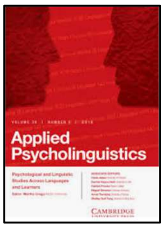
To be published in *Applied Psycholinguistics* (Cambridge University Press)



Domain-General Auditory Processing Determines Success in Second Language Pronunciation Learning in Adulthood: A Longitudinal Study

Kazuya Saito[1]

Hui Sun

Adam Tierney

Abstract

In this study, we propose a hypothesis that domain-general auditory processing, a perceptual-cognitive anchor of L1 acquisition, can serve as an important deciding factor for successful post-pubertal L2 pronunciation learning. To examine this hypothesis, samples of spontaneous speech were elicited from a total of 30 L1 Chinese L2 English learners at two points (outset and endpoint) during an eight-month study-abroad period in the UK. The participants were tested on three different components of auditory processing ability (formant, pitch, and duration discrimination) using behavioural instruments. The auditory processing scores were then linked to the segmental, prosodic and fluency dimensions of their L2 pronunciation proficiency development throughout the project. Overall, most learners' speech became smoother, faster, and more fluent (fewer pauses, faster articulation rate, more optimal perceived tempo). Interestingly, certain learners with high-level auditory processing ability (more precise formant discrimination) appeared to further attain more correct pronunciation of individual sounds and words (greater segmental and word stress accuracy), leading to more advanced L2 phonological skills (fluent *and* accurate). The findings suggest that auditory processing abilities can be a root of language learning throughout the lifespan, and may apply to the initial-to-mid phase of naturalistic L2 pronunciation learning in adulthood.

Key words: Individual Differences, L2 Speech, Pronunciation, Auditory Processing

---

[1] This study was funded by Leverhulme Trust Research Grant (RPG-2019-039) and Arnold Bentley New Initiatives Fund. We gratefully acknowledge insightful comments from Yui Suzukida, Shungo Suzuki, and three anonymous *Applied Psycholinguistics* reviewers on earlier versions of the manuscript. Corresponding Author: Kazuya Saito (k.saito@ucl.ac.uk)

Background

Perceptual-Cognitive Individual Differences in Adult L2 Speech Learning

Over the past 50 years, many scholars have explored how exposure to and practice with a target language can improve the various dimensions of post-pubertal second language (L2) proficiency. An interesting finding emerging from this broad body of research is that the outcomes of L2 acquisition in adulthood are subject to a great deal of individual variation that go beyond mere experience with the L2. In other words, even if the same two individuals engage in the same amount and type of language practice, there is a large chance that they will end up with different levels of L2 proficiency.

To date, there has been a wealth of research examining whether, to what degree, and how the process and product of L2 acquisition can be tied to the individual perceptual and cognitive abilities related to explicit language learning, hereafter referred to as *aptitude* (Li, 2016; Skehan, 2019 for comprehensive overviews). The range of theoretical frameworks and test instruments that have been devised to examine this construct (e.g., Carroll & Sapon, 1959 for Modern Language Aptitude Test; Linck et al., 2013 for Hi-LAB, Meara, 2005 for LLAMA Test) have yielded a robust body of research showing that an individual's aptitude profile can greatly influence, in particular, L2 grammar learning. These aptitude tests generally measure language-specific processing and cognitive abilities, whereas here we focus on domain general processing and perceptual abilities. For instance, certain L2 learners are more cognitively adept at decoding, memorizing and internalizing grammar rules when they are explicitly asked to do so. These abilities have been shown to be linked to faster and more substantial improvement from explicit language-focused instruction (e.g., VanPatten & Smith, 2015; Yalçın & Spada, 2016; Yilmaz, 2013).

More recently, Saito and colleagues have begun to examine the role of aptitude in L2 pronunciation learning (e.g., Saito, 2017, 2019). Drawing on Carroll and Sapon's (1959) model of foreign language aptitude, these studies have isolated one particular component of explicit phonological awareness—phonemic coding—as particularly predictive of pronunciation attainment. The studies within this paradigm have primarily relied on the LLAMA test (Meara, 2005), where participants are given a set of sound strings and corresponding symbols, and asked to memorize the sound-symbol correspondence within two minutes; and are induced to understand what the diacritics of each symbol indicate (e.g., voicing, nasalization). The results of these studies have shown that L2 learners who scored higher in the phonemic coding test had more refined pronunciation skills at the segmental level in particular, arguably because they are able to better analyze the phonemic characteristics of L2 input.

The problem inherent in this line of work is that phonemic coding is a composite construct that covers a range of different memory and analytical abilities, and only broadly represents phonological awareness (see Skehan, 2019). Little is known about what *specific* aspects of phonemic coding contribute to L2 segmental learning. Furthermore, although phonemic coding is somewhat linked to L2 learners' pronunciation abilities, and individual consonantal and vocalic sounds in particular, it is as of yet unknown which perceptual and cognitive abilities affect the prosodic dimensions of L2 pronunciation development (cf. Li & DeKeyser, 2017 for music aptitude). Following the extant literature on the fundamental role of audition in first language (L1) acquisition which is surveyed in the next section of this manuscript, we propose that individual differences in *auditory processing* may offer a perspective for conceptualizing the segmental, prosodic and fluency dimensions of L2 pronunciation learning.

Domain-General Auditory Processing and Language Learning

A major controversy in the field of cognitive psychology concerns whether and to what degree the neural mechanisms underlying L1 acquisition are specifically devoted to language learning or generalizable across various types of learning behaviours (Campbell & Tyler, 2018 vs. Hamrick, Lum, & Ullman, 2018). One such domain-general ability is auditory processing, defined as encoding, representing, and remembering time and frequency characteristics of sounds. This auditory precision is fundamental to many learning behaviours (e.g., Patel, Gibson, Ratner, Besson, & Holcomb, 1998 for auditory processing, language, and music learning). Even though any learning resulting from auditory processing takes place on a domain-specific level (e.g., language, speech, music, and emotion), they may essentially draw upon early auditory processing stages which are domain-general (i.e. precise representation of spectral and temporal details).

To date, there is ample evidence that the capacity to track and translate different acoustic dimensions of speech signals in the brain (i.e., auditory processing) is a significant determinant of L1 learning and language impairment. Within first six to eight months, infants perceptually detect, form, and adjust their speech categories to their native phonetic and phonological systems (Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson, 2006). Then, infants establish lexical and syntactic boundaries in the ambient speech stream by attuning to the prosodic patterns in the language around them (Cutler & Butterfield, 1992; Marslen-Wilson et al., 1992). As their vocabulary size expands around 18 months (i.e., vocabulary spurt), they attend to not only word-sized, but phonemic details of speech once again. Mastery of these sound-level units is subsequently used to refine phonetic representations, which in turn enhances the ability to perceive and produce a number of phonologically similar and complex words (e.g., minimal pairs) (Werker & Tees, 1999), and perceptually non-salient morphosyntactic markers (Koester, Gunter, Wagner, & Friederici, 2004).

Auditory processing ability is commonly evaluated through the use of behavioural instruments, which explicitly ask subjects to discriminate between sounds that differ in spectral (formant shape), melodic (pitch contour) and temporal cues (patterns in time). Using these behavioural paradigms, researchers have demonstrated that individual learners' auditory processing profiles predict L1 learning difficulty (Casini, Pech-Georgel, & Ziegler, 2018; Goswami, Wang, Cruz, Fosker, Mead, & Huss, 2011; Won, Tremblay, Clinard, Wright, Sagi, & Svirsky, 2016), rate of literacy development (Gibson, Hogben, & Fletcher, 2006; White-Schwoch et al., 2015), and individual differences in the impact of age on the perception of speech in noise (Ruggles, Bharadwaj, & Shinn-Cunningham, 2012; Schneider et al., 2002).

Overall, there is a consensus that individual differences in auditory processing are linked to individual differences in the speed of L1 acquisition (see Goswami, 2015 for auditory deficit hypothesis). However, whether auditory processing can *cause* differences in the speed of acquisition is controversial, since many toddlers have been shown to eventually develop normal linguistic proficiency through the use of compensatory mechanisms, despite having different levels of auditory deficiency (Amitay, Ahissar, & Nelken, 2002; Rosen, 2003).[2] It is possible, however, that the link between audition and acquisition is stronger during the initial stages of language learning, when learners begin to encounter, parse and process a new language (McArthur & Bishop, 2005).

Building on the auditory deficit hypothesis in L1 acquisition (Goswami, 2015), we argue that the role of auditory processing may play a more critical role especially in adult L2

[2] Another argument for a lack of causality underlying the auditory-attainment relationship is that the auditory processing problems are a symptom of broader attentional difficulties rather than a cause. However, this argument assumes that "attention" is a broad, unitary construct, whereas "ability to attend to sound" or even "ability to attend to a particular auditory dimension" may be specific skills that are theoretically separable from general executive function.

speech learning (the main focus of the current study) owing to the *quantitative* and *qualitative* differences in L1 and L2 learning experience. During L1 acquisition, the language is practiced intensively and extensively. Even if speakers have poor auditory processing skills, ample practice opportunities may help them overcome auditory deficits by resorting to other strategies. These compensatory mechanisms may either be cognitive (i.e. compensating for poor auditory processing by taxing memory or attention to a greater degree) or perceptual (i.e. compensating for poor auditory processing that is specific to a single perceptual channel by relying to a greater degree on alternative perceptual channels; for an example, see Jasmin, Dick, Holt, & Tierney, 2019). When it comes to a second language, however, the amount of input that a learner can access is relatively limited (compared to L1 acquisition), and thus may not be sufficient for those with auditory deficits to establish successful compensatory strategies. Following the aptitude framework in L2 learning (Skehan, 2019), it is possible that individual differences in language learning aptitude, including auditory processing skills, relate to learning success, especially when learners engage in a relatively difficult task (i.e., speaking the L2) with which a learner has less experience.

Furthermore, we echo and extend McAllister, Flege, and Piske's (2002) theoretical discussion that learners' noticing, encoding, and re-finetuning to individual acoustic dimensions (i.e., auditory processing) could be a more challenging task, and thus a more critical skill, in L2 compared to L1 speech acquisition. While learning a new language, adult learners process all the input through their already-existing auditory representational systems which have been shaped by a great deal of their accumulative L1 experience. In such contexts, they need not only detect new patterns in spectral and temporal information that they usually do not attend to as a primary cue for discriminating and identifying L1 phonological contrasts (e.g., F3 for Japanese speakers' English /r/-/l/ acquisition; Saito & Brajot, 2013), but also adjust and restructure the weights of their perceptual strategies that they are accustomed to (e.g., relying more on durational rather than pitch information for Chinese speakers' English prosody acquisition; Jasmin, Sun, & Tierney, 2020).

Though limited in number, some recent studies have begun to explore the domain-generality of the link between auditory processing and L1 acquisition to the paradigm of novel sound and word learning in laboratory settings. In Wong and Perrachione (2007), for example, English speaking adults were asked to learn an artificial language with unique lexical stress patterns. Learning outcomes were found to be tied to participants' individual differences in prosodic processing abilities. Other studies have shown that native speakers of English with better nonverbal temporal processing abilities are better able to discriminate novel speech sound contrasts which differ in durational cues (Norwegian vowels) (Kempe, Thoresen, Kirk, Schaeffler, & Brooks, 2012). When it comes to L2 speech learning, Lengeris and Hazan (2010) showed that Greek speakers with better formant discrimination abilities not only demonstrated better L2 English segmental proficiency at the outset of the project, but also showed greater gains from auditory training on the acquisition of L2 English vowels (see also Chandrasekaran, Kraus, & Wong, 2011, who used neural methods to demonstrate a link between pitch sensitivity and L2 Mandarin tone learning).

Despite emerging evidence on the relationship between auditory processing and short-term learning of language skills in the laboratory, it remains unclear whether auditory processing ability can help L2 learners acquire new phonetic categories and transfer perceptive to productive skills in immersive naturalistic settings. As a preliminary investigation, Kachlicka, Saito and Tierney (2019) and Saito, Kachlicka, Sun, and Tierney (2020) cross-sectionally compared the aptitude profiles of 100 Polish residents in the UK, their immersion experience, and various dimensions of their L2 English proficiency. The results indicated that participants' explicit auditory processing abilities (formant and pitch discrimination) significantly predicted their L2 speech perception and production performance (for the generalizability of the findings to various dimensions of L2 proficiency among Polish, Spanish, and Chinese speakers of English, see Saito, Sun, Kachlicka, Robert, Nakata, & Tierney, in press). Building on this line of work, the current study takes a very first step towards delving into the extent to which auditory sensitivity can account for the longitudinal development of L2 pronunciation abilities during eight months of immersion (study abroad) in the UK. Given the previous longitudinal evidence that adult L2 learners demonstrates much learning (with their pronunciation forms being intelligible/adequately L2-like) within first few months of immersion (e.g., Munro & Derwing, 2008; Munro, Derwing, & Saito, 2013), we consider the timeframe of the current study (i.e., first eight months) as the *initial* to *mid* phase of L2 speech learning.

Research Questions and Predictions

Drawing on the notion of auditory processing in the cognitive psychology and first language acquisition literature, the primary objective of the current investigation was to propose, test and validate our new explicit phonetic aptitude framework. As defined in Saito and Plonsky's (2019) assessment model, L2 pronunciation proficiency is first considered to comprise three overall abilities: (a) the ability to pronounce consonants and vowels accurately (segmental proficiency); (b) the ability to correctly assign stress at both the word and sentence levels (prosodic proficiency); and (c) the ability to deliver speech at an optimal tempo without too many pauses, repetitions and/or self-corrections (fluency proficiency). A total of three different perceptual-cognitive abilities are assumed to be relevant to these: the ability to capture perceptible differences in (a) higher frequencies (formant discrimination), (b) lower frequencies (pitch discrimination), and (c) sound and silence ratio (duration discrimination). Accordingly, the research question guiding the current study was:

- To what degree does explicit phonetic aptitude, operationalized as three types of auditory processing (formant, pitch and duration discrimination), relate to the segmental, prosodic, and fluency aspects of L2 pronunciation proficiency development?

While our general prediction is that precise auditory processing facilitates various dimensions of L2 pronunciation development, different dimensions of L2 pronunciation proficiency development may involve different types of auditory processing abilities with different weights. Thus, we aim to explore more specific hypotheses in conjunction with the primary acoustic correlates of L2 segmental, prosodic and fluency development. For segmental proficiency, the accurate pronunciation of consonants and vowels requires L2 learners to access all the relevant auditory processing strategies (formant, pitch and duration discrimination). At the same time, however, it has been shown that many L2 learners tend to *overly* rely on melodic and temporal (rather than spectral) cues to perceive and produce new L2 sounds in the early stage of interlanguage development (e.g., Bohn & Flege, 1992 for vowels; Saito & Brajot, 2013 for consonants). Thus, we assume that it is spectral sensitivity (i.e., formant discrimination) that may play a key role in aiding L2 learners in this study to further refine phonetic details with a view of attaining more advanced L2 segmental proficiency.

The acquisition of prosody is highly multi-dimensional in nature. There are both pitch-based and durational correlates of stress at word and sentence levels alike. Thus, we consider both pitch and duration processing to be *equally* instrumental to the attainment of high-level prosodic proficiency. Finally, all the fluency phenomena featured in this study—breakdown and speed fluency—are directly related to L2 learners' sensitivity to the timing of sound and silence in language input. Therefore, participants' temporal processing abilities (i.e., duration discrimination) could be a primary predictor of participants’ L2 fluency development over time.

Though admittedly simplified, our specific hypotheses are summarized in Table 1. L2 pronunciation proficiency development is driven by auditory processing abilities related to the primary acoustic correlates of L2 segmental, prosodic and fluency development: (a) those with greater spectral sensitivity (formant discrimination) will demonstrate more advanced segmental proficiency (consonant and vowel accuracy); (b) those with greater melodic and temporal sensitivity (pitch and duration discrimination) will show better prosodic proficiency (word stress, intonation); and (c) those with greater temporal sensitivity (duration discrimination) will exhibit better fluency proficiency (breakdown and speed fluency).

Table 1 *Predicted Relationship between L2 Pronunciation Proficiency and Auditory Processing*

Proficiency dimensions	Primary Perceptual-Cognitive Correlates
• Segmentals	• Formant discrimination
• Prosody	• Pitch and duration discrimination
• Fluency	• Duration discrimination

Method

The current study featured a total of 30 Chinese international students during an eight-month period of study abroad in the UK. The participants’ pronunciation proficiency was assessed at the outset (T1) and endpoint (T2) of their immersion, and their auditory processing was measured at T2 (for details, see below). From a cross-sectional perspective, we explored the degree to which the participants’ L2 pronunciation proficiency was associated with their past English-as-a-Foreign-Language (EFL) experience profiles (before study abroad) and auditory processing ability. In addition, we conducted a longitudinal investigation of the relationship between participants’ recent English-as-a-Second-Language (ESL) experience, auditory processing and L2 pronunciation proficiency development over time (T1 → T2). The design of the study is visually summarized in Figure 1.

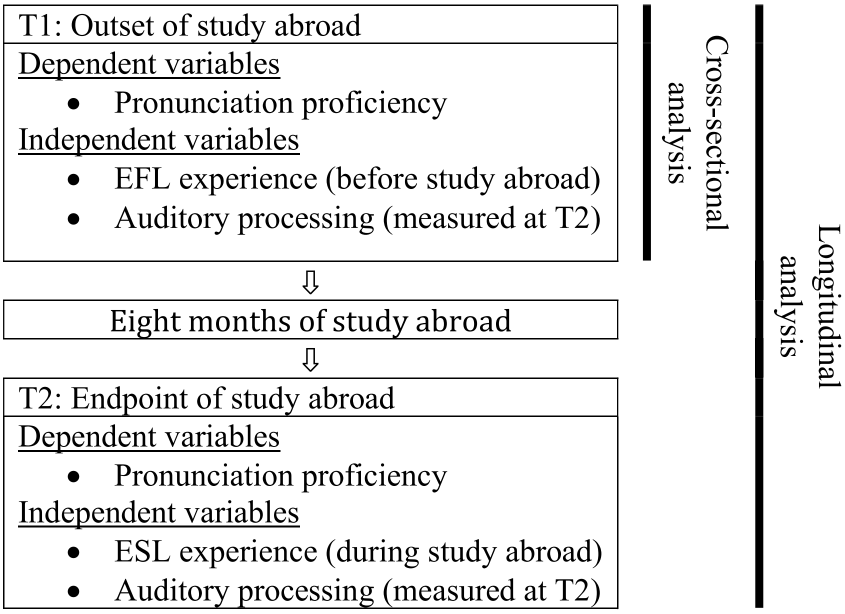


Figure 1 Design of Current Study

Participants

As a part of a larger project which examined the L2 pronunciation proficiency of 100+ learners with a wide range of L1, experience, age and aptitude profiles in London, 30 Chinese learners of English (27 females) were recruited at the beginning of an academic year (T1). Their average age was 23.5 years (*SD* = 2.2, *Range* = 21-34). All of them were graduate students in various areas of sciences, social sciences and humanities. At the outset of the project, they had resided in the UK for less than one month. Their general L2 English proficiency as measured by IELTS spanned between 6 and 8 out of 9 (*M* = 7.1, *SD* = 0.5, *Range* = 6-8), which classifies them as independent and proficient L2 users (i.e., B2 to C2 as per Common European Framework of Reference). All the participants reported normal hearing. Participants participated in a second data collection session eight months after T1, i.e., at the end of their program (T2).

All participants were individually interviewed at T1 to gather information about their past EFL experiences. The participants reported having studied English for an extensive amount of time in China (*M* = 15.3 years, *SD* = 2.1, *Range* = 10-22) with different onsets of learning (*M* = 8.3 years, *SD* = 1.9, *Range* = 6-13). According to their self-reports, the total number of hours they engaged in EFL education widely varied (*M* = 3037.5 hours, *SD* = 1152.1, *Range* = 1520-6840). No participants reported having any extensive immersion experience prior to the project. Because the quantity of L2 use likely differed among the participants during the project, they were interviewed again at T2 to survey their ESL experience (T1 → T2). The results of this interview indicated that their estimated amount of L2 use per week widely varied (*M* = 8.0 hours, *SD* = 9.8, *Range* = 1-51).

Speaking Task

L2 pronunciation scholars have generally relied on controlled speech tasks (e.g., word and sentence reading) in order to lead their participants to pronounce certain phonological features that they are interested in (for research synthesis, see Saito & Plonsky, 2019). However, this task format allows adult L2 learners to monitor their correct pronunciation forms, which may not necessarily index their actual L2 proficiency in communicative settings (Piske, Flege, MacKay, & Meador, 2011). In this regard, a timed picture description task was used to elicit participants' relatively unconscious and unmonitored use of L2 pronunciation forms (i.e., spontaneous L2 pronunciation proficiency). In the task, participants described a series of picture narratives with a primary focus on meaning under time pressure. Four different versions of the task (Versions A, B, C & D) were selected from the Pre-Grade 1 Level of the EIKEN English Test (EIKEN, 2016), and administered randomly at each testing point (Version A or B at T1; Version C or D at T2). In this way, we were able to control for the confounding effects of different topics, as well as test-retest effects. In each version, participants were asked to describe a four-frame cartoon with two minutes of planning time. To avoid false starts, they were given the first sentence that they had to use in the task (for all the task materials used in the study, see **Supporting Information-A**).

All of the speech samples were recorded in a quiet room at a university using a Roland-05 audio recorder, set to 44100 Hz sampling rate and 16-bit quantization, and via a unidirectional condenser microphone. Following the standards in L2 pronunciation research (e.g., Isaacs & Trofimovich, 2012) and to reduce the amount of listener fatigue during the L2 speech rating sessions (see below), the first 30 seconds of each speech sample were cut, normalized for peak intensity, and saved as a WAV file. A total of 60 speech samples (30 participants × 2 data collection points [T1, T2]) were submitted to subjective (rater judgements) analysis and objective acoustic analysis, as described below.

Constructs of L2 Pronunciation Proficiency

L2 pronunciation proficiency is broadly conceptualized on segmental and suprasegmental levels (Saito & Plonsky, 2019). Following Trofimovich and Baker's (2006) framework, the latter construct of L2 pronunciation can be further divided into "melody-based" suprasegmentals, defined as acoustic phenomena signaled via fundamental frequency and intensity (word stress, intonation); and "rhythm-based" suprasegmentals, featuring all the temporal characteristics (breakdown and speed fluency). In this paper, therefore we considered L2 pronunciation to be three-fold: (a) the ability to pronounce consonants and vowels accurately (segmental proficiency); (b) the ability to correctly assign stress at both the word and sentence levels (prosodic proficiency); and (c) the ability to deliver speech at an optimal tempo without too many pauses, repetitions and/or self-corrections (fluency proficiency).

L2 Pronunciation Assessment: Rater Judgements

A total of five expert raters were trained to make impressionistic judgements of four different dimensions of L2 pronunciation: (a) segmentals (substitution, omission, or insertion of individual consonants or vowels); (b) word stress (misplaced or missing primary stress in multisyllabic words); (c) intonation (appropriate, varied versus incorrect and monotonous use of pitch on a sentence-level); and (d) perceived tempo (optimal speed of delivery) (for development and validation of the procedure, see Saito, Trofimovich and Isaacs (2017). All of the speech samples were played for the raters in a randomized order via a MATLAB-based software. For each sample, raters used a moving slider to assess the degree of targetlikeness, which was recorded on a 1000-point scale (*0 = not targetlike at all, 1000 = targetlike*). For onscreen labels for each rated category, see **Supporting Information-B**. The first author administered the rater training and listening sessions.

Raters. As recommended in Saito et al. (2017), five raters (2 males, 3 females) were carefully selected in light of their experience, familiarity, and expertise with L2 speech judgements of this kind. Four raters were native speakers of British English, while one rater was a native speaker of American English. All of them had obtained an MA in Applied Linguistics, had extensive English teaching experience (*M* = 9.6 years, *SD* = 8.9, *Range* = 3-

20 years), and reported high-level familiarity with foreign-accented English (*M* = 6 out of 6: *1 = not at all; 6 = very much*).

Procedure. All the listening sessions took place individually in a quiet room at a university in London. First, the raters received detailed explanation from the first author about the meanings of each rating category, and how to use the scale. For training scripts, see **Supporting Information-C**. Next, they familiarized themselves with the rating procedure by using a set of five practice samples (but not included in the main dataset). For each sample, they explained and justified their segmental, prosodic (word stress, intonation) and temporal (perceived tempo) ratings, with the researcher providing feedback to ensure that the raters had clearly understood and differentiated between the four rated categories. The raters then proceeded to rate the main dataset of 60 audio samples. The entire session lasted for two hours. To avoid listener fatigue, raters were given a break halfway through.

Inter- and Intra-Rater Reliabilities. Cronbach's alpha was calculated to check the agreement among the pronunciation proficiency ratings. In line with previous research (e.g., Saito et al., 2017), the results found relatively high alpha levels for segmentals ($\alpha = .85$), word stress ($\alpha = .80$), intonation ($\alpha = .78$) and speech rate ($\alpha = .85$), indicating that the raters' judgements were overall consistent across the four different dimensions of L2 pronunciation proficiency. In addition, the raters assessed the extent to which they had understood each pronunciation category (*1 = I did not understand at all, 9 = I understand this concept well*); and the extent to which they could use/handle the category while listening spontaneous speech samples (*1 = very difficult, 9 = very easy and comfortable*). The raters reported relatively high level of understanding (*M* = 9.0) and comfort (*M* = 8.6) for all linguistic categories. For the subsequent analyses, all the raters' scores were averaged across for each participant (Participants 1-30) at different time points (T1, T2) for the four L2 pronunciation categories (segmentals, word stress, intonation, perceived tempo).

L2 Pronunciation Assessment: Acoustic Analyses

Following the assessment framework of L2 utterance fluency (Tavakoli & Skehan, 2005), two linguistically trained coders participated and conducted acoustic analyses of two temporal dimensions of L2 speech—(a) breakdown (the ratio of filled and unfilled pauses); and (b) speed (articulation rate)[3].

To conduct the analysis, all speech samples were first transcribed and separated into clauses using the Analysis of Speech unit (Foster, Tonkyn, & Wigglesworth, 2000). As operationalized in Bosker, Pinget, Quené, Sanders and De Jong (2013), the number of filled (e.g., ah, oh, eh) and unfilled pauses were combined. The pauses were divided by the total number of words. For speed fluency, the total number of syllables was divided by the total phonation time (without any unfilled pauses; i.e., articulation rate).

Two experienced, trained coders first separately conducted fluency analyses on 20 similar L2 picture narratives (not included in the main dataset). Their inter-rater reliability was relatively high for breakdown (inter-class correlations: $r = .95$ for pause ratio) and speed ($r = .96$ for articulation rate). Any discrepancies were discussed and solved with the view of obtaining a consistent understanding of the breakdown and speed fluency analyses. Subsequently, the first coder proceeded to the fluency analyses of the main dataset ($n = 60$ speech samples).

Auditory Processing Measures

In light of the existing literature on auditory individual differences in L1 acquisition, audition abilities were operationalized as one's sensitivity to perceptual differences in formant frequency, pitch, and duration (e.g., McArthur & Bishop, 2005). Adopting the procedure/materials established in the precursor research (e.g., Kachlicka et al., 2019), participants took three different types of auditory processing tests—formant, pitch and duration discrimination. The participants took the tests in the following order: formant discrimination, pitch discrimination, and duration discrimination. For practical reasons, the testing sessions took place at the endpoint of the project. Because domain-general abilities such as auditory processing are thought to be relatively *stable* for the same individuals over time (Hornickel, Knowles, & Kraus, 2012), we do not believe that the timing of the assessment (at the endpoint of the project) significantly affected how well the participants'

[3] In the current study, we did not further analyze the repair aspects of L2 fluency (the ratio of repetitions and self-corrections). This decision was based on the growing evidence that repair is deeply associated with L1 traits and behaviors (e.g., personality, attentional control; Zuniga & Simard, 2019) rather than L2-specific variables (e.g., length of immersion, age of acquisition; Saito, Ilkan, Magne, Tran, & Suzuki, 2018).

explicit perceptual acuity was represented. In the Discussion section, however, we revisited this assumption, providing food for thought for future studies of this kind.

The auditory processing tests were designed to assess the participants' thresholds for sound discrimination on three dimensions—(a) frequency of spectral peaks, (b) fundamental frequencies; and (c) duration. For each test a total of 100 stimulus levels were created via custom MATLAB scripts. Participants were presented with three tones with an inter-stimulus interval of 0.5s and were asked to indicate which of the tones differed from the middle one by pressing the number "1" or "3." Using an adaptive threshold procedure (Levitt, 1971), the size of the difference varied from trial to trial based on task performance. Initially, the tests started at Level 50. For every incorrect response, the difficulty of the task decreased (with the difference being wider). When participants correctly responded to three consecutive stimuli, task difficulty increased (with the difference being smaller).

The program also tracked which "direction" the threshold had been moving over the last couple of trials. For example, if the current stimulus difference was smaller than the difference in the last trial, then thresholds had been decreasing. It then looked for trials in which there was a "reversal" in direction. This could be due either to a participant getting the answer right after a string of wrong answers, or to a participant getting the answer wrong after a string of correct answers. The step size was initially set at ten levels. After the first reversal in direction (i.e., an increase in difficulty followed by a decrease or vice versa), the step size decreased to five levels, and decreased to two and one after the second and third reversals. The program stopped after either 50 trials or eight reversals have been reached. Then, to calculate the individual's threshold, the program averaged the stimulus differences levels that occurred across all reversals from the second on. Thus, this program was designed to calculate the level at which participants could just barely do the task.

For the formant discrimination test, stimuli were complex tones with fundamental frequencies (F0) at 100 Hz and harmonics up to 3000 Hz. Stimuli were 0.5s in duration with a 0.015s linear ramp at the beginning and the end. Three formants were created using a parallel formant filter bank (Smith, 2007). The first and third formants were kept constant at 500 Hz (F1) and 2500 Hz (F3). The frequency of the second formant (F2) ranged from 1500 Hz to 1700 Hz with an increment of 2 Hz. For example, if a participant stops at Level 50, this means that he can perceive a minimum difference of 100Hz (2Hz × 50 levels) when two stimuli differ in F2.

For the pitch and duration discrimination tests, a standard four-harmonic complex tone was created with F0 at 330 Hz. Each stimulus was 0.5s in duration, beginning and ending with 0.015s of linear amplitude ramp. While the stimuli in the pitch discrimination test differed on the continuum of F0 between 330 and 360 Hz (with a step of 3 Hz), those in the duration discrimination test ranged from 0.25 to 0.5s (0.0025s steps). For example, if a participant ends at Level 50, this means that he can perceive a minimum difference of 150Hz (3Hz × 50 levels) when two stimuli differ in F0, or a minimum difference of 0.125s (0.0025s × 50 levels) when they differ in duration. As such, formant, pitch and duration processing scores were calculated in terms of how small of a difference participants could perceive on a 100-point continuum. For more methodological details, audio materials, and validation research, see Saito, Sun, and Tierney (2020).

Results

L2 Pronunciation Proficiency

The segmental, prosodic, and fluency dimensions of the participants' speech were assessed using the four rater judgement measures (segmentals, word stress, intonation, perceived tempo) and the two acoustic measures (pause ratio, articulation rate). According to Kolmogorov-Smirnov tests, the distribution of the two acoustic measures (pause ratio, articulation rate) were found to be positively skewed ($p < .05$). Therefore, the log10 function was used to transform and normalize the data. To check if the four types of picture prompts (Versions A, B, C, vs. D) influenced participants' L2 pronunciation proficiency, a set of repeated-measure ANOVAs were conducted with each dimension of participants' pronunciation proficiency scores relative to task prompts at T1 (Versions A vs. B) and T2 (Versions C vs. D). The results did not find any significant relationship between L2 pronunciation proficiency and task prompts in any contexts ($p > .05$). The descriptive statistics of L2 pronunciation proficiency scores were summarized in Table 2.

Table 2 *Descriptive Information of L2 Pronunciation Proficiency Scores at T1 and T2*

	T1			T2		
	M	*SD*	*Range*	*M*	*SD*	*Range*
A. Rater judgements						
Segmentals (1000 points)	458	131	213-754	514	130	293-749
Word stress (1000 points)	582	104	368-755	577	119	331-897
Intonation (1000 points)	449	135	207-727	487	158	277-856
Perceived tempo (1000 points)	521	135	256-841	602	121	415-893
B. Acoustic analyses						
Pause ratio (%)	31.1	9.2	15.2-62.1	26.8	8.5	16.5-58.3
Articulation rate (syllables per second)	2.9	0.4	2.1-4.1	3.2	0.4	2.4-4.3

To delve into the relationship among the participants' pronunciation scores (elicited from the six measures), a set of Pearson correlations were conducted with their T1 scores. Due to multiple comparisons (i.e., each measure was compared against the other five measures), the alpha level was adjusted and set at $p < .010$ (Bonferroni corrected). As summarized in Table 3, all subcategories of the rater judgements were significantly correlated with each other except for the link between segmentals and perceived tempo ($r = .375$, $p = .041$). Interestingly, while one acoustic measure (articulation rate) was significantly associated with raters' judgements of suprasegmentals in particular (word stress, intonation, perceived tempo), the other acoustic measures appeared to be rather unrelated and independent (breakdown vs. speed). Thus, a total of six measures were assumed to tap into the following three overall domains of L2 pronunciation proficiency: (a) correct pronunciation of individual sounds and words (segmentals, word stress, intonation); (b) overall prosody (word stress, intonation, perceived tempo, articulation rate); and (c) temporal fluency (pause ratio, articulation rate).

Table 3 *Inter-Relationships between L2 Pronunciation Proficiency Scores (T1)*

	Word stress		Intonation		Perceived tempo		Pause ratio		Articulation rate	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Segmentals (1000 points)	.708*	< .001	.714*	< .001	.375	.041	.067	.726	.445	.014
Word stress (1000 points)			.743*	< .001	.652*	< .001	.116	.541	.534*	.002
Intonation (1000 points)					.628*	< .001	.149	.432	.514*	.004
Perceived tempo (1000 points)							-.433	.017	.595*	.001
Pause ratio (%)									-.467*	.009

Note. * for statistically significant ($p < .010$) (Bonferroni corrected)

L2 Auditory Processing Profiles

Auditory processing scores were recorded on a 100-point scale (smaller values indicating greater spectral, melodic and rhythmic sensitivity). The raw data were summarized in Table 4. For the sake of interpretation, smaller values indicate more sensitive and better auditory processing abilities. As for formant discrimination, for example, some could perceive very subtle differences when F2 shifted to a very small degree (e.g., 7 Hz), but the formant acuity of others could be much less sensitive (e.g., 117 Hz). As observed in the precursor research (e.g., Kachlicka et al., 2019; Saito et al., 2020), the results of Kolmogorov-Smirnov tests found the pitch and duration discrimination scores to be positively skewed ($p < .001$, $p = .041$). To approximate normal distribution, these auditory scores were thus transformed via the log10 function. For the purpose of comparisons and subsequent analyses, all the auditory processing scores were converted to z-scores.

Table 4 *Summary of Raw Auditory Processing Scores*

	M	*SD*	*Range*
Formant discrimination	55.4 Hz	29.4	7-117
Pitch discrimination	5.0 Hz	3.8	1.4-16.9
Duration discrimination	51.4 ms	35.1	11.2-127.5

To check the inter-relationships between the three auditory processing measures (formant, pitch, and duration) and the two experience factors (past EFL and recent ESL experience), Pearson correlation analyses were performed. Considering the number of comparisons (each independent variable was compared against the other four variables), an alpha level was set to .012 (Bonferroni corrected). As summarized in Table 5, no significant correlations were identified in any contexts, except for the relationship between past and recent experience (total hours of L2 use in ESL settings vs. total hours of L2 education in EFL settings). In terms of the degree of independence of auditory processing factors, these results indicate that (a) formant, pitch and duration discrimination measures may represent three distinct aspects of auditory processing abilities; and (b) such auditory processing factors may be outside of the influence of past/recent L2 learning experience.

Table 5 *Inter-Relationships between Auditory Processing and Experience Scores*

	Pitch discrimination		Duration discrimination		Past EFL experience		Recent ESL experience	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*
A. Auditory processing								
Formant discrimination	133	.483	.061	.748	-.128	.501	-.247	.188
Pitch discrimination			.169	.371	.108	.569	-.020	.915
Duration discrimination					.259	.167	-.045	.814
B. Experience								
Past EFL experience							.552	.002*

Note. * for statistically significant ($p < .012$) (Bonferroni corrected)

Cross-Sectional Investigations (Onset of Study Abroad: T1)

We next examined the relationships between participants' auditory processing (formant, pitch, and duration discrimination), past experience (the amount of L2 use prior to study abroad) and L2 pronunciation proficiency at the outset of the project. A set of Pearson correlation analyses were performed with an alpha level set to .012 via Bonferroni corrections (each pronunciation measure was compared against three audition measures and one experience measure) (see Table 6). Three areas of participants' L2 pronunciation proficiency (word stress, perceived tempo, articulation rate) yielded significant associations with the total number of hours they had spent on practicing the target language during 10-20 years of EFL education in China (r = .493-.518). According to Plonsky and Ghanbar's (2018) field-specific benchmarks, the strength of the correlations could be considered relatively large. Interestingly, the link between auditory processing and any dimension of L2 pronunciation proficiency remains unclear, as none of the contrasts reached statistical significance ($p > .012$).

Table 6 *Relationship between Auditory Processing, Experience and L2 Pronunciation Proficiency at the Beginning of Study Abroad (T1)*

	A. Auditory processing						B. Experience	
	Formant discrimination		Pitch discrimination		Duration discrimination		Total hours of L2 before study abroad	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Segmentals	-.226	.229	-.189	.317	.113	.554	.344	.063
Word stress	.062	.745	-.195	.300	.190	.316	.493*	.006
Intonation	-.026	.890	-.083	.662	.024	.898	.299	.109
Perceived speed	-.054	.777	.019	.920	.111	.559	.501*	.005
Pause ratio	.233	.215	.278	.138	.110	.563	-.098	.606
Articulation rate	-.141	.457	-.295	.114	.175	.354	.518*	.003

Note. * for statistically significant ($p < .012$) (Bonferroni corrected)

Longitudinal Investigations (T1 to T2)

Our final objective was to conduct a *longitudinal* examination of the extent to which the participants' diverse aptitude and experience profiles could influence their L2 pronunciation proficiency over a period of eight months studying abroad in the UK. To obtain an overall picture of how both aptitude and experience factors as a composite model explained variance in L2 pronunciation development, linear mixed-effects modelling analyses were conducted using the lmer functions from the lme4 package (Version 1.1-21; Bates, Maechler, Bolker, & Walker, 2015) in the R statistical environment (R Core Team, 2018). While participants' L2 scores were used as dependent variables, a total of six separate models were constructed for six dimensions of pronunciation —i.e., segmentals, word stress, intonation, perceived tempo, pause ratio, and articulation rate.[4]

[4] We did not combine all pronunciation measures within the same model as composite dependent variables, because each measure is hypothesized to tap into essentially different dimensions of L2 pronunciation proficiency (as we explained and justified in the Method section), and the resulting scores cannot be directly compared across different dimension conditions. For example, positive gains indicate more development in raters' judgements of segmentals, word stress, intonation, and perceived speed; and reduction indexes more development in the ratio of pauses. Although raters' perceived tempo and articulation rate appear to concern similar dimensions of L2 pronunciation proficiency (speed fluency), their inter-relationship could be complex (curvilinear rather than linear) (Munro & Derwing, 2001).

For each model, participants' pronunciation scores at T1 and T2 were used as dependent variables. To track their gains over the course of eight months of immersion, Time was treatment coded by assigning "-.5" to participants' pronunciation scores at T1, and ".5" to their scores at T2. The fixed effects included five predictor variables—i.e., (a) the main effect of time (T1, T2), (b) the interaction of time and three auditory processing measures (formant discrimination, pitch discrimination, duration discrimination), and (c) the interaction of time and the total amount of L2 use (hours) during study abroad (between T1 and T2). The random effects included participants' ID (1-30). Due to the size of sample, participants were included as random intercepts only.

To avoid type 1 errors, the following planned comparison was made for the construction of the most optimal model in response to our research question. The main objective of the longitudinal analyses was to examine not only the extent to which L2 participants' proficiency significantly changed due to the eight months of immersion (i.e., main effects of time), but also how such development over time could be associated with auditory processing (interaction effects of time and auditory processing) and with L2 use (interaction effects of time and recent L2 use). Thus, the following model was used to examine the source of variances among the participants' L2 pronunciation developmental patterns (T1 → T2). For R code, see **Supporting Information-D**.

- Pronunciation proficiency = Intercepts + Time + Time × Formant_Disctimination + Time × Pitch_Disctimination + Time × Duration_Disctimination + Time × Recent_L2_Use + Random effects of participants

Table 7 *Results of Mixed Effects Modeling Analyses on Factors Affecting L2 Pronunciation Development*

Pronunciation proficiency	Fixed effects: Factor	*Estimate*	*SE*	*t*	*p*	Random effects: Factor	Variance	*SD*	Conditional R^2	Marginal R^2
Segmentals	Intercepts	485.95	22.01	22.07	< .001*	Participants	0.559	0.748	.766	.245
	Time	59.20	22.73	2.60	.015*					
	Time × Formant	-40.81	18.02	-2.26	.032*					
	Time × Pitch	-25.22	17.70	-1.43	.167					
	Time × Duration	-13.54	17.58	-0.77	.448					
	Time × L2_Use	-0.02	0.06	-0.27	.792					
Word stress	Intercepts	579.41	17.08	33.93	< .001*	Participants	0.472	0.687	.653	.271
	Time	20.00	20.62	0.97	.341					
	Time × Formant	-94.22	16.61	-5.67	< .001*					
	Time × Pitch	7.46	16.31	0.46	.651					
	Time × Duration	-24.82	16.20	-1.53	.138					
	Time × L2_Use	-0.10	0.05	-1.84	.077					
Intonation	Intercepts	467.83	23.67	19.77	< .001*	Participants	0.5361	0.7322	.569	.073
	Time	67.06	33.92	1.98	.059					
	Time × Formant	-18.07	26.88	-0.67	.508					
	Time × Pitch	-1.21	26.40	-0.05	.964					
	Time × Duration	-25.02	26.22	-0.95	.349					
	Time × L2_Use	-0.11	0.08	-1.32	.199					
Perceived speed	Intercepts	561.58	20.97	26.77	< .001*	Participants	0.5646	0.7514	.674	.157
	Time	94.54	26.46	3.57	.001*					
	Time × Formant	-26.43	21.21	-1.25	.224					
	Time × Pitch	-23.83	20.83	-1.14	.263					
	Time × Duration	-20.60	20.69	-1.00	.329					
	Time × L2_Use	-0.05	0.07	-0.79	.436					
Pause ratio	Intercepts	0.2901	0.0153	18.89	< .001*	Participants	0.7269	0.8526	.805	.132
	Time	-0.0433	0.0139	-3.11	.005*					
	Time × Formant	0.0043	0.0111	0.39	.698					
	Time × Pitch	2.95E-05	0.0109	0.00	.998					
	Time × Duration	-0.0243	0.0109	-1.776	.081					
	Time × L2_Use	-8.69E-07	3.52E-05	-0.03	.980					
Articulation rate	Intercepts	3.1268	0.074473	41.98	< .001*	Participants	0.6197	0.7872	.740	.160
	Time	0.4028	0.079877	5.04	< .001*					
	Time × Formant	-0.0477	0.063423	-0.75	.458					
	Time × Pitch	0.0799	0.06228	1.28	.211					
	Time × Duration	0.0143	0.061865	0.23	.818					
	Time × L2_Use	-0.0004	0.0002	-0.34	.734					

As summarized in Table 7, significant main effects of time were identified for perceived speed, pause ratio, and articulation rate. The findings here imply that the eight months of study abroad were facilitative of the temporal aspects of L2 pronunciation proficiency development (i.e., more optimal tempo), and the gains were attributable to reduced pause ratio and increased articulation rate. Furthermore, significant interaction effects of time and auditory processing (formant discrimination) were found for segmentals and word stress. The results indicate that auditory processing may play a key role in determining the amount of improvement especially in the accuracy aspects of L2 pronunciation proficiency; and that certain participants with more precise auditory processing could demonstrate not only more fluent, but also more accurate L2 pronunciation proficiency development over time. The relationships between participants' gain (from T1 to T2) and formant discrimination residuals were visually summarized in Figure 2.

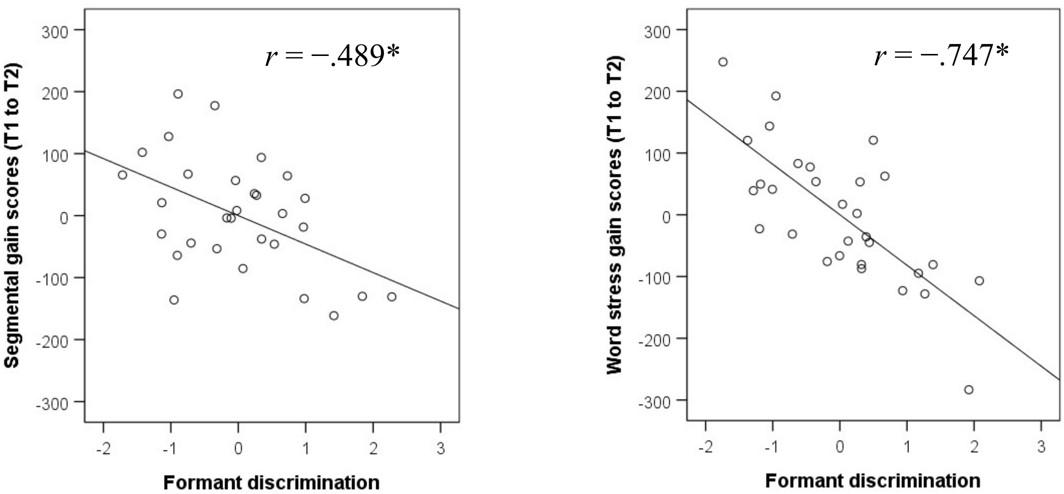


Figure 2. Correlation between L2 pronunciation gain scores and auditory processing residuals with initial proficiency partialled out

Discussion, Conclusion, and Future Directions

The current study proposed a hypothesis that a perceptual-cognitive foundation of successful L1 acquisition, domain-general auditory processing, can account for some variance in post-pubertal L2 pronunciation learning. Specifically, we predicted that learners with greater sensitivity to the formant, pitch, and duration aspects of acoustic information could make the most of L2 input and output opportunities to improve the segmental, prosodic, and fluency aspects of their L2 pronunciation proficiency. To test this hypothesis, we conducted cross-sectional and longitudinal investigations on the pronunciation development of 30 L1 Chinese students with diverse profiles of auditory processing abilities (formant, pitch, and duration discrimination), when they engaged in eight months of study abroad in the UK. A total of six pronunciation measures were adopted to tap into three different dimensions of the participants' L2 pronunciation proficiency: (a) correct pronunciation of sounds and words, (b) overall prosodic accuracy, and (c) temporal fluency.

Results of the analyses revealed that the participants' L2 pronunciation proficiency was not significantly related to their auditory processing profiles at the outset of the project. Rather, the quality of their suprasegmental performance was significantly tied to the extent to which they had previously practiced the target language in China prior to their arrival in the UK. Over the course of the eight-month immersion period, however, most learners were able to attain smoother, faster, and more fluent speech (fewer pauses, faster articulation rate, more optimal perceived tempo). Additionally, it was certain learners with high-level auditory processing abilities (i.e., more precise formant discrimination) that further enhance their L2 pronunciation proficiency via greater segmental and word stress accuracy (more correct pronunciation of individual sounds and words), leading to more advanced L2 pronunciation (fluent *and* accurate).

From a theoretical perspective, our findings lend some support for the claim that auditory processing abilities can act as a bottleneck for post-pubertal L2 pronunciation learning especially at the initial to mid stage of immersion—i.e., first eight months of study abroad for L1 Chinese students without any experience overseas. While some studies have noted the relationship between auditory processing and L2 speech perception development through intensive laboratory training (e.g., Lengeris & Hazan, 2010), our study provided the very first empirical evidence from a longitudinal perspective that auditory processing is tied to L2 pronunciation learning in an immersive, naturalistic context.

In line with the multifaceted view of L2 aptitude and acquisition (Li, 2016; Skehan, 2019), it is important to point out that different dimensions of L2 learning outcomes were driven by factors related to different aptitude profiles (e.g., auditory processing). Notably, perceptual and cognitive individual differences are thought to play less of a role for relatively learnable features such as better perceived tempo and articulation rate (Lahmann, Steinkrauss, & Schmid, 2017). By comparison, individual differences in auditory processing may be more influential in determining the attainment of advanced-level L2 pronunciation accuracy and fluency, characterized as greater segmental and prosodic refinement (Granena & Long, 2013; Saito, 2019).

On the one hand, participants were able to significantly enhance the speed fluency aspects of their L2 speech (perceived tempo, articulation rate)—relatively salient and learnable features—as a function of increased experience alone (eight months of study abroad). Although the participants' reported a different amount of L2 use over during study abroad, the significant main effects of time in the mixed effects modeling analyses showed that the speed of their L2 speech became significantly faster on the whole. This finding concurs with a range of previous studies showing that L2 fluency development is highly susceptible to the quantity of L2 experience (e.g., Derwing, Munro, Thomson, & Rossiter, 2009; Saito, Ilkan, Magne, Tran, & Suzuki, 2018; Trofimovich & Baker, 2006;). In other words, as long as L2 learners practice the target language on a regular basis, a great deal of improvement is likely to be observed in various areas of fluency, not only within a short period of study (e.g., Mora & Valls-Ferrer, 2012; Segalowitz & Freed, 2004), but over a longer period of immersion (e.g., Derwing et al., 2009; Saito et al., 2018; Trofimovich & Baker, 2006).

On the other hand, the findings suggest that learners may require high-level aptitude on top of extensive experience when the target features are relatively complex, non-salient, or difficult (Doughty, 2019; Li, 2016; Skehan, 2019). Our findings suggest that for pronunciation, formant discrimination may anchor the degree of segmental and word stress (rather than perceived tempo and fluency) acquisition. These accuracy dimensions are considered to be relatively complex and difficult aspects of L2 pronunciation proficiency, as we will further discuss and elaborate below (Saito, 2015a, 2015b).

Previous research has indicated that most L2 segmental learning happens within the first few months of immersion, as learners' pronunciation becomes readily intelligible with frequent and familiar words. However, this is quickly followed by a plateau (Munro & Derwing, 2008; Munro et al., 2013), and leads to a degree of foreign accentedness that is highly resistant to change, especially on a segmental level (Muñoz & Llanes, 2014; Saito, 2015a). To further refine segmental accuracy, L2 learners may require extensive amounts of explicit instruction (Thomson & Derwing, 2015) and/or naturalistic exposure to the L2 (e.g., 10+ years of residence: Flege, Takagi, & Mann, 1995; 2015b). Similarly, L2 learners appeared to show tremendous difficulty in the correct assignment of word stress, arguably because it requires the learners to decode the timing and location of vowel reduction by using all spectral, prosodic, and durational cues in speech at the same time (Trofimovich & Baker, 2006). It is in the acquisition of such difficult features that aptitude may make a large difference. Though few in number, there is emerging evidence that those with greater phonemic coding ability likely show more nativelike segmental pronunciation skills on a global (e.g., Granena & Long, 2013 for global foreign accentedness) and specific level (e.g., Saito, 2019 for English /r/-/l/).

In the current investigation, two auditory processing measure (i.e., pitch and duration discrimination) did not show significant associations with any aspects of L2 pronunciation proficiency development, as opposed to our earlier predictions. The findings here were surprising, because pitch information plays a key role in other instances of L2 speech acquisition (cf. Li & DeKeyser, 2017 for American learners of Mandarin Chinese). As one possible interpretation, the findings could be ascribed to the nature of the participants' L1 background—i.e., Mandarin Chinese. In the previous literature, it has been shown that tonal language speakers tend to demonstrate stronger brainstem representation of fundamental frequencies than non-tonal language users (Bidelman, Gandour, & Krishnan, 2011) as well as more accurate pitch discrimination (Giuliano et al. 2011). As a result, Mandarin L1 speakers may be generally more capable of sufficiently precise pitch perception to extract prosodic structure from speech as compared to speakers of other languages, weakening relationships between pitch perception and prosodic and phonetic processing. This explanation could be tested in future research examining links between auditory processing and pronunciation in speakers of other languages.

To close, the exploratory nature of the project, its design, and findings lead us to suggest a set of intriguing directions for future aptitude research. Drawing on the L1 literature (McArthur & Bishop, 2005), we assumed that domain-general auditory processing could be considered as a rather *stable* trait. Indeed, the results of the current study found participants' auditory processing to be unrelated to the influence of any experience factors at least within the time framework of the current study (i.e., first eight months of study-abroad). However, there is some cross-sectional evidence that experience does play a role in individual differences in auditory processing, especially when researchers compare participants who have substantially different biographical backgrounds, such as total vs. non-tonal language users (e.g., Bidelman, Gandour, & Krishnan, 2011), musicians vs. non-musicians (Tierney, Krizman, Kraus, & Tallal, 2015), and sequential vs. simultaneous bilinguals (Krizman et al., 2015). Whereas we evaluated participants' audition ability only at the endpoint of the current project, we urge future studies to take a *longitudinal* look at the extent to which late bilinguals' linguistic *and* auditory abilities change over the course of long-term immersion.

To follow up our tentative hypothesis that immersion experience may enhance both auditory and linguistic skills, we recently conducted another longitudinal study with a very similar population (n = 50 Chinese international students in London) (Sun, Saito, & Tierney, forthcoming). In this project, we used both behavioural and neural measures to track their auditory processing ability at the outset and endpoint of their short-term immersion (three months), finding no significant change over time (despite the fact that some participants with more precise auditory precision demonstrated greater improvement in their L2 skills). Here, our follow-up study concurs with L1 acquisition literature that auditory processing is relatively stable (e.g., Hornickel et al., 2012), or that more extensive second language exposure or second language exposure early in life may be needed to enhance auditory processing (e.g., Bidelman et al., 2011). To further test the *plasticity* of auditory processing, we may need longitudinal studies with a much longer timeframe (e.g., 1-2 years), or training studies, wherein we can investigate the presence and absence of any change in auditory skills when participants receive hours of intensive auditory training (see Hayes, Warrier, Nicol, Zecker, & Kraus, 2003 for 35-40 hours of commercial auditory processing training programs).

Another topic worthy of attention concerns different types of auditory processing ability, i.e., the explicit-implicit distinction. In the field of L2 acquisition, some scholars have begun to emphasize that certain L2 learners can attain high-level L2 proficiency due to their ability to process language with (explicit learning) and without (implicit learning) awareness (Doughty, 2019). A growing number of studies have indicated that the incidence of high-level L2 grammar attainment is linked to a series of implicit, statistical learning mechanisms (e.g., Link et al., 2013; Suzuki & DeKeyser, 2017). It is interesting to note that L1 acquisition scholars have primarily measured the auditory processing phenomenon using two different methodological paradigms—behavioural and neurophysiological. Behavioral measures (e.g., formant, pitch and duration discrimination) are primarily used to tap into explicit auditory processing, while neurophysiological instruments are used to tap into implicit auditory processing. To our knowledge, little empirical investigation has been conducted on the relationship between explicit and implicit auditory processing. In the context of post-pubertal L2 speech learning, however, Diaz, Mitterer, Broersma, Escera, and Sebastian-Galles (2016) have provided preliminary evidence that behavioural and neurophysiological measures seem to tap into two different constructs of perceptual acuity, and that they can explain independent variance in post-pubertal L2 pronunciation learning (see also Kachlicka et al., 2019; Saito, Sun, & Tierney, 2019; Sun et al., forthcoming).

Given this promising finding, we call for future studies to interface L1 and L2 research perspectives and examine the role of explicit and implicit auditory processing in different types of L2 speech learning. In the present study we have provided evidence to suggest that explicit auditory processing may be instrumental to the initial stage of naturalistic L2 speech learning (eight-month of study abroad). It would be intriguing to further this line of inquiry by probing the extent to which explicit and implicit auditory processing can help even *highly experienced* L2 learners (those with 10+ years of immersion) ultimately improve their L2 pronunciation proficiency (cf. Saito et al., 2020).

The independent contribution of explicit and implicit auditory processing to L2 speech leaning can be further tested via the aptitude-interaction paradigm. While learners with explicit aptitude are hypothesized to benefit more from explicit, language-focused instruction, those with implicit aptitude are thought to make better use of implicit, meaning-focused instruction (DeKeyser, 2012). By adopting a quasi-experimental design, therefore, future studies could track how L2 learners with different explicit and implicit auditory processing abilities can differentially improve their proficiency when they engage in explicit (e.g., Saito & Plonsky, 2019 for articulatory and auditory training) and implicit pronunciation training (e.g., Lee & Lyster, 2016).

References

Amitay, S., Ahissar, M., & Nelken, I. (2002). Auditory processing deficits in reading disabled adults. *Journal of the Association for Research in Otolaryngology*, *3*, 302-320.

Baayen, R. H. (2008). Exploratory data analysis: an introduction to R for the language sciences. *Cambridge: Cambridge UP*.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-21.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and cognition, 77*(1), 1–10.

Bohn, O. S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition, 14*(2), 131–158.

Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing, 30*, 159–175.

Campbell, K. L., & Tyler, L. K. (2018). Language-related domain-specific and domain-general systems in the human brain. *Current Opinion in Behavioral Sciences*, *21*, 132–137.

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

Casini, L., Pech-Georgel, C., & Ziegler, J. C. (2018). It's about time: revisiting temporal processing deficits in dyslexia. *Developmental Science*, *21*, e12530.

Chandrasekaran, B., Kraus, N., & Wong, P. C. (2011). Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology*, *107*, 1325–1336.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*, 218–236.

DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, *62*, 189–200.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*, 533–557.

Diaz, B., Mitterer, H., Broersma, M., Escera, C., & Sebastian-Galles, N. (2016). Variability in L2 phonemic learning originates from speech-specific capabilities: An MMN study on late bilinguals. *Bilingualism: Language and Cognition*, *19*, 955-970.

Doughty, C. J. (2019). Cognitive language aptitude. *Language Learning*, *69*, 101–126.

EIKEN Foundation of Japan. (2016). EIKEN Pre–1 level: Complete questions collection. Tokyo: Oubunsha.

Flege, J. (2016, June). *The role of phonetic category formation in second language speech acquisition*. Plenary address delivered at New Sounds, Aarhus, Denmark.

Flege, J., Takagi, N., & Mann, V. (1995). Japanese adults learn to produce English /ɹ/ and /l/ accurately. *Language and Speech, 38*, 25–55.

Foster, P, Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language. *Applied Linguistics, 21*, 354–375.

Gibson, L. Y., Hogben, J. H., & Fletcher, J. (2006). Visual and auditory processing and component reading skills in developmental dyslexia. *Cognitive Neuropsychology*, *23*, 621-642.

Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S., & Wicha, N. Y. (2011). Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. *Frontiers in psychology*, *2*, 146.

Goswami, U. (2015). Sensory theories of developmental dyslexia: three challenges for research. *Nature Reviews Neuroscience, 16*, 43–54.

Goswami, U., Wang, H. L. S., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2011). Language-universal sensory deficits in developmental dyslexia: English, Spanish, and Chinese. *Journal of Cognitive Neuroscience*, *23*, 325-337.

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*, 311-343.

Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences*, *115*, 1487–1492.

Hornickel, J., Knowles, E., & Kraus, N. (2012). Test-retest consistency of speech-evoked auditory brainstem responses in typically-developing children. *Hearing research*, *284*, 52–58.

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, *34,* 475–505.

Jasmin, K., Dick, F., Holt, L., & Tierney, A. (2019). Tailored perception: Individuals' speech and music perception strategies fit their perceptual abilities. *Journal of Experimental Psychology: General*. DOI: 10.1037/xge0000688

Jasmin K, Sun H, Tierney A (2020) Effects of language experience on domain-general perceptual strategies. bioRxiv, 892943. DOI: 10.1101/2020.01.02.892943

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.

Kempe, V., Thoresen, J. C., Kirk, N. W., Schaeffler, F., & Brooks, P. J. (2012). Individual differences in the discrimination of novel speech sounds: effects of sex, temporal processing, musical and cognitive abilities. *PloS one*, *7*(11), e48623.

Koester, D., Gunter, T. C., Wagner, S., & Friederici, A. D. (2004). Morphosyntax, prosody, and linking elements: The auditory processing of German nominal compounds. *Journal of Cognitive Neuroscience*, *16*, 1647–1668.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental science, 9*(2), F13–F21.

Krizman, J., Slater, J., Skoe, E., Marian, V., & Kraus, N. (2015). Neural processing of speech in children is influenced by extent of bilingual experience. *Neuroscience Letters*, *585*, 48–53.

Lahmann, C., Steinkrauss, R., & Schmid, M. S. (2017). Speed, breakdown, and repair: An investigation of fluency in long-term second-language speakers of English. *International Journal of Bilingualism*, *21*, 228–242.

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition, 39*(1), 167–196.

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

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*, 3757-3768.

Levitt, H. C. C. H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, *49*, 467-477.

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition, 38*(4), 801–842.

Li, M., & DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition*, *39*, 593-620.

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., & Doughty, C. J. (2013). Hi-LAB: A New Measure of Aptitude for High-Level Language Proficiency. *Language Learning*, *63*, 530–566.

Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *The Quarterly Journal of experimental psychology*, *45*, 73-87.

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of phonetics, 30*(2), 229–258.

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.

Meara, P. (2005). Llama language aptitude tests: The manual. *Swansea: Lognostics*.

Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, *46*, 610-641.

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.

Muñoz, C., & Llanes, Á. (2014). Study abroad and changes in degree of foreign accent in children and adults. *The Modern Language Journal*, *98*, 432-449.

Munro, M., & Derwing, T. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, *23*, 451–468.

Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language learning*, *58*, 479–502.

Munro, M. J., Derwing, T. M., & Saito, K. (2013). English L2 vowel acqusition over seven years. In. J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*. Ames, IA: Iowa State University.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of cognitive neuroscience, 10*(6), 717–733.

Piske, T., Flege, J., MacKay, & Meador, D. (2011). Investigating native and non-native vowels produced in conversational speech. In M. Wrembel, M. Kul & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010* (pp. 195–205). Frankfurt am Main: Peter Lang.

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal*, *102*, 713-731.

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*, 509–527.

Ruggles, D., Bharadwaj, H., & Shinn-Cunningham, B. G. (2012). Why middle-aged listeners have trouble hearing in everyday settings. *Current Biology*, *22*, 1417–1422.

Saito, K. (2015a). Experience effects on the development of late second language learners' oral proficiency. *Language Learning, 65*, 563-595.

Saito, K. (2015b). The role of age of acquisition in late second language oral proficiency attainment. *Studies in Second Language Acquisition, 37*, 713-743.

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

Saito, K. (2019). The role of aptitude in second language segmental learning: The case of Japanese learners' English/ɹ/pronunciation attainment in classroom settings. *Applied Psycholinguistics*, *40*, 183–204.

Saito, K., & Brajot, F. X. (2013). Scrutinizing the role of length of residence and age of acquisition in the interlanguage pronunciation development of English/ɹ/by late Japanese bilinguals. *Bilingualism: Language and Cognition, 16*(4), 847–863.

Saito, K., Kachlicka, M., Sun, H., & Tierney, A. (2020). Domain-general auditory processing as an anchor of post-pubertal second language pronunciation learning: Behavioural and neurophysiological investigations of perceptual acuity, age, experience, development, and attainment. *Journal of Memory and Language*.

Saito, K., Ilkan, M., Magne, V., Tran, M. N., & Suzuki, S. (2018). Acoustic characteristics and learner profiles of low-, mid-and high-level second language fluency. *Applied Psycholinguistics, 39,* 593–617.

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learnin*g, 69, 652−708.

Saito, K., Sun, H., & Tierney, A. (2019). Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental, prosodic and temporal sensitivity and performance via behavioral and neurophysiological measures. *Bilingualism: Language and Cognition, 22*, 1123-1140.

Saito, K., Sun, H., & Tierney, A. (2020). Brief report: Test-retest reliability of explicit auditory processing measures. *bioRxiv.* https://doi.org/10.1101/2020.06.12.149484

Saito, K., Sun, H., Kachlicka, M., Robert, J., Nakata, N., & Tierney, A. (in press). Domain-general auditory processing explains multiple dimensions of L2 acquisition in adulthood. *Studies in Second Language Acquisition*.

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*, 439-462.

Schneider, B. A., Daneman, M., & Pichora-Fuller, M. K. (2002). Listening in aging adults: from discourse comprehension to psychoacoustics. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *56*, 139.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in second language acquisition*, *26*, 173–199.

Skehan, P. (2019). Language aptitude implicates language and cognitive skills. In Z. Weng, P. Skehan, A. Biedron, S. Li & R. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (Chapter 4). New York: Routledge.

Smith, J. O. (2007). *Introduction to digital filters: with audio applications* (Vol. 2). Julius Smith.

Sun, H., Saito, K., & Tierney, A. (forthcoming). A longitudinal investigation of explicit and implicit auditory processing in L2 segmental and suprasegmental acquisition. *Studies in Second Language Acquisition*.

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*, 747-790.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–277). Amsterdam: Benjamins.

Tierney, A. T., Krizman, J., Kraus, N., & Tallal, P. (2015). Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 10062–10067.

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*, 1-30.

VanPatten, B., & Smith, M. (2015). Aptitude as grammatical sensitivity and the initial stages of learning Japanese as a L2: Parametric variation and case marking. *Studies in Second Language Acquisition*, *37*, 135–165.

Hayes, E. A., Warrier, C. M., Nicol, T. G., Zecker, S. G., & Kraus, N. (2003). Neural plasticity following auditory training in children with learning problems. *Clinical neurophysiology, 114*(4), 673–684.

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual review of psychology*, *50*, 509–535.

White-Schwoch, T., Carr, K. W., Thompson, E. C., Anderson, S., Nicol, T., Bradlow, A. R., Zecker, S., & Kraus, N. (2015). Auditory processing in noise: A preschool biomarker for literacy. *PLoS biology*, *13*, e1002196.

Won, J. H., Tremblay, K., Clinard, C. G., Wright, R. A., Sagi, E., & Svirsky, M. (2016). The neural encoding of formant frequencies contributing to vowel identification in normal-hearing listeners. *The Journal of the Acoustical Society of America*, *139*, 1–11.

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

Yalçın, Ş., & Spada, N. (2016). Language aptitude and grammatical difficulty. *Studies in Second Language Acquisition*, *38*, 239–263.

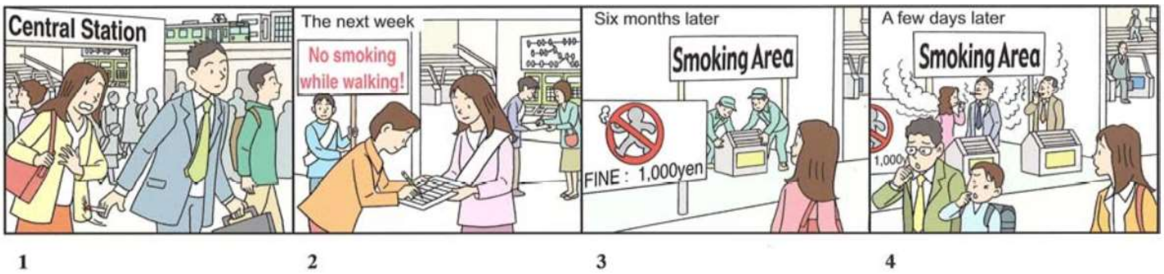
Yilmaz, Y. (2013). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics*, *34*, 344-368.

Zuniga, M., & Simard, D. (2019). Factors influencing L2 self-repair behavior: The role of L2 proficiency, attentional control and L1 self-repair behavior. *Journal of Psycholinguistic Research*, *48*(1), 43-59.

Supporting Information-A: Speaking Tasks (Versions A, B, C and D)

Version A

You have one minute to prepare. This is a story about a woman who wanted to stop people from smoking on the street. You have two minutes to narrate the story. Your story should begin with the following sentence: *One day, a woman was on her way to work.*

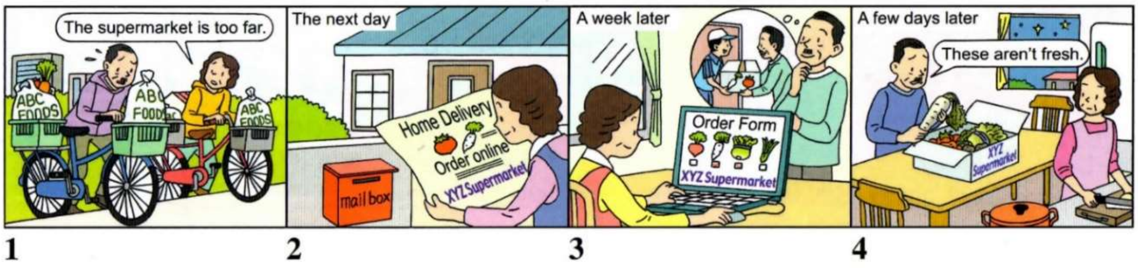


Version B

You have one minute to prepare. This is a story about a couple who ran an open-air restaurant on a beach. You have two minutes to narrate the story. Your story should begin with the following sentence: *One day, a couple was at their restaurant.*

Version C

You have one minute to prepare. This is a story about an elderly couple who lives far away from the nearest supermarket. You have two minutes to narrate the story. Your story should begin with the following sentence: *One day, an elderly couple was coming home from the supermarket.*



Version D

You have one minute to prepare. This is a story about a girl who wanted a smartphone. You have two minutes to narrate the story. Your story should begin with the following sentence: *One day, a girl was at home with her parents.*

All the materials adapted from EIKEN Foundation of Japan. (2016). EIKEN Pre-1 level: Complete questions collection. Tokyo: Oubunsha.

Supporting Information-B: Onscreen Labels for Pronunciation Measures

1. Vowel and/or consonant errors

Frequent		**Infrequent or absent**

2. Word stress errors affecting stressed and unstressed syllables

Frequent		**Infrequent or absent**

3. Intonation (i.e., pitch variation)

Too varied or not varied enough		**Appropriate across stretches of speech**

4. Perceived tempo

Too slow or too fast		**Optimal**

Supporting Information-C: Training Materials for Pronunciation Measures

Segmental errors	This refers to errors in individual sounds. For example, perhaps somebody says "*road" "rain"* but you hear an "l" sound instead of an "r" sound. This would be a consonant error. If you hear someone say "*fan*" "*boat*" but you hear "*fun*" "*bought*," that is a vowel error. You may also hear sounds missing from words, or extra sounds added to words. These are also consonant and vowel errors.
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.
Perceived tempo	Perceived tempo is simply how quickly or slowly someone speaks. Speaking very quickly can make speech harder to follow, but speaking too slowly can as well. A good speech rate should sound natural and be comfortable to listen to.

Adapted from Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics, 38,* 439-462.

Supporting Information-D: R Code for Mixed Effects Modeling Analyses

We constructed a set of linear mixed-effects models using the lmer functions from the lme4 package (Version 1.1-21; Bates, Maechler, Bolker, & Walker, 2015) in the R statistical environment (R Core Team, 2018). To this end, the following code was used:

- Model <- lmer(formula = Pronunciation_Proficiency ~ 1 + Time + Time: Formant_Discrimination + Time: Pitch_Disctimination + Time: Duration_Disctimination + Time: Recent_L2_Use + (1|ID), data = data)

Dependent variables: "Pronunciation_Proficiency" refers to the participants' pronunciation scores at T1 and T2. A total of six separate models were created for segmentals, word stress, intonation, perceived speed, pause ratio, and articulation rate, respectively.

Fixed effects: Each model consists of (a) one main effect of Time, (b) three interaction effects of Time and auditory processing scores ("Formant_Discrimination," "Pitch_Discrimination," "Duration_Discrimination"), and (c) one interaction effect of Time and the amount of participants' L2 use throughout the project ("Recent_L2_Use"). For Time, "-.5" was given to participants' pronunciation performance at T1 (the beginning of the project), and ".5" was given to their performance at T2 (the end of the project). When the participants' increased pronunciation scores were considered as improvement (segmentals, word stress, intonation, optimal speed, articulation rate), the estimate could be positive. When their decreased pronunciation scores were considered as improvement (pause ratio), the estimate could be negative).

Random effects: Participants' ID (1-30) were entered into the model as random effects.