Less precise auditory processing limits instructed L2 speech learning: Communicative focus on phonetic form revisited

Yaoyao Ruan, Kazuya Saito [*]

University College London, United Kingdom

ABSTRACT

Drawing learners' attention to phonetic form during meaning-oriented activities (i.e., communicative focus on form) has been found to be an optimal method of L2 speech training. In the context of 55 Chinese speakers' English [ɪ] and [i] acquisition (e.g., "sit" vs. "seat"), the current study set out to examine the extent to which the outcomes of such instructed L2 speech learning are tied to individual differences in domain-general auditory processing (precisely representing the acoustic properties of sounds). All the participants engaged in 1.5 h of meaning-oriented instruction. The treatment was carefully manipulated to induce those in the experimental group to attend to the accurate distinction of English [ɪ] and [i]. According to the pre- and post-tests, the experimental group significantly enhanced their English [ɪ] and [i] identification accuracy. However, the follow-up analyses demonstrated that instructional effectiveness was observed especially among those with high levels of auditory processing, but *unclear* among those with low levels of auditory processing. The findings suggest that the source of individual variation in instructed L2 speech learning stems at least partially from domain-general auditory processing, and that auditory training could help facilitate language learning in some individuals with specific auditory difficulties.

Whereas there is a consensus that instruction facilitates the development of second language (L2) speech proficiency especially when it is delivered in communicatively authentic contexts, instructional effectiveness is subject to a great deal of individual variation. Many scholars have ascribed the source of the variation to learners' different levels of perceptual-cognitive aptitude. Although the majority of the existing aptitude literature has exclusively focused on *lexicogrammar* (for a comprehensive overview, Wen & Skehan, 2021), scholars have begun to examine the associations between outcomes of L2 *speech* learning and domain-general auditory processing (i.e., precise encoding of sounds; Saito & Tierney, 2023). The current investigation aims to examine (a) how a total of 55 college-level Chinese students enhanced their English [ɪ] and [i] perception accuracy (e.g., "sit" vs. "seat") while working on a range of meaning-oriented tasks; and (b) the extent to which such gains could be differentially associated with the participants' auditory processing profiles. Taken together, the study was the first attempt to disentangle the complex relationship between auditory processing and meaning-oriented L2 speech learning under instructed conditions.

---

* Corresponding author. University College London, Institute of Education, 20 Bedford Way, WC1H0AL, United Kingdom.
E-mail addresses: yaoyao.ruan.20@ucl.ac.uk (Y. Ruan), k.saito@ucl.ac.uk (K. Saito).

1. Background

1.1. Instructed L2 speech learning and individual differences

Over the past five decades, scholars have examined how instruction can help L2 learners develop their linguistic abilities in the most efficient and effective way. Although language-focused, decontextualized, and explicit instruction has remained dominant in many classrooms worldwide, the generalizability of such instructional gains to real-life contexts remains controversial (Spada & Tomita, 2010). Scholars have agreed on the importance of drawing learners' attention to linguistic accuracy, especially when they engage in meaning-oriented activities in classroom settings (i.e., communicative focus on form; for a comprehensive overview, see Ellis, 2016). From a theoretical standpoint, maintaining learners' attention on both meaning and form at the same time is believed to help L2 learners proceduralize and automatize their explicit and declarative knowledge in the long-term memory (DeKeyser, 2017) and then help transfer what they have learned from instruction to future communicative settings (Spada & Tomita, 2010).

To achieve this goal, there are two different types of focus-on-form techniques—(a) the proactive approach (i.e., creating tasks wherein learners are required to use certain linguistic structures accurately with a view of successful task completion) and (b) the reactive approach (i.e., providing corrective feedback in response to the occurrence of linguistic errors). While the existing literature has been mainly concerned with the impact of proactive and reactive focus-on-form on L2 lexicogrammar learning (see Li, 2010), a growing number of scholars have examined the applicability of these approaches to L2 phonological learning (e.g., Saito, 2021 for phonological recasts; Mora & Levkina, 2017 for task-based speech learning).

Overall, communicative focus on *phonetic* form significantly impacts various dimensions of L2 speech learning (Lee & Lyster, 2016 for perception; Saito & Lyster, 2012; Parlak & Ziegler, 2016 for controlled and spontaneous production; for a comprehensive overview, Saito, 2021). Much scholarly attention has also been directed towards examining how instructional effectiveness can be associated with a range of input-related variables, such as length of instruction (Munoz, 2014), focus of instruction (segmentals vs. prosody; Derwing, Munro, & Wiebe, 1998), type of instruction (form-vs. meaning-oriented; Saito, 2012; intentional vs. incidental; Saito, Hanzawa, et al., 2022), and mode of instruction (face-to-face vs. computer-mediated; Parlak & Ziegler, 2016). At the same time, there is evidence that even when individuals engage in the same type of instruction for the same amount of time, their L2 outcomes often differ, with some achieving advanced L2 proficiency and others demonstrating much difficulty (Saito, 2017). This is arguably because individuals differ in the extent to which they can notice, elaborate, and make the most of every input opportunity; such relevant perceptual and cognitive abilities are generally termed as "aptitude". Whereas a range of conceptual frameworks have been proposed for L2 lexicogrammar learning aptitude (e.g., Carroll & Sapon, 1959 for MLAT; Linck et al., 2013 for Hi-Lab; for a comprehensive review, see Wen & Skehan, 2021), very few studies have worked on the role of aptitude factors in post-pubertal L2 speech learning.

1.2. Auditory processing and L2 speech learning

Recently, scholars have begun to propose auditory processing as a key aptitude which drives language learning throughout the lifespan (Mueller, Friederici, & Männel, 2012). Auditory processing is a basic, lower-order, perceptual ability to represent the spectral and temporal characteristics of sounds. Since learners chiefly rely on auditory input at every stage of linguistic processing (e.g., formant, pitch and duration analyses for the development of phonetic, phonemic, and prosodic categories; pitch and duration analyses for the detection of word and sentence stress and boundaries), auditory processing is considered a bottleneck for language learning (Goswami, 2015). This ability is believed to anchor a range of other perceptual phenomenon beyond speech where acoustic signals are accessed, orchestrated, and weighted in many different ways (e.g., music, emotion, and environmental sounds). This aligns with an influential view in cognitive psychology which states that auditory processing can be considered domain-general and this ability forms the basis of multiple domain-specific phenomena.

In the context of first language (L1) acquisition, there is ample evidence showing that those with language disorders likely have auditory deficits (e.g., Hämäläinen, Salminen, & Leppänen, 2012 for dyslexia) and that those with low levels of auditory processing likely experience slower language learning (e.g., Kalashnikova, Goswami, & Burnham, 2019 for longitudinal evidence on the link between auditory processing and literacy development within the first three years of life). Based on these findings, some researchers have suggested that auditory processing could be used as a diagnostic tool for certain types of language delay or impairment, such as dyslexia (Hornickel & Kraus, 2013).

However, whether auditory processing plays a significant role in *ultimate* L1 attainment remains controversial given that some infants with auditory deficits have been found to demonstrate normal language proficiency levels in the long run (Rosen & Manganari, 2001). This is arguably because auditory problems can be compensated for through the use of other cognitive abilities (e.g., attention, memory, and control; McArthur & Bishop, 2005) and/or because dimension-specific auditory difficulties can be compensated through the use of other perceptual cues (Jasmin, Dick, Holt, & Tierney, 2020). Thus, there is a stronger consensus that auditory precision influences the *rate/speed* of acquisition, especially when learners begin encoding and parsing sounds, words, and sentences in the initial stages of language learning.

To test the generalizability of the hypothesis, certain scholars have examined the generalizability of the topic to the context of post-pubertal L2 speech learning (Mueller et al., 2012). Overall, a medium-to-strong relationship between auditory processing and L2 speech acquisition has been reported in the literature (Saito, 2023; Saito & Tierney, 2023). A growing amount of cross-sectional investigation has shown that those who have attained a high level of L2 speech proficiency likely have both ample immersion experience and precise auditory processing (e.g., Kachlicka, Saito, & Tierney, 2019). Longitudinal investigations have demonstrated that those with higher levels of auditory processing tend to yield more improvement when they engage in immersion experience (e.g.,

Sun, Saito, & Tierney, 2021). Notably, the existing literature has exclusively drawn on observational data of naturalistic L2 learners. Little is known about how auditory processing can impact instructed L2 speech learning, which is the main focus of the current study.

2. Motivation for the current study

Communicative focus on form has been suggested as an optimal method of L2 instruction (Ellis, 2016), and found to help post-pubertal learners enhance L2 lexicogrammar *and* phonological learning (Saito & Lyster, 2012). Notably, the magnitude and rate of improvement vary to a great degree. Given that auditory processing has been proposed as an anchor of language learning throughout the lifespan (Goswami, 2015) and germane to post-pubertal L2 (Kachlicka et al., 2019), we argued that one source of variation in the focus-on-form instruction could be learners' aptitude related to this ability. To date, there is some emerging evidence that auditory processing may mediate the influence of explicit phonetic training (e.g., Chandrasekaran, Sampath, & Wong, 2010 for intensive exposure to target sounds). More importantly, as in the discussion in L1 acquisition (e.g., Hornickel & Kraus, 2013), some scholars have further pointed out that having auditory processing difficulties overrides the benefits of explicit phonetic training (Perrachione, Lee, Ha, & Wong, 2011).

Examining the aptitude-treatment interaction is a crucial initiative with ample practical and theoretical relevance as the findings would identify not only high-aptitude learners who can make the most of instruction, but also low-aptitude learners who may fail to benefit from instruction. Such findings will help us conceptualize how to provide remedial treatment (e.g., auditory training) to those with perceptual and cognitive difficulties so that all L2 students can *equally* benefit from instruction and learn/master a target language regardless of aptitude profiles. However, no empirical studies have ever explored how those with varied auditory processing abilities can differentially benefit from more communicatively-authentic, meaning-oriented instruction (i.e., communicative focus on form). The current study set out to address the following research questions.

1. To what extent can communicative focus-on-form instruction help Chinese learners of English improve their L2 vowel acquisition (English [i] and [ɪ])?
2. To what extent do individual differences in domain-general auditory processing mediate learning outcomes?

In terms of RQ1, following the existing literature (e.g., Saito & Lyster, 2012), we predicted that Chinese learners would significantly enhance their English [i] and [ɪ] accuracy owing to focus-on-form instruction. As for RQ2, there were two possible predictions. First, the aptitude-treatment interaction could be linear (i.e., the amount of instructional effectiveness is tied to learners' aptitude scores; Skehan, 2016). Alternatively, the aptitude-acquisition link could be dichotomous (i.e., the benefit of instruction could be limited among those with low-aptitude; Perrachione et al., 2011).

3. Method

3.1. Design

This study employed a quasi-experimental pre- and post-test design, with a total of 55 participants randomly assigned to the experimental group ($n = 39$) or the control group ($n = 16$). We intentionally recruited more participants for the experimental group. We later divided them into subgroup conditions (high vs. low auditory processing) and aimed to examine how their aptitude profiles could lead to different instructional gains. Due to the pandemic, all of the experiments had to be moved to Zoom. However, we made a lot of effort to ensure the quality of the training. The participants were divided into smaller subgroups of two to five. During each session, the instructor gave each participant individual attention and provided recasts whenever they made vowel errors. The setup could be considered more like tutoring than a traditional teacher-fronted classroom.
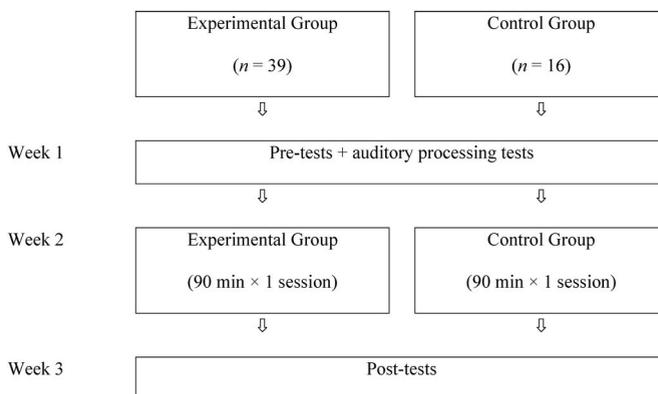
Fig. 1. Summary of research design.

Both groups (Experimental, Control) received 1.5 h of meaning-oriented instruction on the development of English argumentative skills. In the experimental group, the tasks were designed to induce learners to notice and practice the English [i] and [ɪ] contrast (difficult phonological features for Chinese learners; for details, see below) while learning how to construct logical arguments in English. An instructor provided recasts in response to the occurrence of pronunciation errors in English [i] and [ɪ]. There was no focus-on-form component in the control treatment. In pre- and post-tests, participants took a forced-choice identification task in order to evaluate the impact of instruction on participants' English [i] and [ɪ] perception abilities under trained vs. untrained lexical conditions. Such perception abilities were assessed via a forced-choice identification task. The design of the study was visually summarized in Fig. 1.

3.2. Participants

Participants of this study were young adult Chinese English-as-a-Foreign-Language students at a university in China. They were native speakers of Mandarin Chinese. During the participant recruitment phase, ads were distributed to non-language and non-linguistics major undergraduate students within the university online by the researcher or in-person by their English teacher. Interested participants contacted the researcher to sign the consent form and set up a date for the pre-test via a social media application. A total of 58 volunteers initially participated in the current study, and three of them withdrew from the project for personal reasons. The mean age of the whole sample ($N = 55$) was 19.1 years old, ranging from 18 to 22. Forty-three learners had taken the IELTS test in the past two years, with an average overall score of 6.3, ranging from 5.5 to 7, and an average score of 5.7 for speaking. According to the descriptors of the levels of Common European Framework of Reference for Languages, the learners' English proficiency at the time of the current study was considered to be B2. None of the participants reported pronunciation training and more than one month of immersion experience in an English-speaking country. Given that the participants were in China, the opportunities to communicate with other native and non-native speakers of English were highly limited. After the pre-tests, the participants were randomly assigned to the experimental ($n = 39$) or control ($n = 16$) group. As we described in the Results section, those in the experimental group were further divided into two subgroups corresponding to different levels of their auditory processing ($n = 20$ for high audition; $n = 19$ for low audition).

3.3. Instructional treatment

Instructor. The instructor for both groups (the first author) was a native Mandarin speaker with advanced L2 English proficiency who was an author of this paper. The instructor had extensive English teaching experience in China (more than 5 years) with an MA degree in TESOL.

Target of Instruction. The tense vowel [i] (as in "seat") and the lax vowel [ɪ] (as in "sit") in British English were the target sounds in the current study. The former sound can be pronounced with less centralization (more front), longer duration, and clear lip spreading; but the latter sound with centralization, shorter duration, and lip neutralization. Thus, the perception of English [i] and [ɪ] can be achieved by orchestrating both temporal and spectral information (longer, lower F1, and higher F2 for English [i] vs. shorter, higher F1, and lower F2; (Hillenbrand, Getty, Clark, & Wheeler, 1995). In general, native speakers tend to use first and second formant values (Hz) as a primary cue especially when it comes to General American (wherein the durational differences can be neutralized due to the influence of individual speakers' speech rate; Gottfried, Miller, Payton, 1990).

At the same time, native listeners do rely on the phonemic length (ms) as a secondary information especially when the primary cue is unavailable or unclear in synthetic speech (Hillenbrand, Clark, & Houde, 2000). The durational differences can be more clearly observed in Received Pronunciation (the main focus of the study; Ladefoged & Johnson, 2014). In the International Phonetic Alphabet guidelines, for example, the high front vocalic contrast [i] and [ɪ] in Received Pronunciation features the long-short distinction, i.e., [i:] and [ɪ]; but such vowel length contrast is not distinctive in General American, i.e., [i] and [ɪ].

Extensive studies have found that L1 Mandarin learners of English have difficulty distinguishing between [i] and [ɪ] in English and

Table 1
Summary of 42 target words (26 Trained words, 16 untrained words).

Trained Items	Untrained Items
1. bean-bin	1. feet-fit
2. beat-bit	2. keys-kiss
3. cheap-chip	3. lead-lid
4. feel-fill	4. least-list
5. heel-hill	5. meal-mill
6. heat-hit	6. peak-pick
7. leave-live	7. sleep-slip
8. reach-rich	8. steel-still
9. read-rid	
10. scene-sin	
11. seat-sit	
12. seek-sick	
13. sheep-ship	

tend to categorize both vowels as [i] (e.g., Wang & Munro, 2004). There is some evidence that many L2 speakers overly rely on duration (phonemic length) rather than spectral cues (first and second formants; Flege, Bohn, & Jang, 1997). This could be a problem especially in the case of General American ([i] and [ɪ]) but less so Received Pronunciation ([i:] and [ɪ]). According to the perceptual assimilation model (Best & Tyler, 2007), given that only English [i] exists in the Mandarin vowel inventory, L1 Mandarin learners of English initially perceive English [i] and [ɪ] as equivalents of the Mandarin [i]. As a function of increased L2 experience, Mandarin listeners start classifying English [i] as the Mandarin [i], and then create a new category for English [ɪ] (Category Goodness Assimilation).

Importantly, the mispronunciation of English [i] and [ɪ] is considered highly problematic with a view of successful communication. According to the functional load principle (Brown, 1988), the contrast has a number of minimal pairs. The presence of phonological errors in English [i] and [ɪ] may greatly hinder listeners' global understanding of L2 speech (Munro & Derwing, 2006; Suzukida & Saito, 2021).

Treatment. As for the experimental group, the focus-on-form components (focused tasks, recasts) were embedded in the 1.5-h meaning-oriented lesson on English argumentation skills. All the materials were originally developed in Saito (2013) where a range of meaning-oriented activities were created to guide Japanese listeners to learn English [r] and [l]. These activities were revised and adjusted to Chinese listeners' English [i] and [ɪ] acquisition. A total of 26 minimal pairs differing only in [i] and [ɪ] (see "trained items" in Table 1) were embedded in the two activities: (a) awareness raising task ($n = 16$ words) and (b) debate task ($n = 10$ words).

First, participants engaged in the awareness raising task (10 min). To help students who may have lacked sufficient phonetic knowledge in the target of instruction (English [i] and [ɪ]), they received metalinguistic instruction about how the vowel contrast differs in terms of tongue positions and phonemic length. Then, they proceeded to the debate task (80 min). The main objective of the task was to help students to develop critical thinking, express opinions, and provide counterarguments via a range of debate activities in English. All the topics were on current phenomena or issues familiar to Chinese university students (e.g., "*We should _seek_ help from doctors immediately when we feel _sick_ because it is better for us*"). To promote the noticeability of the target words, they were highlighted in red (i.e., typographically enhanced input). Throughout the activities, the participants were encouraged to pay attention to the differences between English [i] and [ɪ] with their primary focus remaining on meaning (i.e., English argumentative skills). For the details of the activities and the topics that were discussed and debated in class, see **Supporting Information A.**

Finally, following the procedure in Saito (2013), the instructor provided recasts when participants mispronounced English [i] and/ or [ɪ]. Recasts are believed to provide both positive evidence (the presentation of model forms) and negative evidence (the indication of errors), both of which play a key role in L2 speech acquisition (Goo & Mackey, 2013). To increase the saliency of phonological recasts, the instructor recasted only target words (i.e., partial recasts). For examples, see below and **Supporting Information B.**

Episode 1
Student: People shouldn't seek doctor's help immediately when they feel seek* [sick].
Teacher: Sick. (RECAST)
Student: Sick. (REPAIR)
(Recast stops.)

As for the control group, the 90-min session also comprised the same content (i.e., English argumentation skills). However, all the target words were carefully replaced with substitute words (e.g., "We should consult a doctor as soon as we feel unwell because it is better for us."). The instructor provided feedback on the content of participants' speech. We did not video-record each instructional session. In retrospect, the instructor (an author of this project) provided recasts whenever the vowel errors were identified. The number of recasts may have greatly varied between participants (5–20 times).

As all the participants were based in China (wherein their English learning opportunities were exclusively limited to grammar-based instruction in classrooms), no participants reported any exposure to extra pronunciation training (featuring English [i] and [ɪ]) between pre- and post-tests.

3.4. Measures of L2 speech development

Following Flege's speech learning model (Flege, 1995), we assumed that L2 phonetic representations are auditory-based (i.e., characterized by acoustic information such as formants, pitch, and phonemic duration) and learning initially takes place perceptually before transferring to production. To capture how instruction impacted participants' representations and behaviors for the target phonetic structures (English [i] and [ɪ]), a forced-choice identification task was adopted for the pre- and post-tests.

Materials. The testing materials included both trained and untrained items. The untrained words were used to check the extent to which the effects of instruction can be generalized to novel words. All of the lexical items ($n = 42$ target tokens) in the testing materials are Consonant-Vowel-Consonant (CVC) single words (see Table 1). They all fall within the first 4,000 most frequent word families according to vocabulary profiling via Lextutor (Cobb, 2012). In an investigation into English vocabulary size across CEFR levels, Milton (2010) stated that around 3,000 of the most frequently occurring words might be required to achieve basic proficiency levels (i.e., A1 and A2), and advanced CEFR levels (i.e., C1 and C2) were associated with full recognition of the 5,000 most frequently used words. Thus, for intermediate-level learners (CEFR B1 and B2) in the current study, the impacts of lexical frequency and familiarity on test performance were minimized.

All speech samples were recorded by one male and one female professional voice-over artist, who were both native speakers of Standard British English, in isolated studios with their own professional recording equipment. Each sample was digitalized at a 44,100

Hz sampling rate and normalized for peak intensity.

Procedure. After they received instruction from the investigator, the participants used their own computer with a headset to access the test materials available in the online platform *Gorilla* (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020). They listened to a total of 84 stimuli in a randomized order (42 trained and untrained target words read by two native speakers). For each stimulus, they identified which word they heard by clicking one of the two orthographic options presented on the computer screen (see Fig. 2). Their accuracy scores were recorded on a 100-point scale.

3.5. Measures of auditory processing

As in the previous literature (e.g., Kachlicka et al., 2019), three different AXB discrimination tasks were adopted to assess three different dimensions of participants' auditory sensitivities, i.e., perception of fundamental frequencies (pitch), formants, and duration. To reflect the domain generality of this ability, the tasks in the current study (and elsewhere) adopted a number of synthesized, nonverbal stimuli with very simple acoustic characteristics (e.g., completely flat fundamental frequencies, formant contours, and harmonic spectrum). Normal hearing listeners will not perceive the stimuli as speech (for the results of listener judgments of the nonverbal stimuli, see Saito, Kachlicka, et al., 2022; for a methodological summary of the auditory processing tests, Saito & Tierney, 2023).[1]

In each trial, participants listened to three non-verbal sounds and were required to indicate which one (either the first or third) sounded different from the other two by clicking on the number "1" or "3" on the screen with a mouse (see Fig. 3). Since each sound differed in only one acoustic dimension (pitch, formants or duration), each task was designed to test how small of a difference participants could hear in the target parameter.

For each test, one reference stimulus and 100 target stimuli were created using custom MATLAB scripts. The stimuli for pitch and duration discrimination were 250-ms-long complex tones of four harmonics with a fundamental frequency (F0) of 330 Hz. For pitch discrimination, the reference stimulus was created with an F0 of 330 Hz, whereas the target stimuli varied in frequency from 330.3 to 360 Hz with a 0.3-Hz increment. For duration discrimination, the length of the reference stimulus was 250 ms. The target stimuli ranged from 252.5 ms to 500 ms (step size of 2.5 ms). The stimuli for formant discrimination were complex tones with an F0 of 100 Hz and three formants (F1 = 500 Hz, F2 = 1500–1700, F3 = 2500 Hz). Whereas the F2 of the reference stimulus was set at 1500 Hz, the F2 of the target stimuli ranged from 1502 Hz to 1700 Hz (step size of 2 Hz).

Using Levitt's (1971) adaptive up-down procedure, the test began at Level 50 (out of 100 levels) and automatically adjusted the difficulty level in response to the participant's performance: it became 10 steps more difficult when the participant provided three correct responses in a row, or 10 steps easier when the participant gave one incorrect response. The step size decreased to five after the first reversal, from five to two following the second, and from two to one after the third until the task was completed. Each test terminated after 70 trials or 8 reversals (see Kachlicka et al., 2019 for more details of the method). Participants' final scores were recorded on a 100-point scale. Smaller values indicated participants had more precise auditory processing for formants, pitch, and duration.

To check the construct of the auditory processing test scores, participants' individual subtest scores (formants, pitch, and duration) were submitted to an exploratory factor analysis with Varimax rotation and an eigen value set to 1.0. The results showed that the three subtests were clustered as one single factor (explaining 75.5% of the variances), suggesting that the three subtests may have tapped into one single aspect of participants' auditory processing ability. For the purpose of the statistical analyses, we did not use subtest scores to avoid multicollinearity problems. Instead, their subtest scores were standardized and averaged to generate overall auditory processing scores per participant.

All the training and assessment materials are deposited in open science platform for future replicaiton, L2 Speech Tools (Mora-Plaza, Saito, Suzukida, Dewaele, & Tierney, 2022)

4. Results

4.1. Overall improvement (Pre-to Post-Tests)

The pre- and post-test correct identifications scores (%) were summarized in Table 2 and visually plotted in Fig. 4 as per three different conditions: overall ($n = 42$ items), trained ($n = 26$ items), and untrained ($n = 16$ items). The results of a Kolmogorov-Smirnov test showed that participants' pre-test scores did not significantly differ from normal distribution to trained conditions ($D = 0.100$, $p = .603$) and untrained conditions ($D = 0.081$, $p = .835$). To find any pre-existing difference in perceptual accuracy of the target sounds (i. e., English [i] and [ɪ]), participants' total pre-test scores were submitted to a one-way analysis of variance (ANOVA) with Group (Experimental vs. Control) as the between-group factor. No significant Group difference was found at the time of the pre-test ($p > .05$). The results suggest that the experimental and control participants' vowel performance were comparable prior to the treatment.

---

[1] A reviewer problematized the domain-generality of the auditory processing tests as they included formants which are crucial acoustic properties in human speech sounds. While this is true, formats can also be observed in many other acoustic and perceptual phenomena (beyond speech). We need to point out that formants can be defined as a broad peak in the spectrum and generated where there is acoustic resonance. Thus, formants can occur when sounds travel through not only a vocal tract (speech) but also musical instruments (e.g., violin, flute) and a room (as per its particular shapes of walls).

Fig. 2. A screenshot of the forced-choice identification test (English [i] and [ɪ] accuracy).

Fig. 3. A screenshot of auditory processing discrimination tasks.

Table 2
Summary of L2 speech and auditory processing scores.

	Experimental ($n = 39$)				Control ($n = 16$)			
	M	*SD*	95% CI		*M*	*SD*	95% CI	
			Lower	Upper			Lower	Upper
A. L2 vowel identification (%)								
Overall (pre)	70.2	14	65.7	74.6	73.9	12	67.3	80.5
Trained (pre)	68.3	15	63.6	73.1	71.3	14	63.9	78.6
Untrained (pre)	73.2	14	68.7	77.6	78.1	12	72.0	84.2
Overall (post)	74.6	16	70.0	80.0	72.5	15	64.7	80.2
Trained (post)	73.9	16	68.9	78.9	71.2	15	63.2	79.1
Untrained (post)	75.6	17	70.2	81.1	74.6	16	66.1	83.2
B. Aptitude factors								
Auditory processing (z scores)[a]	.02	.68	−.18	.22	−.05	.56	−.35	.24

Note. [a] lower values for more precise auditory processing.

In order to assess the effects of Time (pre-/post-tests), Group (Experimental vs. Control) and Lexis (trained vs. untrained), repeated-measures ANOVAs were used. A three-way ANOVA with Group as a between-group factor and Time and Lexis as within-group factors yielded a significant Group × Time interaction effect, $F(1, 53) = 6.185$, $p = .016$, $\eta_p^2 = 0.105$. The analyses of multiple comparisons showed that the control group's performance did not reach statistical significance ($M = 73.9 = 72.5\%$, $p = .363$, $\eta_p^2 = 0.016$). This suggests that there was no test-retest effect in the current study. Regarding the experimental group, their overall scores demonstrated a significant improvement over time ($M = 70.2 \rightarrow 74.6\%$, $p = .002$, $\eta_p^2 = 0.184$). In line with Cohen's (1988) benchmarks, the effect size could be considered medium to large ($\eta_p^2 = 0.13$-0.26). Yet, The three-way Group × Time × Lexis interaction did not reach statistical significance, $F(1, 53) = 0.008$, $p = .927$, $\eta_p^2 < 0.001$. Interestingly, a significant main effect for Lexis was found, $F(1, 53) = 20.883$, $p < .001$, $\eta_p^2 = 0.283$. According to the pairwise comparisons, learners in both experimental and control groups performed significantly better on untrained lexical items than trained items at both testing time points (Trained $M = 71.2\%$, $SD = 14.0\%$; Untrained $M = 75.0\%$, $SD = 14.0\%$).

Taken together, the results indicate (a) that the experimental group demonstrated a significant improvement in their abilities to identify English [i] and [ɪ] regardless of lexical context (trained and untrained); and (b) that the gains here were not due to test-retest
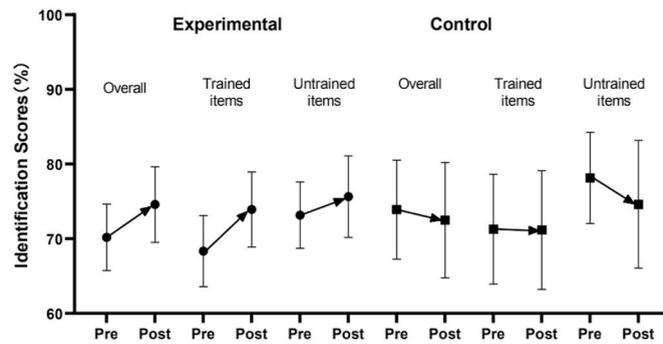
Fig. 4. 95% confidence intervals and mean values of the learners' perception scores.

effects.

4.2. Auditory processing and L2 speech learning

The next objective of the statistical analyses was to further examine the extent to which the improvement of the experimental group could be tied to participants' auditory processing scores. According to the results of the normality tests (Kolmogorov-Smirnov), auditory processing scores were comparable to normal distribution ($D = .101$, $p = .780$). Following the analyses of the aptitude-treatment interaction in previous studies, we conducted two different types of analyses.

Variance-Based Analyses. Given that each learner differed in her auditory processing ability to a great degree, we tested an assumption that learners' gains were correlated with their individual auditory profiles. To this end, a partial correlation analysis was conducted to check how participants' gain scores (post minus pre-test scores) were related to their auditory processing scores as per two group conditions (Experimental, Control), when their pre-test performance was controlled for (an alpha set to .025 via Bonferroni corrections). The correlation coefficients did not reach significance for Experimental, $r = -0.147$, $p = .379$, and Control, $r = 0.054$, $p = .850$.[2]

Mean-Based Analyses. Some scholars argued that whereas adult L2 learners can learn new sounds, the improvement could be limited to certain individuals with auditory difficulties (Perrachione et al., 2011). To this end, we conducted a group analysis to examine whether those with high vs. low-level audition differentially benefited from instruction (for a similar analysis decision [the median-split approach], see Chandrasekaran et al., 2010). The participants in the experimental group were further divided into two subgroups, high-vs. low-audition, by using the group's median value as a cut-off point. A total of 39 participants in the experimental group were divided into two subgroups, high-audition ($n = 20$; $M = -0.43$, $SD = 0.28$, $Range = -0.96$ to -0.01) vs. low-audition ($n = 19$; $M = 0.50$, $SD = 0.49$, $Range = 0.06$ to 2.00).

The participants' vowel perception scores were submitted to a two-way repeated-measures ANOVA with Group (High-Audition, Low-Audition, Control) as a between-group factor, and Time (pre, post) as a within-group factor. The analysis showed significant main effects of Time, $F(1, 52) = 5.209$, $p = .027$, $\eta_p^2 = 0.091$, but not Group, $F(1, 52) = 2.322$, $p = .108$, $\eta_p^2 = 0.082$. The Group × Time interaction effects reached statistical significance, $F(2, 52) = 4.311$, $p = .018$, $\eta_p^2 = 0.143$.

According to the results of post-hoc multiple comparison analyses (an alpha set to 0.017 via Bonferroni corrections), the three groups' vowel proficiency was comparable at the time of the pre-tests ($p = .092$ to .965). After the participants engaged in the treatment, however, the high-audition participants significantly enhanced their accuracy scores over time with large effects ($M = 73.7\% \rightarrow 80.0\%$, $F(1, 37) = 12.934$, $p = .001$, $\eta_p^2 = 0.237$). They also outperformed the low-audition participants in the post-tests at a marginally significant level ($M = 80.0\%$ vs. 68.9%, $F(1, 37) = 5.574$, $p = .024$, $\eta_p^2 = 0.131$). Both the low-audition and control participants' performance remained unchanged between pre- and post-tests ($p < .05$). For a visual summary, see Fig. 5 (group performance) and 6 (individual performance) (see Fig. 6).

5. Discussion

In the context of 55 Chinese speakers' English [i] and [ɪ] acquisition, the current study examined how the provision of communicative focus-on-form instruction can facilitate L2 speech learning, and how such instructional gains can be tied to domain-general auditory processing. The statistical analyses provided three primary findings. First, 1.5 h of focus-on-form instruction significantly helped improve L2 vowel perception with medium-to-large effects for both trained and untrained lexical items. Secondly, such instructional gains were observed especially among those with a high level of auditory processing (with large effects). Third, the benefits of focus-on-form was *limited* and *unclear* among those with low levels of auditory processing.

---

[2] The null results of the main analyses did not change whether we used the participants' overall auditory processing scores or their subtest scores (formant, pitch, and duration). This is because the subtest scores overlapped with each other to a great degree and may have represented one single aspect of auditory processing (shown in the exploratory factor analysis).
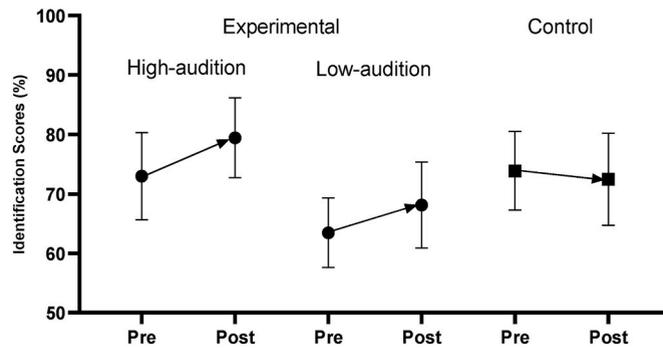
Fig. 5. 95% confidence intervals and mean values of the learners' perception scores as per auditory processing conditions relative to control group.
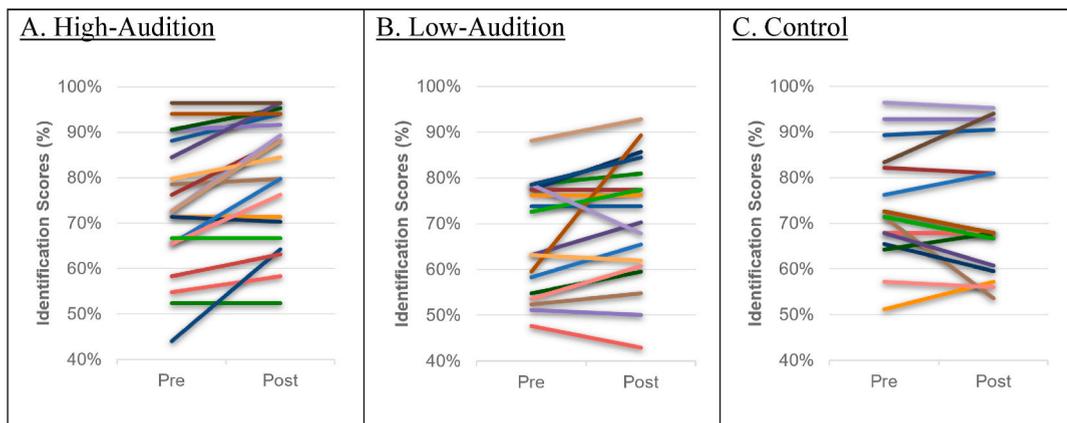

Fig. 6. Individual learner performance in high-audition, low-audition, and control groups.

The findings echo the previous literature that communicative focus-on-form can significantly facilitate L2 speech acquisition (e.g., Saito & Lyster, 2012) and that the effectiveness of instruction can be tied to participants' aptitude factors (e.g., Kissling, 2013). Extending the existing evidence on aptitude-treatment interaction effects (e.g., see Chandrasekaran et al., 2010 for the role of auditory processing in explicit L2 speech training), the current study further added that domain-general auditory processing mediates the effectiveness of more meaning-oriented, communicatively-authentic instruction (communicative focus on form). To date, aptitude has been conceptualized as the abilities *specific* to foreign language learning via explicit instruction (e.g., MLAT). Here, our findings suggest that such an aptitude framework should include domain-general auditory processing abilities in order to explain individual differences in L2 learning in various settings (form-oriented *and* meaning-oriented).

This argument is in line with emerging cross-sectional and longitudinal data indicating that both auditory processing and working memory are predictive of the rate of success in L2 speech learning in immersion settings where learners use, practice, and develop a target language for conversational purposes. For example, significant effects of aptitude relating to domain-general auditory processing among naturalistic, but not classroom L2 learners, have previously been reported (Saito, Suzukida, Tran, & Tierney, 2021; see Faretta-Stutenberg & Morgan-Short, 2018 for similar findings on working memory). The findings here shed some light on the idea that different types of aptitude can uniquely relate to different types/stages of L2 learning (Skehan, 2016). In the field of L2 grammar, there is some emerging evidence for the aptitude-treatment interaction: (a) Those with higher domain-*specific* analytic abilities (e.g., grammatical sensitivity, rote memorization) can benefit more from explicit instruction and correction and (b) those with implicit domain-*general* abilities (e.g., implicit and procedural memory) can demonstrate more gains from implicit, incidental instruction and correction (e.g., Yilmaz & Granena, 2021). To test this hypothesis, future studies can replicate the current study but using participants with varied aptitude profiles (e.g., auditory processing, working memory, implicit memory, grammatical sensitivity, rote memorization) and devising different types of L2 speech training methods (communicative focus on form vs. explicit training).

Having said that, however, the most important finding worthy of discussion is that the aptitude-instruction link was *dichotomous* rather than *linear*. Namely, the significant improvement resulting from focus-on-form was clear among those with high-audition but not among those with low-audition. The latter participants may have failed to make the most of the communicative focus-on-form treatment because they struggled to encode the acoustic properties of the target sounds, especially when their primary focus lay in conveying meaning. Given that all the target sounds are masked in communicative oriented input and thus appear to be non-salient, the low-aptitude learners may not notice, remember, and internalize L2 input in long-term representations (for the ambiguity of implicit and less-intrusive feedback, recasts, in L2 grammar learning, see Lyster & Ranta, 2013). The discussion here concurs with the

auditory deficit view in L1 (Goswami, 2015) and L2 (Perrachione et al., 2011) that individual differences in language learning success and difficulties stem at least in part from domain-general difficulties with auditory perception.

As Skehan (2016) argued, it is such low-aptitude L2 learners that may need remedial interventions (e.g., provision of explicit metalinguistic instruction and feedback). As such, low-aptitude L2 learners can be guided to not only notice but also understand and internalize linguistic targets which would otherwise be difficult to spot (Suzukida & Saito, 2023).

Another potential remedy particularly for those with auditory difficulties concerns the provision of auditory training which will in turn increase the efficacy of existing methods of L2 instruction. In auditory training, learners are induced to focus on perceiving particular acoustic characteristics of sounds (formants, pitch, and duration) via identification and discrimination of nonverbal synthesized sounds. According to L1 hearing research, for example, auditory processing can be remedied after a few hours of training especially when it is combined with language training (Micheyl, Delhommeau, Perrot, & Oxenham, 2006). In the context of L2 speech learning, there is emerging evidence that auditory processing training can enhance not only auditory processing but also phonetic performance (Saito, Petrova, Suzukida, Kachlicka, & Tierney, 2022).

Given the theoretical and practical relevance of the topic, future studies are strongly called for in order to further examine how provision of auditory training can help boost the impact of instructed L2 speech learning especially for those with low-level auditory processing, and ensure that all students can equally benefit from instruction and learn/master a second language regardless of aptitude profiles.

6. Future directions

Given that the current investigation took the first step towards examining aptitude effects in meaning-oriented L2 phonetic training (communicative focus on form), the study led to a number of future directions with a view of achieving a full-fledged picture of the mechanisms underlying successful instructed L2 speech learning.

- The AXB discrimination task was adopted to assess participants' auditory processing in the current study. Such aptitude abilities could be considered as relatively explicit as the outcome measure inevitably evokes their awareness. Future studies could include more incidental and implicit kinds of aptitude, such as procedural memory (Linck et al., 2013) and frequency following response (Saito et al., 2021).
- It has been shown that native listeners use the spectral cues as a primary acoustic correlate of English [i] and [ɪ] especially in General American. In contrast, the temporal information could be an unreliable cue as it is highly subject to the influence of individual speakers' speech rate (Gottfried et al., 1990). However, L2 listeners have been found to overly rely on duration rather than spectral cues to differentiate English [i] and [ɪ] (Flege et al., 1997). In the current study, participants were encouraged to attend to both spectral and temporal information both of which play an equally important role in Received Pronunciation, [i:] and [ɪ]. It would be interesting to see how much phonological recasts can affect L2 listeners' cue weighting patterns. To do so, however, future studies need to use synthetic (rather than natural) tokens in pre- and post-tests wherein spectral and temporal cues can be controlled for (Hillenbrand et al., 2000).
- The instruction targeted Chinese speakers' acquisition of English [i] and [ɪ]. The learning difficulty of which could be considered "medium" because their L1 phonetic inventories have at least one counterpart sound (Chinse [i]). As we discussed earlier, this particular instance is categorized as "Category Goodness Assimilation" within the perceptual assimilation model (Best & Tyler, 2007). However, existing aptitude literature has shown that aptitude matters especially when it comes to relatively difficult instances of L2 learning (e.g., Japanese speakers' English [r] and [l] acquisition; neither sound exists in their L1 system). According to the perceptual assimilation model, Japanese speakers' acquisition of English [r] and [l] is an example of Single Category acquisition; the two phones are perceived as poor exemplars of the Japanese alveolar tap [ɾ] (Single Category Assimilation). The perceptual assimilation model assumes the more learning difficulty for "Single Category Assimilation" than "Category Goodness Assimilation."
- The impact of instruction was examined at the perception level only. Future studies should further explore where the aptitude-acquisition link can be most clearly observed when a range of outcome measures are adopted (perception, controlled production, and spontaneous production).
- Although we argued that those with less precise auditory precision did not benefit from instruction, such findings were based on the relatively brief amount of training (1.5 h). There is a possibility that low-auditory learners can still demonstrate improvement if they engage in longer instruction.
- We conducted only one post-test one week after the instruction due to practical reasons. According to Spada and Tomita's (2010) synthesis of instructed L2 acquisition research, the test interval of one week could be considered a delayed post-test rather than an immediate one. However, we acknowledge the absence of a second post-test which future studies should include to examine the long-term sustainability of instructional effectiveness.

Author statement

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.system.2023.103020.

References

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

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn, & M. Munro (Eds.), *Second-language speech learning: The role of language experience in speech perception and production. A Festschrift in Honour of James E. Flege* (pp. 13–34). Amsterdam: John Benjamins.

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL quarterly, 22*(4), 593–606.

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test*.

Chandrasekaran, B., Sampath, P., & Wong, P. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America, 128*(1), 456–465. https://doi.org/10.1121/1.3445785

Cobb, T. (2012). *The compleat lexical tutor*. http://www.lextutor.ca/vp/.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

DeKeyser, R. M. (2017). Knowledge and skill in SLA. In S. Loewen, & M. Sato (Eds.), *The Routledge handbook of second language acquisition* (pp. 15–32). New York, NY: Routledge.

Derwing, T., Munro, M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*(3), 393–410. https://doi.org/10.1111/0023-8333.00047

Ellis, R. (2016). Focus on form: A critical review. *Language Teaching Research, 20*(3), 405–428. https://doi.org/10.1177/1362168816628627

Faretta-Stutenberg, M., & Morgan-Short, K. (2018). The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research, 34*(1), 67–101. https://doi.org/10.1177/0267658316684903

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Balti more: York Press.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics, 25*(4), 437–470.

Goo, J., & Mackey, A. (2013). The case against the case against recasts. *Studies in Second Language Acquisition, 35*(1), 127–165. https://doi.org/10.1017/s0272263112000708

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

Hämäläinen, J., Salminen, H., & Leppänen, P. (2012). Basic auditory processing deficits in dyslexia. *Journal of Learning Disabilities, 46*(5), 413–427. https://doi.org/10.1177/0022219411436213

Hillenbrand, J., Clark, M., & Houde, R. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America, 108*(6), 3013–3022. https://doi.org/10.1121/1.1323463

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America, 97*(5), 3099–3111.

Hornickel, J., & Kraus, N. (2013). Unstable representation of sound: A biological marker of dyslexia. *Journal of Neuroscience, 33*(8), 3500–3504. https://www.jneurosci.org/content/33/8/3500.short

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

Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and Language, 192*, 15–24. https://doi.org/10.1016/j.bandl.2019.02.004

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

Kissling, E. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *The Modern Language Journal, 97*(3), 720–744. https://doi.org/10.1111/j.1540-4781.2013.12029.x

Ladefoged, P., & Johnson, K. (2014). *A course in phonetics*. Cengage learning.

Lee, A. H., & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition, 38*(1), 35–64.

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

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*, 309–365. https://doi.org/10.1111/j.1467-9922.2010.00561.x

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

Lyster, R., & Ranta, L. (2013). Counterpoint piece: The case for variety in corrective feedback research. *Studies in Second Language Acquisition, 35*(1), 167–184. https://doi.org/10.1017/S027226311200071X

McArthur, G., & Bishop, D. (2005). Speech and non-speech processing in people with specific language impairment: A behavioural and electrophysiological study. *Brain and Language, 94*(3), 260–273. https://doi.org/10.1016/j.bandl.2005.01.002

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*(1–2), 36–47. https://doi.org/10.1016/j.heares.2006.05.004

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. A common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, and textbooks across Europe. In I. Bartning, M. Martin, & I. Uedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 211–232). European Second Language Association.

Mora, J., & Levkina, M. (2017). Task-based pronunciation teaching and research: Key issue and future directions. *Studies in Second Language Acquisition, 39*(2), 381–399. https://doi.org/10.1017/s0272263117000183

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.

Munoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics, 35*(4), 463–482. https://doi.org/10.1093/applin/amu024

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System, 34*(4), 520–531.

Parlak, Ö., & Ziegler, N. (2016). The impact of recasts on the development of primary stress in a synchronous computer-mediated environment. *Studies in Second Language Acquisition, 39*(2), 257–285. https://doi.org/10.1017/S0272263116000310

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America*, (130), 461–472. https://doi.org/10.1121/1.3593366

Rosen, S., & Manganari, E. (2001). Is there a relationship between speech and nonspeech auditory processing in children with dyslexia? *Journal of Speech, Language, and Hearing Research, 44*(4), 720–736. https://doi.org/10.1044/1092-4388(2001/057

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *Tesol Quarterly, 46*(4), 842–854. https://doi.org/10.1002/tesq.67

Saito, K. (2017). Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning, 67*(3), 665–693.

Saito, K. (2021). Corrective feedback and the development of L2 pronunciation. In H. Nassaji, & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in language learning and teaching* (pp. 407–429). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/9781108589789.020.

Saito, K. (2023). How does having a good ear promote successful second language speech acquisition in adulthood? Introducing auditory precision hypothesis-L2. *Language Teaching.* https://doi.org/10.1017/S0261444822000453

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., Petrova, K., Lee, B. J., & Tierney, A. (2022). Auditory precision hypothesis-L2: Dimension-specific relationships between auditory processing and second language segmental learning. *Cognition, 229*, Article 105236.

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of/ɹ/by Japanese learners of English. *Language Learning, 62*(2), 595–633. https://doi.org/10.1111/j.1467-9922.2011.00639.x

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., 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. *Language Learning, 71*(3), 669–715.

Saito, K., & Tierney, A. (2023). 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.* https://doi.org/10.1017/S027226312200047X

Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in L2 processing and acquisition* (pp. 15–38). Amsterdam: John Benjamins.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*(2), 263–308. https://doi.org/10.1111/j.1467-9922.2010.00562.x

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/s0272263120000649

Suzukida, Y., & Saito, K. (2021). Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Language Teaching Research, 25*(3), 431–450. https://doi.org/10.1177/1362168819858

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.* https://doi.org/10.1017/S1366728922000700

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System, 32*(4), 539–552.

Wen, Z., & Skehan, P. (2021). Stages of acquisition and the P/E model of working memory: Complementary or contrasting approaches to foreign language aptitude? *Annual Review of Applied Linguistics, 41*, 6–24. https://doi.org/10.1017/S026719052100001

Yilmaz, Y., & Granena, G. (2021). Implicitness and explicitnes in cognitive abilities and corrective feedback: A double discussion? *Studies in Second Language Acquisition, 43*(3), 523–550. https://doi.org/10.1017/S0272263120000601