of research suggests that domain-general auditory processing mechanisms underpin the rate of adult L2 speech learning (Mueller et al., 2012). This is particularly relevant when learners engage with the target language in real-life, communicatively authentic contexts in a similar way to L1 acquisition (Saito et al., 2022). In this study, we built on the emerging hypothesis that, while auditory precision generally relates to language learning, its influence would be most evident among those with lower auditory precision, suggesting that reduced auditory precision could impede their L2 speech development.

Our analysis focused on the dimension-specific relationship between the development of English prosody proficiency in Chinese learners and their individual differences in pitch processing abilities. Study 1 tracked 46 Chinese learners in study-abroad contexts and found a significant interaction between L2 usage and improvement in the group with normative auditory processing abilities. Echoing prior literature (Flege & Bohn, 2021), this suggests that learners with normative pitch processing abilities continued to improve their English prosody production as a result of increased daily L2 usage.

In contrast, learners with low pitch processing abilities exhibited regression, likely because these perceptually unprepared learners processed more authentic L2 input through the prosodic patterns of their L1, leading to outputs that were more characteristic of the L1 and less representative of the L2. As noted by Munro et al. (2024), and conceptualized as the U-shaped learning trajectory under the usage-based account of L2 learning (Ellis, 2015), the phenomenon of regression could be found among certain L2 learners, especially during the early-to-mid phase of L2 speech learning. Here, we add that the regression could be linked to individual differences in imprecise auditory processing. Drawing on a large corpus of pitch discrimination data from a similar population ($n = 400$ university-level Chinese learners of English), Study 2 indicated that approximately the bottom 1.5 quartiles (35.2%) of individuals with normal hearing fell into this low auditory precision category.

The interplay of experience, auditory processing, and L2 speech learning offers multiple implications for both practitioners and researchers. First, exposing learners with suboptimal auditory processing abilities to naturalistic, real-life L2 speech input might hinder—or at least not promote—their L2 speech learning. Such individuals may benefit from modified, tailored L2 input and training to effectively discern and comprehend perceptual differences between the L1 and L2, thereby developing new L2 phonetic representations relative to their L1 counterparts. Those considered perceptually unready (i.e., below the normative auditory precision range) should be encouraged to engage in explicit phonetic instruction, wherein they can hear acoustically enhanced speech, particularly in face-to-face settings (Suzukida & Saito, 2023).

In contrast, individuals with normative auditory processing should be encouraged to use and practice the target language as often as possible. They are considered perceptually ready to make the most of ample and communicatively authentic L2 practice opportunities. These individuals should be encouraged not only to develop and revise their L2 phonetic representations but also to automatize them through such opportunities (e.g., study abroad, task-based language learning).

Final revised version accepted 26 February 2025

Open Research Badges

This article has earned Open Data and Open Materials badges for making publicly available the digitally-shareable data and the components of the research methods needed to reproduce the reported procedure and results. All data and materials that the authors have used and have the right to share are available at http://sla-speech-tools.com and http://doi.org/10.17616/R31NJNAX. All proprietary materials have been precisely identified in the manuscript.

Notes

1 It is important to note that the overreliance on pitch cues may not be an optimal strategy because other aspects of acoustic information equally play important roles in English prosody perception (e.g., formant processing for vowel reduction; Y. Zhang & Francis, 2010).

2 In the main analysis, we adopted a less conservative threshold, suggesting that participants with pitch processing scores above 14 exhibited significant time effects, indicative of slower L2 speech learning. However, as shown in Figure 4, the time effects became nonsignificant when participants' pitch processing scores ranged between 25 and 30. These findings suggest a more conservative threshold of 30, with significant time effects observable only when participants' pitch processing scores exceeded this value. Referring to the corpus data results from Study 2, approximately 15% of the normal hearing population may fall into the category of learners with low auditory precision based on the more conservative threshold. By contrast, using the less conservative threshold, approximately 35% of the normal hearing population could be classified as learners with low auditory precision.

3 The low-auditory-precision group showed comparable L2 use profiles at T1 and T2 ($M = 27.1\%$, $SD = 18.1\%$, range = 3.5–80%). A one-way analysis of variance (ANOVA) indicated no significant group difference in L2 use, $F = 0.164$, $p = .687$. To examine whether L2 use predicted speech development within the low-auditory-precision group, we conducted a similar analysis: DV low auditory group ~ time*dimension*L2_use + (1|ID). The interaction effect between L2 use and time was not statistically significant, $F = 0.002$, $p = .958$. These findings suggest that, unlike in the normative group, L2 speech development among low-auditory-precision learners may not be influenced by their amount of L2 use during the study period (LOR = 5–10 months).

References

Abrahamsson, N. (2012). Age of onset and nativelike L2 ultimate attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition*, *34*(2), 187–214. https://doi.org/10.1017/S0272263112000022

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N. Z., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral science platform. *Behavior Research Methods*, *52*(1), 388–407.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2023). lme4: Linear mixed-effects models using Eigen and S4 (Version 1.1-37) [R package]. R Foundation for Statistical Computing. https://cran.r-project.org/package=lme4

Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquiere, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities*, *32*(2), 560–570. https://doi.org/10.1016/j.ridd.2010.12.020

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2339–2349. https://doi.org/10.1121/1.2642103

Chandrasekaran, B., Sampath, P. D., & Wong, P. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*(1), 456–465. https://doi.org/10.1121/1.3436524

Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1468–1479. https://doi.org/10.1044/2014_JSLHR-L-13-0279

Correia, S., Rato, A., Ge, Y., Fernandes, J. D., Kachlicka, M., Saito, K. & Rebuschat, P. (in press). Effects of phonetic training and cognitive aptitude on the perception and production of non-native speech contrasts. *Studies in Second Language Acquisition*.

Darcy, I., Mora, J. C., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*, *66*(4), 741–773. https://doi.org/10.1111/lang.12161

DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, *62*, 189–200. https://doi.org/10.1111/j.1467-9922.2012.00712.x

Ellis, N. C. (2015). Cognitive and social aspects of learning from usage. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 49–73). Mouton de Gruyter.

Flege, J. E, & Bohn, O.S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, *26*(2), 275–301. https://doi.org/10.1017/S0272263104262064

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, *1*, 126–152. https://doi.org/10.1177/002383095800100207

Goswami, U. (2015). Sensory theories of developmental dyslexia: Three challenges for research. *Nature Reviews Neuroscience*, *16*(1), 43–54. https://doi.org/10.1038/nrn3836

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311–343. https://doi.org/10.1177/0267658312461497

Holt, L. L., Tierney, A. T., Guerra, G., Laffere, A., & Dick, F. (2018). Dimension-selective attention as a possible driver of dynamic, context-dependent re-weighting in speech processing. *Hearing Research*, *366*, 50–64. https://doi.org/10.1016/j.heares.2018.06.014

Huensch, A. (2024). Clarifying the role of inhibitory control in L2 phonological processing: A preregistered, close replication of Darcy et al. (2016). *Studies in Second Language Acquisition*, 1–21. https://doi.org/10.1017/S0272263124000238

Hyde, K., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, *15*, 356–360. https://doi.org/10.1111/j.0956-7976.2004.00683.x

Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2012). Can native Japanese listeners learn to differentiate /r–l/on the basis of F3 onset frequency? *Bilingualism: Language and Cognition*, *15*, 255–274. https://doi.org/10.1017/S1366728912000041

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267–3278. https://doi.org/10.1121/1.2062307

Jasmin, K., Dick, F., Holt, L. L., & Tierney, A. (2020). Tailored perception: Individuals' speech and music perception strategies fit their perceptual abilities. *Journal of Experimental Psychology: General*, *149*(5), 914–934. https://doi.org/10.1037/xge0000688

Jasmin, K., Sun, H., & Tierney, A. T. (2021). Effects of language experience on domain-general perceptual strategies. *Cognition*, *206*, 104481. https://doi.org/10.1016/j.cognition.2020.104481

Jongman, A., & Tremblay, A. (2020). Word prosody in L2. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 594–604). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198832232.013.34

Kalashnikova, M., Goswami, U., & Burnham, D. (2019). Sensitivity to amplitude envelope rise time in infancy and vocabulary development at 3 years: A significant relationship. *Developmental Science*, *22*(6), Article e12836. https://doi.org/10.1111/desc.12836

Kang, O., Rubin, D. O. N., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, *94*, 554–566. https://doi.org/10.1111/j.1540-4781.2010.01091.x

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, *128*(6), 3757–3768. https://doi.org/10.1121/1.3506351

Lenth, R. V. (2025). *emmeans package* (R package; Version 1.10.7) [Computer software]. https://cran.r-project.org/package=emmeans

Levitt, H. C. C. H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*, 467–477. https://doi.org/10.1121/1.1912375

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, *32*, 451–454. https://doi.org/10.1121/1.1908095

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., Smith, B. K., Bunting, M. F., & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*(3), 530–566. https://doi.org/10.1111/lang.12011

Liu, F., Patel, A., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain*, *133*, 1682–1693. https://doi.org/10.1093/brain/awq089

Major, R. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410604293

McArthur, G. M., & Bishop, D. V. (2005). Speech and non-speech processing in people with specific language impairment: A behavioural and electrophysiological study. *Brain and Language*, *94*, 260–273. https://doi.org/10.1016/j.bandl.2005.01.002

McLaughlin, B. (1990). Restructuring. *Applied Linguistics*, *11*, 113–128. https://doi.org/10.1093/applin/11.2.113

McMurray, B., Baxelbaum, K. S., Colby, S., & Tomblin, J. B. (2023). Understanding language processing in variable populations on their own terms: Towards a functionalist psycholinguistics of individual differences, development, and disorders. *Applied Psycholinguistics*, *44*(4), 565–592. https://doi.org/10.1017/S0142716423000255

McWeeny, S., & Norton, E. S. (2024). Auditory processing and reading disability: A systematic review and meta-analysis. *Scientific Studies of Reading*, *28*(2), 167–189. https://doi.org/10.1080/10888438.2022.2044371

Mora-Plaza, I., Saito, K., Suzukida, Y., Dewaele, J-M., & Tierney, A. (2022). *Tools for second language speech research and teaching*. http://sla-speech-tools.com. http://doi.org/10.17616/R31NJNAX

Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences*, *109*(39), 15953–15958. https://doi.org/10.1073/pnas.1204319109

Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, *58*(3), 479–502. https://doi.org/10.1111/j.1467-9922.2008.00446.x

Munro, M. J., Derwing, T. M., & Saito, K. (2024). Aspiring to aspirate: L2 acquisition of English word-initial /p/over 10 years. *Language and Speech*. Advance online publication. https://doi.org/10.1177/00238309241264296
Neijenhuis, K., Campbell, N. G., Cromb, M., Luinge, M. R., Moore, D. R., Rosen, S., & de Wit, E. (2019). An evidence-based perspective on "misconceptions" regarding pediatric auditory processing disorder. *Frontiers in Neurology*, *10*, Article 287. https://doi.org/10.3389/fneur.2019.00287
Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573
Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*(1), 58–75. https://doi.org/10.1196/annals.1284.006
Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472. https://doi.org/10.1121/1.3593366
Potthoff, R. F. (1964). On the Johnson-Neyman technique and some extensions thereof. *Psychometrika*, *29*, 241–256. https://doi.org/10.1007/BF02289721
R Core Team. (2023). *R: A language and environment for statistical computing* (Version 4.3.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/
Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything? *Journal of Phonetics*, *31*(3–4), 509–527. https://doi.org/10.1016/S0095-4470(03)00046-9
Ruan, Y., & Saito, K. (2023). Less precise auditory processing limits instructed L2 speech learning: Communicative focus on phonetic form revisited. *System*, *114*, Article 103020. https://doi.org/10.1016/j.system.2023.103020
Saito, K. (2013). Age effects on late bilingualism: The production development of /ɹ/by high-proficiency Japanese learners of English. *Journal of Memory and Language*, *69*, 546–562. https://doi.org/10.1016/j.jml.2013.07.003
Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning*, *65*(3), 563–595. https://doi.org/10.1111/lang.12120
Saito, K., Hanzawa, K., Petrova, K., Kachlicka, M., Suzukida, Y., & Tierney, A. (2022). Incidental and multimodal high variability phonetic training: Potential, limits, and future directions. *Language Learning*, *72*(4), 1049–1091. https://doi.org/10.1111/lang.12503
Saito, K., Kachlicka, M., Suzukida, Y., Mora-Plaza, I., Ruan, Y., & Tierney, A. (2024). Auditory processing as perceptual, cognitive, and motoric abilities underlying

successful second language acquisition: Interaction model. *Journal of Experimental Psychology: Human Perception and Performance*, *50*(1), 119. https://doi.org/10.1037/xhp0001166

Saito, K., & Munro, M. J. (2014). The early phase of /ɹ/production development in adult Japanese learners of English. *Language and Speech*, *57*(4), 451–469. https://doi.org/10.1177/00238309135132

Saito, K., Petrova, K., Suzukida, Y., Kachlicka, M., & Tierney, A. (2022). Training auditory processing promotes second language speech acquisition. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(12), 1410. https://doi.org/10.1037/xhp0001042

Saito, K., Sun, H., Kachlicka, M., Alayo, J. R. C., Nakata, T., & Tierney, A. (2022). Domain-general auditory processing explains multiple dimensions of L2 acquisition in adulthood. *Studies in Second Language Acquisition*, *44*(1), 57–86. https://doi.org/10.1017/S0272263120000490

Saito, K., Sun, H., & Tierney, A. (2020). Domain-general auditory processing determines success in second language pronunciation learning in adulthood: A longitudinal study. *Applied Psycholinguistics*, *41*(5), 1083–1112. https://doi.org/10.1017/S0142716420000491

Saito, K., Suzukida, Y., Tran, M., & Tierney, A. (2021). Domain-general auditory processing partially explains second language speech learning in classroom settings: A review and generalization study. https://doi.org/10.1111/lang.12447

Saito, K., & Tierney, A. (2024). Domain-general auditory processing as a conceptual and measurement framework for second language speech learning aptitude: A test-retest reliability study. *Studies in Second Language Acquisition*, *46*(4), 1206–1230. https://doi.org/10.1017/S027226312200047X

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*(4), 439–462. https://doi.org/10.1093/applin/amv047

Silbert, N. H., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., & Tare, M. (2015). Non-native phonemic discrimination, phonological short term memory, and word learning. *Journal of Phonetics*, *50*, 99–119. https://doi.org/10.1016/j.wocn.2015.03.002

Smith, J. (2007). *Introduction to digital filters with audio applications*. W3K Publishing.

Snowling, M. J. (2013). Early identification and interventions for dyslexia: A contemporary view. *Journal of Research in Special Educational Needs*, *13*(1), 7–14. https://doi.org/10.1111/j.1471-3802.2012.01262.x

Snowling, M. J., Gooch, D., McArthur, G., & Hulme, C. (2018). Language skills, but not frequency discrimination, predict reading skills in children at risk of dyslexia. *Psychological Science*, *29*(8), 1270–1282. https://doi.org/10.1177/0956797618763090

Sun, H., Saito, K., & Tierney, A. (2021). A longitudinal investigation of explicit and implicit auditory processing in L2 segmental and suprasegmental acquisition. *Studies in Second Language Acquisition*, *43*(3), 551–573. https://doi.org/10.1017/S0272263120000481

Surprenant, A. M., & Watson, C. S. (2001). Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *The Journal of the Acoustical Society of America*, *110*(4), 2085–2095. https://doi.org/10.1121/1.1396320

Suzuki, Y. (Ed.). (2023). *Practice and automatization in second language research: Perspectives from skill acquisition theory and cognitive psychology*. Routledge: Taylor & Francis.

Suzukida, Y., & Saito, K. (2023). Detangling experiential, cognitive, and sociopsychological individual differences in second language speech learning: Cross-sectional and longitudinal investigations. *Bilingualism: Language and Cognition*, *26*(4), 762–775. https://doi.org/10.1017/S136672892200059X

Tallal, P., & Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends in Neurosciences*, *29*(7), 382–390. https://doi.org/10.1016/j.tins.2006.05.003

Tierney, A., & Kraus, N. (2014). Auditory-motor entrainment and phonological skills: Precise auditory timing hypothesis (PATH). *Frontiers in Human Neuroscience*, *8*, Article 949. https://doi.org/10.3389/fnhum.2014.00949

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 1–30. https://doi.org/10.1017/S0272263106060013

Varghese, L., Bharadwaj, H. M., & Shinn-Cunningham, B. G. (2015). Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. *Brain Research*, *1626*, 146–164. https://doi.org/10.1016/j.brainres.2015.08.042

Vuvan, D., Nunes-Silva, M., Pretz, I. (2015). Meta-analytic evidence for the non-modularity of pitch processing in congenital amusia. *Cortex*, *69*, 186–200. https://doi.org/10.1016/j.cortex.2015.05.002

Whiteford, K. L., & Oxenham, A. J. (2018). Learning for pitch and melody discrimination in congenital amusia. *Cortex*, *103*, 164–178. https://doi.org/10.1016/j.cortex.2018.03.001

Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*, 565–585. https://doi.org/10.1017/S0142716407070312

Xu, J., Saito, K., & Mora-Plaza, I. (in press). Task-based pronunciation teaching: Lack of auditory precision but not memory hinders learning. *System*.

Zhang, J., Meng, Y., Fan, X., Ortega-Llebaria, M., & Ieong, S. L. (2018). Stress typicality effect in Chinese advanced and intermediate ESL learners. *Educational Psychology*, *38*(5), 617–632. https://doi.org/10.1080/01443410.2017.1380169
Zhang, J. D., Susino, M., McPherson, G. E., & Schubert, E. (2020). The definition of a musician in music psychology: A literature review and the six-year rule. *Psychology of Music*, *48*(3), 389–409. https://doi.org/10.1177/0305735618804038
Zhang, W., Liao, Y., & Truong, H. T. (2024). High variability phonetic training facilitates categorical perception of Mandarin lexical tones in L2 older adults: A link to auditory processing. *Language Teaching Research*. https://doi.org/10.1177/136216882412822
Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics*, *38*(2), 260–271. https://doi.org/10.1016/j.wocn.2009.11.002
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

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Accessible Summary

Appendix S1. Second Language Speech Materials.