Check for updates

Original Article

second language research

Second Language Research
1–21
© The Author(s) 2020

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0267658320978493
journals.sagepub.com/home/slr
SAGE

Successful second language pronunciation learning is linked to domain-general auditory processing rather than music aptitude

Chaoqun Zheng

Kazuya Saito
University College London, London, UK

Adam Tierney
Birkbeck, University of London, London, UK

Abstract

Whereas a growing amount of attention has been directed to the role of perceptual-cognitive aptitude in successful second language (L2) lexicogrammar learning, scholars have begun to investigate the same topic in the context of L2 pronunciation learning. To date, there is ongoing discussion on the mechanism underlying L2 speech learning in relation to music aptitude and domain-general auditory processing ability. Situated within 48 moderately experienced Chinese learners of English in the UK (length of residence = one year), the current study examined the relationship between music aptitude, auditory perception, and L2 pronunciation proficiency. Results revealed that music aptitude and auditory processing were partially overlapping, whereas both abilities were independent of participants' past and current L2 language learning experience. Whereas individual differences in auditory processing demonstrated significant associations with various dimensions of L2 pronunciation proficiency, music aptitude was only weakly predictive of prosodic aspects of L2 pronunciation proficiency. Comparatively, none of the experience variables were related to acquisition within the current dataset.

Keywords

aptitude, auditory processing, immersion, music, second language pronunciation

Corresponding author:
Kazuya Saito, Institute of Education, University College London, 20 Bedford Way, London, WC1H 0AL, UK
Email: k.saito@ucl.ac.uk

I Introduction

Learning a second language (L2) after puberty is well-known to be subject to a great deal of individual variation. Even if two individuals spend the same amount of time practicing a target language, their final outcomes may differ greatly. This could be in part due to the fact that certain individuals are perceptually and cognitively adept at making the most of every practice opportunity (i.e. they possess higher L2 learning aptitude), resulting in more advanced L2 proficiency (Doughty, 2019). Whereas scholars have extensively examined which perceptual-cognitive abilities relate to successful L2 learning, most of the existing literature has been exclusively concerned with lexicogrammar aspects of language learning (Li, 2016). In light of the ongoing discussion regarding the similarities between speech, music and language learning (Tierney et al., 2015), we highlight two overlapping abilities, music aptitude and auditory processing, as a framework of aptitude relevant to successful L2 pronunciation learning. Music aptitude is defined as a set of composite, domain-specific abilities to remember and reproduce music phrases that are no longer physically present, generally measured through standardized tests, and found to relate to L2 pronunciation development to some degree (Slevc and Miyake, 2006). More specifically, scholars have been interested in domain-general sensitivity to more fine-grained properties of acoustic signals (formants, pitch, duration, amplitude), which we refer to here collectively as auditory processing. The ability has been linked to first language (L1) acquisition (e.g. Goswami, 2015), and to L2 acquisition (e.g. Kachlicka et al., 2019). To test whether music aptitude and auditory processing abilities explain variation in adult L2 pronunciation learning, we investigated the complex relationship between music aptitude, auditory processing, and biographical profiles of 48 Chinese learners of English in the UK.

II Background

1 Second language pronunciation and aptitude

On a broad level, second language pronunciation proficiency comprises one's ability to produce new individual sounds without L1 substitutions (segmental proficiency), form words and sentences with adequate stress patterns (prosodic proficiency), and deliver speech at an optimal speed (temporal proficiency). According to Saito and Plonsky's (2019) framework, L2 pronunciation proficiency can be further considered as a multifaceted phenomenon characterized by different constructs of measurement focus (global vs. specific), scoring method (human judgments vs. acoustic analyses), and processing type (controlled vs. spontaneous). In the current investigation, L2 pronunciation proficiency was assessed via expert raters' judgments of the accurate and fluent use of segmentals and prosody, when L2 learners' speech was elicited via a picture narrative task (see the sections of Current Study). As per Saito and Plonsky's (2019) framework, L2 pronunciation proficiency as referenced in the current article concerns global, subjective and spontaneous constructs of L2 pronunciation abilities. In terms of fluency judgments, raters were trained to pay primary attention to temporal aspects of L2 speech. Traditionally, fluency is conceptualized via acoustic analyses of speech properties related to speed

(speech and articulation rate) and breakdown (filled and unfilled pauses) (Segalowitz, 2016), or native listeners' perception of optimal speed (i.e. not too slow or fast; Munro and Derwing, 2001). Since our raters assessed fluency by using the rubric of optimal speed (i.e. perceived fluency; Bosker et al., 2013), the terms 'fluency' and 'optimal speed' are used interchangeably for the rest of the article.

From theoretical standpoints, L2 speech learning initially takes place on a perception level. New phonetic categories are formed when L2 learners can become capable of distinguishing multiple acoustic dimensions of L2 sounds (e.g. the height, contour and length of pitch and formants) from L1 counterparts. Subsequently, such perception-based categories stimulate relevant motor movements to produce these sounds (Flege and Bohn, in press). To learn L2 pronunciation in a more efficient and effective fashion, therefore, the relevant abilities are considered to comprise two broad constructs: (1) perceiving spectral and temporal features in acoustic signals (perceptual acuity); and (2) sequencing and timing motor actions to produce these sounds (audio-motor integration). For a similar discussion of the relationship between perceptual acuity, audio-motor integration, and L1 speech acquisition, see Tierney et al. (2015).

To date, there is ample evidence that L2 learners continue to improve their pronunciation accuracy and fluency as they receive more input through more interaction opportunities in a target language (Derwing and Munro, 2013). However, these experience-related factors alone cannot fully explain the outcomes of L2 pronunciation learning in the long run. Examining the linguistic and biographical profiles of late Japanese-English bilinguals in Canada, for example, Saito (2015) showed that only 20%–30% of variance in their L2 English proficiency was explained by length of immersion. Many scholars have argued that certain individuals are more perceptually and cognitively adept at internalizing input and output, resulting in more gains even within a limited amount of immersion experience (Doughty, 2019). While much scholarly attention has been given towards exploring what comprises such aptitude for L2 lexicogrammar learning (for a meta-analysis, see Li, 2016), some have begun to investigate the topic in the context of L2 pronunciation learning.

Some form of aptitude has been suggested to relate to the process and product of L2 pronunciation. For example, Reiterer and her colleagues have shown that phonological working memory is a key construct of the initial phase of novel sound learning as it is directly tied to the activation of the left supramarginal gyrus and Broca's area (e.g. Reiterer et al., 2011). As learners have gained more experience, other cognitive (and domain-specific) abilities, such as phonemic coding, seem to play a key role in determining the acquisition of more advanced L2 proficiency, evidenced in the activation of speech motor control and auditory-perceptual areas in the brains (e.g. Hu et al., 2013).

However, it is important to point out that these abilities feature and confound a range of perceptual-cognitive constructs spanning phonological sensitivity, memory, awareness, analysis, and reproduction. Following the literature in cognitive psychology and L1 acquisition (e.g. Tierney et al., 2015) and the emerging paradigm in L2 acquisition (e.g. Kachlicka et al., 2019), we operationalized phonetic aptitude as auditory acuity and audio-motor integration. The former was measured via the auditory processing tests (i.e. AXB discrimination); and the latter was measured via the music aptitude tests (i.e. tone reproduction).

In the current investigation, we specifically focused on pronunciation as an outcome measure, as opposed to lexicogrammar or other aspects of language. This was because we expected the relationship between auditory processing and L2 pronunciation to be particularly strong. For other aspects of language (e.g. lexicogrammar), the acoustic signal is just one of many possible sources of information about linguistic structure (orthographic information for reading). However, pronunciation requires participants to precisely perceive characteristics of sound so that they can be produced.

2 Domain-general auditory processing

Auditory processing comprises one's ability to encode, represent, and internalize various dimensions of sounds (formants, fundamental frequencies, duration, and amplitude). In the current study, we focused on one component of auditory processing (i.e. perceptual acuity). This ability was measured using an AXB discrimination task (see the sections of Current Study). Individual differences in auditory processing have been found to be associated with the incidence of specific language impairment and developmental dyslexia (e.g. Casini et al., 2018; Goswami et al., 2011). This relationship suggests that auditory deficits may prevent learners from encoding phonetic, phonological and morphosyntactic information from incoming aural input in an efficient and effective manner, which in turn could lead to more global problems in the long run (Kraus and Chandrasekaran, 2010). A similar mechanism could function in second language learning, potentially leading to a relationship between auditory processing skills and language outcomes: the ability to precisely encode auditory input may be a bottleneck for the establishment of knowledge about segmental and suprasegmental linguistic categories.

Interestingly, there is a growing amount of evidence that auditory processing can explain variance in the outcomes of post-pubertal L2 speech perception learning through laboratory training (Lengeris and Hazan, 2010) and naturalistic immersion (Kachlicka et al., 2019; Saito et al., 2020d). A growing amount of evidence has also shown that the development of such audition effects can be generalized to the development of L2 pronunciation proficiency (Saito et al., 2020a, 2020c). These findings are in line with several influential theoretical accounts of L1 and L2 speech acquisition. For example, the Speech Learning Model states that the same mechanisms used for L1 acquisition are active throughout one's lifespan, and germane to post-pubertal language learning (Flege and Bohn, in press). Building on this line of thought, it is reasonable to hypothesize that domain-general auditory processing, which prior work has linked to L1 acquisition and delay, may serve as a bottleneck of L2 speech acquisition in adulthood.

3 Music aptitude

Music aptitude has been measured via a composite battery of perception and production tasks (e.g. the Wing Measures of Musical Talents; Wing, 1968). For the perception tasks, participants listened to and discriminated two musical phrases that could differ in tone, intensity, rhythm, timbre, and timing. They were presented in a similar format to the aforementioned auditory processing tests. For the production tasks, participants listened to, remembered, and replicated (sang or played back) musical phrases. In the current study, while auditory processing tasks were used to tap into participants' perceptual

acuity, the production component of the music aptitude test (i.e. tone reproduction) was used to index participants' audio-motor integration abilities (see the sections of Current Study).

The association between musical aptitude and L2 speech learning has long been discussed. One explanation is based on the shared perceptual-cognitive mechanisms between music and language learning (Milovanov et al., 2010; Patel, 2003), such as the need for precise auditory processing (Slevc and Miyake, 2006). Indeed, there is ample empirical evidence that musicians (e.g. Schellenberg, 2015) and individuals with higher music aptitude (e.g. Strait et al., 2011) may demonstrate more precise perception of certain acoustic dimensions.

To date, previous empirical research has typically investigated the relationship between musical aptitude and non-native phonological competence in laboratory settings. It has been shown that music aptitude could be tied to both segmental and suprasegmental performance in a novel/foreign language on both perception (e.g. Delogu et al., 2010; Li and DeKeyser, 2017) and production levels (e.g. Milovanov et al., 2010; Pei et al., 2016).

Notably, very few studies have ever delved into the relationship between music aptitude and perception and production of naturalistic L2 speech (for a critical review, see Trofimovich et al., 2015). Slevc and Miyake (2006) compared the relationship between music aptitude, phonological short-term memory, and L2 English proficiency in 50 Japanese residents with varied immersion experience in the USA. The results showed that music aptitude explained 8%–12% of the variance in participants' speech perception and production abilities. Focusing on 48 Chinese–English bilinguals in the UK, Saito et al. (2019) similarly found that those with greater music aptitude likely attained more fluent, advanced L2 pronunciation proficiency.

In the current study, we revisited the role of music aptitude in naturalistic L2 speech learning. More specifically, we examined how music aptitude (operationalized via the tone reproduction task) could be associated with auditory processing (operationalized via the AXB discrimination task) and how both music aptitude and auditory processing can differentially relate to L2 pronunciation proficiency (i.e. the production correlates of L2 speech acquisition).

III Current study

Focusing on $N = 48$ L1 Mandarin Chinese users of English in the UK, the current study scrutinized the relationship between auditory processing, music aptitude, L2 learning experience (age of learning, past/pre-departure L2 learning experience in China, current L2 use in the UK), and L2 pronunciation proficiency. Two research questions were formulated:

- Is there a significant association between domain-general auditory perception (formant, pitch, and duration discrimination), music aptitude (perception, memorization and reproduction of melody and rhythm) and L2 learning experience (onset and length of practice)?
- Are auditory perception, music aptitude and L2 English learning experience factors significantly related to segmental and suprasegmental dimensions of L2 pronunciation proficiency?

1 Participants

Originally, 50 native speakers of Mandarin Chinese with similar length of residence (LOR) backgrounds (around 1 year) were recruited from a university in London. Later, two participants were eliminated from the dataset as outliers. One participant demonstrated unusually deviated auditory processing scores (standardized value > 2), which we will detail in the Auditory Processing section. Another participant demonstrated extensive immersion experience (> 2 years). Therefore, the valid sample constituted 48 participants aged from 22 to 29 ($M = 23.77$; $SD = 1.79$). All the participants were graduate students at the time of the project (majoring in different programs in social sciences, such as education and psychology). Before studying in London, none had been abroad for more than a month. Thus, their LOR in English-speaking countries (the UK) was homogeneous ($M = 9.15$ months; $SD = 1.52$; *Range* 8–12 months). According to self-reports, they had been learning English in a learning context of English as a foreign language (EFL) for around six to 21 years ($M_{age\ of\ learning} = 8.73$ years; $SD = 2.66$; *Range* $= 4$–16 years). None of them reported a hearing problem or any English pronunciation training.[1] Participants were not asked if they were diagnosed with language impairments in childhood.

2 L2 pronunciation proficiency

a Speaking materials. Traditionally, scholars have exclusively relied on controlled speaking tasks to analyse L2 pronunciation proficiency (e.g. for word and sentence reading, Slevc and Miyake, 2006). However, such methodological practice has been questioned: Controlled tasks of this kind, which minimize the semantic and syntactic demands of production, allow adult L2 learners to carefully monitor the correctness of their pronunciation, which may not index their ability to produce correct pronunciations in daily-life settings (Piske et al., 2011). To elicit participants' more *spontaneous* speech, a picture narration task was adapted from the EIKEN English Test Pre-Grade 1 Level (EIKEN, 2016). To eliminate the effect of materials on speaking performance, two versions were used (Versions A and B).[2] Half of the participants were randomly assigned to Version A, while the other half to Version B. Each version comprised a four-frame picture. For each task, participants had fixed planning time (i.e. one minute to prepare and two minutes to speak). To keep the speeches on the right track, the first sentence of each story was given. All the speech samples were recorded in a quiet room using Praat with a 44,100 Hz sampling rate. Following the research standard in L2 pronunciation research (e.g. Isaacs and Trofimovich, 2012), the first 30 seconds of each recording were cut and saved as a WAV file as a representation of the whole speech. These speech samples were then submitted to expert raters for subjective ratings.

b Expert rating. Following the L2 pronunciation proficiency assessment procedure developed and validated in Saito et al. (2017), we adopted the expert judgment approach by which to examine four different aspects of L2 speech: (1) segmentals (consonantal and vocalic accuracy), (2) word stress (correct assignment of emphasis in multisyllabic words), (3) intonation (adequate and varied intonation), and (4) optimal speed (not too fast nor slow speech rate).

c *Procedure.* A total of three female native speakers of English were recruited. While two of three raters were originally from the US, they had resided in the UK for their MA degree in TESOL. The other rater was originally from the UK. All of them held Certificate of English Language Teaching to Adults, and reported extensive experience in teaching English ($M_{\textit{years of teaching}}$ = 9.7 years), including EFL experience in China ($M_{\textit{years of teaching in China}}$ = 1.7 years). They demonstrated a relatively strong familiarity with Chinese-accented English speech (M = 5.3) on a 6-point scale (1 = not familiar at all, 6 = very familiar). None of them reported hearing problems.

Each rater received instruction from a trained researcher (the first author of the current project). First, they were given detailed explanation on the four different categories of L2 pronunciation proficiency- segmentals, word stress, intonation, and optimal speed (for training scripts, see Appendix 1). Second, they practiced the rating procedure with three practice samples (not included in the main dataset). For each response, the raters were asked to justify their decisions. After the researcher ensured that the raters fully understood the procedure, they moved onto the judgments of 48 speech samples.

All the samples were played in a randomized order via a MATLAB-based program. Upon hearing each sample, the raters were asked to rate for segmentals, word stress, intonation and optimal speed by moving a slider. Depending on where a cursor was located, pronunciation ratings were automatically recorded on a 1,000-point scale (0 = not targetlike, 1,000 = targetlike). The raters were encouraged to adjust their ratings until they felt satisfied with their judgments. The entire session took about one hour.

Inter-rater reliability was checked after the rating sessions were completed. According to the results of Cronbach alpha analysis, there was a high agreement as to word stress (α = .762) and optimal speed (α = .802). Nonetheless, the alpha level was relatively low as to segmentals (α = .698) and intonation (α = .646). According to Larson-Hall's (2010) field-specific standard, these alpha values were slightly lower than the acceptable level (i.e. α = .70). As a remedy, the raters engaged in another follow-up session in which they focused on listening to five samples on which they had initially showed disagreement, discussed their judgments (segmentals, word stress, intonation, optimal speed), and gave out new scores. Ultimately, the Cronbach alpha rose to α = .763 for segmentals, α = .756 for word stress, α = .732 for intonation, and α = .809 for optimal speed. For the subsequent analyses, the raters' scores were averaged across raters, generating one score for each token as per the four different dimensions of L2 pronunciation proficiency.[3]

3 Auditory processing measures

Following the literature in cognitive psychology and L1 acquisition (e.g. Surprenant and Watson, 2001), and using the same procedure developed in Kachlicka et al. (2019) and validated in Saito et al. (2020b), participants' auditory processing ability in this test was examined via three different types of psychoacoustic AXB discrimination tests: formant, pitch and duration discrimination. In this current investigation, we used the same MATLAB-based test materials in Kachlicka et al. (2019), wherein participants listened to three non-verbal sounds, and chose which one sounded different from the other samples (1st vs. 3rd) by either pressing the number '1' or '3' on a keyboard (for

more methodological details, see Kachlicka et al., 2019). For each of the three tasks (i.e. formant, pitch, and duration discrimination) we created a total of 100 synthesized complex tone stimuli.

For formant discrimination, three formants were created and set at 500 Hz (F1), 1,500 Hz (F2), and 2,500 Hz (F3). The frequency of the second formant (F2) ranged from 1,502 Hz to 1,700 Hz with an increment of 2 Hz. For duration and pitch discrimination, a standard four-harmonic complex tone was created with F0 at 330 Hz. The target acoustic dimension for each test ranged with a step of 2.5 ms in duration (252.5–500 ms) and 0.3 Hz in F0 (330.3–360 Hz), respectively. Using Levitt's (1971) adaptive discrimination procedure, the level of difficulty changed from trial to trial according to participants' performance. The tests started from Level 50 (out of 100). The test began by presenting at stimulus level 50, and it became easier (bigger difference) by 10 steps after an incorrect response, or became more difficult (smaller difference) by 10 steps after every third correct response. The step size changed when a reversal happened. This could be due either to a participant getting the correct answer right after a string of wrong answers, or to a participant getting the answer wrong after a string of correct answers. After a first reversal, the step size changed to 5, and then it changed to 2 after a second reversal. After a third reversal, the step size remained at 1 until the end of the test. Each test stopped after 70 trials or 8 reversals.

4 Music aptitude measures

a Materials. The music aptitude test battery was developed as a part of the Wing Measures of Musical Talents (Wing, 1968), and modified, tested, and validated among Chinese learners of English by Pei and Ting (2013). The materials here are very much similar to those used in Slevc and Miyake (2006). The test was designed to tap into participants' abilities to perceive, memorize, and reproduce melodic and rhythmic aspects of music (for similar melody and rhythm tests used in L2 speech research, see Li and DeKeyser, 2017).

For the melody test, three melodies were created, each of a different length: one consisted of three notes, another of five notes, and the longest of seven notes (for an example of one of these melodies, see Figure 1). Each note was 0.7 seconds in duration. Hence, there were a total of 15 notes across the three melodies. For all the melodies used in the current study, see Appendix 1.

Similar to the melody test, the rhythm test consisted of three stimuli, each containing a sequence of notes with the same F0 (392 Hz) but different durations. Each stimulus was four measures in duration, for a total of 12 measures across the whole test. These three rhythms had three different time signatures: 2/4, 3/4, and 4/4. The duration of each stimulus was exactly 5 seconds; for an example of one rhythmic pattern, see Figure 2. For all the rhythm patterns used in the current study, see Appendix 1.

b Procedure. During test sessions, each of the six musical patterns was played twice. After listening to the patterns using earphones in a quiet room, participants were asked to reproduce the melodies or rhythms by singing 'la la la' as accurately as possible. All the singing patterns were recorded by Praat, and then saved as WAV files. The recordings

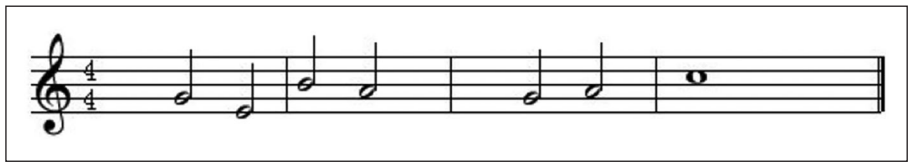

Figure 1. Example of test melody.

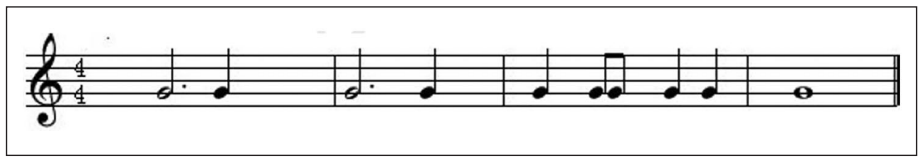

Figure 2. Example of staves for a rhythm pattern.

of this study were sent to expert human raters for scoring. Two female musically trained coders were recruited to evaluate the productive music aptitude tests ($M_{age} = 23.5$). Both raters were graduates from the Shanghai Conservatory of Music, and they specialized in piano and had been learning piano for more than 15 years.

The participants' music productions were scored by the two raters respectively. Before the evaluation session, the criteria were explained to raters: (1) For the tonality test, each accurate production of the relative F0 of a note (relative to previous or subsequent notes) earned one point, while the rhythm was disregarded. However, a globally lower or higher pitch was allowed, given that different participants had different voice ranges. In other words, raters were encouraged to focus on the correctness of the relative tonal contours. (2) For the rhythm test, participants could receive one point when they correctly produced all note durations in a bar while disregarding the F0 of the notes. The assessment sessions took place in a quiet place, and lasted for approximately three hours, during which time raters could take a rest whenever they requested.

For the melody test, the first sample had three notes, the second sample had five notes, and the third sample had seven notes. In total, there were 15 (3+5+7) notes. Since the scores were all relative evaluations, each note was evaluated against one of the notes in the test as a standard note. The standard note was usually the first note, if a participant produced it correctly. If the production of the first one was inaccurate, the second note would be the standard note. For example, a participant who produced a melody with the note of 'Si Sol Do Mi Re' when imitating 'La Sol Do Mi Re' could get four points. Since the standard score was also worth one point, the minimum score was three and the maximum possible score was 15 points.

For the rhythm test, the total number of measures for each sample was four, and in total there were 12 measures, so that the maximum possible score was 12 points. The inter-rater reliability was checked through Cronbach alpha analyses, identifying a high agreement for both the melody ($\alpha = .837$) and rhythm subtests ($\alpha = .902$). Scores from the two raters were then averaged to create an index of each participant's melody aptitude and rhythm aptitude.

Table 1. Descriptive statistics of second language segmental and suprasegmental scores.

	M	SD	Range
Segmentals	422.12	122.53	222.67–712.00
Stress	482.34	104.35	229.67–761.67
Intonation	479.79	115.50	238.67–753.67
Optimal speed	562.13	142.99	304.33–824.33

Table 2. Interrelationships between four constructs of pronunciation.

	Stress		Intonation		Optimal speed	
	r	*p*	*r*	*p*	*r*	*P*
Segmentals	.757*	< .001	.647*	< .001	.369*	< .010
Stress			.621*	< .001	.469*	< .001
Intonation					.571*	< .001

Note. * for statistically significant ($p < .016$) (Bonferroni corrected).

IV Results

1 Pronunciation, aptitude and experience profiles

a Pronunciation profiles. Table 1 summarizes the descriptive statistics of participants' pronunciation proficiency, assessed by three raters according to a subjective rating method in terms of segmental (vowel and consonant accuracy) and suprasegmental (word stress, intonation, optimal speed) dimensions. According to the score range, it is clear that participants' phonological competencies were subject to a great deal of individual variability. The distributions of pronunciation scores were checked by a set of Kolmogorov–Smirnov tests. According to the results, the four dimensions of pronunciation scores were all normally distributed ($p > .05$).

To delve into the relationship among the four pronunciation measures, a set of Pearson correlation analyses were conducted. Due to multiple comparisons (i.e. each measure was compared to the other three measures), the alpha level was set to $p < .016$ according to Bonferroni corrections. As demonstrated in Table 2, there were three emerging patterns: (1) three dimensions of segmental and prosodic accuracy were strongly correlated with each other ($r = .621$–$.757$), (2) optimal speed was moderately related to segmentals and word stress ($r = .369$, $.469$, respectively), and (3) there was a medium-to-strong relationship between intonation and optimal speed ($r = .571$). In light of the strength of the correlation coefficients, the four measures of pronunciations were assumed to tap into two broad domains of L2 pronunciation proficiency: (1) accurate articulation of sounds, words and sentences (segmentals, word stress, and intonation); and (2) fluent delivery of speech with adequate and varied intonation (intonation, optimal speed).

b Auditory processing, music aptitude and experience profiles. In order to answer RQ1 (the relationship between auditory processing, music aptitude, and L2 learning experience),

Table 3. Descriptive statistics of auditory processing, music aptitude scores, and second language learning experience.

	M	SD	Range
A. Auditory processing:			
Pitch discrimination	5.59 Hz	3.50	1.4–16.7
Duration discrimination	57.19 ms	34.86	11–167.08
Formant discrimination	66.96 Hz	30.77	10–121
B. Music aptitude:			
Melody production	63.40%	2.52	36.67–100
Rhythm production	73.18%	2.13	29.17–100
C. Experience:			
Age of learning	8.73 years	2.66	4–16
Total speaking hours in UK (Current Use)	476.33 hours	527.42	40–3,072
Total EFL learning hours in China (Past Experience)	9,340.50 hours	4,497.51	2,288–22,984

we first present the results of descriptive statistics, and then the results of correlation analyses. Table 3 summarizes the raw scores of the three auditory processing scores (formant, pitch, duration), the two aptitude tests (melody, rhythm), and L2 learning experience profiles. Whereas auditory processing indexes the smallest acoustic differences that participants could perceive (smaller is better), music aptitude reflects how accurately they could reproduce melodic and rhythmic information (greater % is better). The participants' L2 English learning experience demonstrated a wide range of variation. The results suggest that although the participants were recruited from a similar cohort (international students at a university in London), the current dataset comprised relatively heterogeneous experience profiles, and, by extension, proficiency levels.

Since the pitch and duration discrimination scores were skewed (Kolmogorov–Smirnov tests, $p < .05$), they were transformed using a log10 function, after which they were approximately normally distributed. A set of Pearson correlation analyses were conducted to probe the interrelationships between auditory processing, music aptitudes and experience. The alpha value was set to $p < .007$ (each factor was compared with the other seven factors). As shown in Table 4, whereas the correlation coefficients did not reach statistical significance among the three auditory processing measures (formant, pitch, duration), both of the music aptitude (melody, rhythm) scores were significantly correlated with each other ($p < .007$). Not surprisingly, there was some significant overlap between auditory processing (pitch discrimination) and music aptitude (melody production) ($r = –.455$, $p = .001$). It is probably more important and intriguing to point out that certain aspects of auditory processing and music aptitude were unrelated and independent. For example, none of the instances related to formant discrimination and rhythm production reached statistical significance.

2 Predictors of L2 pronunciation proficiency

To answer the second research question, we examined how participants' profiles of auditory processing, music aptitude and L2 learning experience jointly interacted to relate to

Table 4. Correlations between auditory processing, music aptitudes, and experience.

	Auditory processing								Experience factors					
	Pitch discrimination		Duration discrimination		Formant discrimination		Melody production		Rhythm production		Age of learning		Past experience		Current L2 use	
			r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Pitch discrimination			.085	.568	.327	.023	–.455*	.001	–.250	.087	–.151	.307	–.017	.907	–.066	.655
Duration discrimination					.027	.855	–.187	.202	.044	.767	.002	.990	.052	.727	.058	.694
Formant discrimination							–.361	.012	–.287	.048	.158	.284	.063	.672	.040	.789
Melody production									.390*	.006	.116	.431	–.121	.414	.067	.650
Rhythm production											–.243	.095	–.014	.924	.162	.273
Age of learning													.011	.941	–.093	.529
Past experience															.230	.116

Notes. Current = current (in the UK) English learning hours. Past = past (in China) English learning hours. * for statistically significant ($p < .007$) (Bonferroni corrected).

L2 pronunciation scores via a set of correlation and multiple regression analyses. Notably, our sample size ($n = 48$) was too small relative to a total of eight potential predictors (resulting in a weak power of .585). To conduct robust statistical analyses, we used the following two-step process. First, we conducted a set of Pearson correlation analyses to look at the overall relationship between the eight predictors and L2 pronunciation proficiency scores. Subsequently, we identified and entered only significant or marginal predictors into multiple regression models.

a Correlations analyses. A set of Pearson correlation analyses were performed with the eight predictors (pitch discrimination, duration discrimination, formant discrimination, melody production, rhythm production, age of acquisition, past experience, current L2 use) and L2 pronunciation scores (segmentals, word stress, intonation, speed). The alpha level was set to .0125 (each independent variable was compared against four variables). As summarized in Table 5, pitch discrimination was statistically associated with intonation ($r = –.419$, $p = .003$), and marginally connected with segmental pronunciation ($r = –.357$, $p = .013$). Duration discrimination also had a statistically significantly negative correlation with segmental pronunciation ($r = –.389$; $p = .006$), and a marginal connection with stress ($r = –.340$, $p = .018$); for scatterplots displaying the relationships between pitch discrimination and intonation rating and between duration discrimination and segmentals rating, see Figure 3. Formant discrimination, on the other hand, had no significant relationship with any dimensions of pronunciation. With respect to the relationship between music aptitude and pronunciation, there was a trend for melodic aptitude to be linked to segmentals, stress and intonation, and for rhythmic aptitude to be connected with stress and fluency, but no significant relationships emerged. No statistically significant correlation was found between formant discrimination, experience, and participants' pronunciation. None of the experience variables were significantly associated with any of the L2 pronunciation scores (for the non-linear relationship between experience and L2 outcomes, see Doughty, 2019; Saito, 2015).

b Multiple regression analyses. To further examine the relative importance of aptitude factors (auditory processing vs. music aptitude) in L2 pronunciation proficiency attainment, a set of stepwise multiple regression analyses were performed with participants' pronunciation scores as dependent variables. Due to the small sample size ($n = 48$), we reduced the number of predictors. Given that formant discrimination and experience-related factors did not show any significant associations with L2 pronunciation scores, they were eliminated in the subsequent analyses. According to the results of power analyses, the sample size of 48 participants together with four predictors (pitch discrimination, duration discrimination, rhythm production, melody production) gained the power of .751, which is above the field-specific benchmark of power size (Larson-Hall, 2010 for .700).

As summarized in Table 6, duration and pitch discrimination were significant predictors for segmental pronunciation. Duration and pitch discrimination could altogether explain around 22.4% ($R^2 = .224$) of the variance of segmental articulation, which can be considered moderate in the SLA field ($R^2 = .18–.51$) based on the guidance provided by Plonsky and Ghanbar (2018). For word stress, duration discrimination ($R^2 = .097$)

Table 5. Inter-relationships between L2 phonological competences, aptitudes and experiences.

	Auditory processing						Music aptitude				Experience factors					
	Pitch discrimination		Duration discrimination		Formant discrimination		Melody production		Rhythm production		Age of acquisition		Past experience		Current L2 use	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Segmentals	−.357†	.013	−.389*	.006	−.077	.602	.349†	.015	.108	.464	.084	.572	.020	.894	.012	.938
Stress	−.255	.080	−.340†	.018	.035	.811	.305†	.035	.315†	.029	−.079	.593	−.084	.569	−.069	.640
Intonation	−.419*	.003	−.099	.505	−.184	.211	.342†	.017	.159	.279	−.026	.862	−.110	.455	−.167	.257
Optimal speed	−.168	.254	−.004	.981	.006	.969	.162	.270	.303†	.036	−.212	.149	−.022	.884	−.126	.394

Notes. Current = current (in the UK) English learning hours. Past = past (in China) English learning hours.
† for marginally significant ($p < .05$). * for statistically significant ($p < .0125$) (Bonferroni corrected).

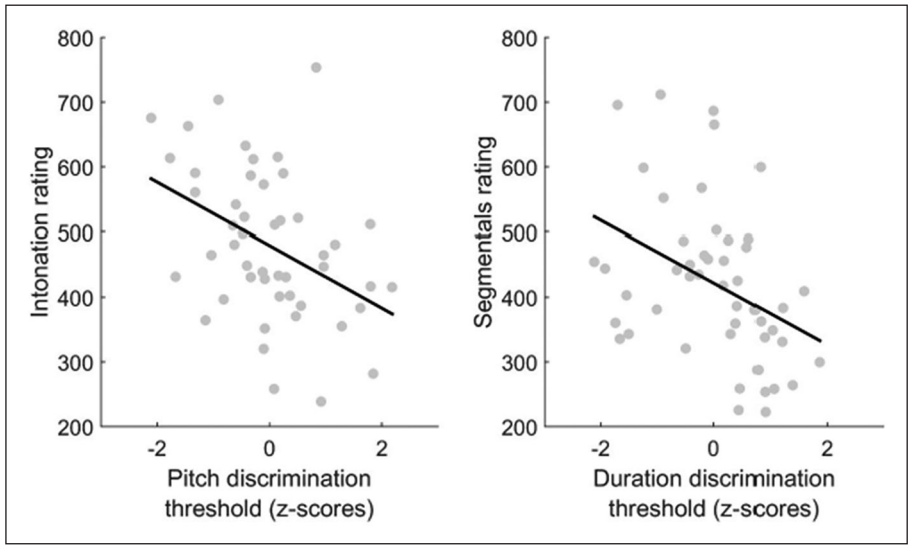

Figure 3. Scatterplots. (Left) Scatterplot displaying relationship between pitch discrimination (log-transformed and converted to z-scores) and intonation rating. (Right) Scatterplot displaying relationship between duration discrimination (log-transformed and converted to z-scores) and segmentals rating.

Table 6. Significant predictors of L2 pronunciation.

Predicted variables	Predictors	Adjusted R^2	Standardized coefficient beta	*t*	*p*
Segmentals	Duration discrimination	.133	–.362	–2.805	.007
	Pitch discrimination	.091	–.327	–2.534	.015
Word stress	Duration discrimination	.097	–.355	–2.701	.010
	Rhythm production	.094	.331	2.519	.015
Intonation	Pitch discrimination	.158	–.419	–3.129	.003
Optimal speed	Rhythm production	.072	.303	2.160	.036

and rhythmic production ($R^2 = .094$) emerged as significant predictors. Intonation could only be predicted by pitch discrimination ($R^2 = .158$), and fluency could be predicted by rhythm production ($R^2 = .072$).

V Discussion and future directions

As for the relationship between auditory discrimination, music aptitude and experience (RQ1), results of the Pearson correlation (see Table 4) revealed that melody production in music aptitude was significantly connected to pitch discrimination and marginally linked to formant discrimination, while rhythm production was not related to any dimensions of auditory processing abilities. These connections suggest that auditory processing abilities

and music aptitude may be partially overlapping especially on spectral levels (pitch discrimination and melody production), but not on temporal levels (duration discrimination and rhythm production) (for similar findings, see Kempe et al., 2015).

In terms of the role of auditory processing, music aptitude, and experience in L2 pronunciation proficiency (RQ2), a set of Pearson correlation and multiple regression analyses revealed that auditory processing was the significant predictor for various dimensions of participants' L2 pronunciation proficiency. Comparatively, music aptitude was only secondarily related to prosodic aspects of L2 pronunciation proficiency, especially when the relative weights of auditory processing and music aptitude were considered in stepwise multiple regression analyses (see Table 6). Finally, none of the experience variables demonstrated any significant associations with L2 pronunciation proficiency within the current dataset. Specifically, participants' segmental articulations were primarily predicted by their auditory processing, while suprasegmentals relied on a combination of auditory processing and music aptitude. These findings successfully validated the framework that regarded auditory processing abilities as L2 phonetic aptitude (e.g. Kachlicka et al., 2019).

At the same time, the predictive function of music aptitude for pronunciation appears to be somewhat marginal, at least in this population. This could be arguably because such music aptitude taps into a range of perceptual and cognitive abilities, including short-term memory, and as such is not a very pure measure of auditory processing (for their critical review on the ambiguous relationship between music aptitude and L2 pronunciation learning, see Trofimovich et al., 2015). In addition, our music aptitude test used in the current study was productive rather than perceptual.

The results showed that it is our comparatively unique and perceptual measure of auditory processing (rather than the composite and productive construct of music aptitude) which showed a significant relationship with L2 pronunciation proficiency. Thus, we argue that the relationship between auditory discrimination and L2 proficiency reflects a specific role for auditory processing in language learning, rather than the influence of domain-general cognitive factors (Tierney and Kraus, 2014). Further, the stronger and clearer effects of discrimination (auditory processing) than production measures (music aptitude) suggest that the mechanisms underlying L2 speech learning (segmental and prosodic accuracy in particular) could be perceptual in nature (Flege and Bohn, in press).

Finally, we would like to point out that our findings are in line with prior studies demonstrating the significant effects of perceptual-cognitive aptitude in L2 speech learning (e.g. Hu et al., 2013; Saito et al., 2020c). These results together concur with Doughty's (2019) theoretical discussion that it is a combination of aptitude and experience that interact to affect the degree of success in various dimensions of L2 learning. That is, the rate and ultimate attainment of adult L2 speech learning could be intrinsically determined by the extent to which individuals can make the most of received input via their access to auditory precision, encoding and integration (Kachlicka et al., 2019).

Overall, our findings provide some empirical support to a theoretical view that a range of perceptual-cognitive mechanisms used for successful L1 acquisition remain intact throughout one's lifespan, and are germane to the context of adult L2 speech learning (Flege and Bohn, in press). More specifically, our study has revealed that domain-general

auditory processing, which prior work has linked to L1 acquisition and delay, may be a primary determinant of L2 speech acquisition as well. As shown in previous literature, many late L2 learners continue to enhance L2 speech proficiency, as long as they regularly use a target language over an extensive period of immersion (for vowels, see Munro et al., 2013; for prosody, see Trofimovich and Baker, 2006). However, we argue that experience may be a necessary but not a sufficient condition. According to our data, it is possible that individuals with more precise auditory processing likely make the most of every input and output opportunity, as they can better decode, memorize and integrate acoustic information for L2 phonological learning. In the long run, individual differences in auditory processing may predict the extent to which L2 learners can enhance their L2 pronunciation proficiency (for a comprehensive overview on the relationship auditory perception, experience, and acquisition, see Saito et al., 2020e).

To close, we would like to point out some limitations of the current study, and suggest directions for future studies. First of all, the dataset in the current study was cross-sectional in nature. In order to further scrutinize the role of perceptual-cognitive individual differences in L2 speech learning, future studies need to adopt a longitudinal design (see Saito et al., 2020a).

Another limitation of the current study is that although the participants demonstrated a great deal of variation in terms of the onset and length of past L2 English learning experience (shown in Table 3), all of them had very similar immersion experience: The length of their residence in the UK was 1 year. Prior studies have found that Chinese learners tend to have a stronger tie to their L1 community, and consequently a lower degree of self-confidence in their English abilities compared to speakers from other countries, which may influence their overall pronunciation performance (Derwing and Munro, 2013). Therefore, to further examine the correlations between aptitudes, experience and ultimate language proficiency, future studies should recruit a larger number of learners with diverse backgrounds, language proficiencies and experience profiles (see Saito et al., 2020c).

Third, the findings reported in the current study were based on Chinese learners of English. It would be interesting to examine how these results from Chinese individuals could differ from other groups of L2 speakers learning alphabetic languages without lexical tones. On the one hand, it has been shown that auditory processing profiles could differ between tonal vs. non-tonal language users (Giuliano et al., 2011). On the other hand, there is some emerging evidence that the predictive power of auditory processing for multiple dimensions of L2 English learning could be found across different groups of L1 speakers (i.e. Chinese, Spanish vs. Polish; Saito et al., 2020d). To further examine this topic, future studies should recruit more participants and more linguistic measures.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The project was funded by Arnold Bentley New Initiatives Fund,

Spencer Foundation Research Grant (202100074), and Leverhulme Trust Research Grant (RPG-2019-039).

ORCID iD

Kazuya Saito https://orcid.org/0000-0002-4718-2943

Notes

1. We surveyed the participants' music training (for details, see Appendix 1). The average length of training was 2.7 years (SD = 3.89 years; $Range$ = 0–18 years). We decided not to use the information in the subsequent statistical analyses for the following reasons. First, it was difficult to determine precisely how to define and quantify 'music training' as participants reported various types of training (e.g. instruments vs. singing) with varied degree of intensity and formality (e.g. with vs. without tutors). Second, the data was skewed and deviant from normal distribution with many participants reporting no experience (24 out of 48). Third, following the methodological norm in the field of music perception research (Zhang et al., 2020), we defined musicians as those who self-report 6+ years of music training. In total, 11 out of 48 were identified as musicians in our dataset. According to the analyses of independent sample t-tests, we failed to find significant group differences (Musicians vs. Non-Musicians) in any contexts of L2 pronunciation proficiency and auditory processing profiles (p = .128–.883).
2. A set of independent t-tests was performed to examine the effect of task version on four different dimensions of L2 pronunciation proficiency (segmentals, word stress, intonation, optimal speed). The results did not find any significant differences between task versions for any of the four L2 pronunciation proficiency measures ($p > .05$). This indicates that the effect of speaking task version was minimal in the current investigation.
3. We did not normalize the raters' pronunciation scores. It was important to maintain their original scores so that we could average not only different rating patterns, but also different levels of leniency.

References

Bosker HR, Pinget A-F, Quene H, Sanders T, and de Jong NH (2013) What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing* 30: 159–75.

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

Delogu F, Lampis G, and Belardinelli MO (2010) From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology* 22: 46–61.

Derwing TM and Munro MJ (2013) The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning* 63: 163–85.

Doughty CJ (2019) Cognitive language aptitude. *Language Learning* 69: 101–26.

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

Flege J and Bohn O-S (in press) The revised speech learning model. In: R Wayland (ed) *Second Language Speech Learning: Theoretical and Empirical Progress*. Cambridge: Cambridge University Press.

Flege JE and Bohn O-S (2020) *The revised Speech Learning Model (SLM-r)*.

Giuliano RJ, Pfordresher PQ, Stanley EM, Narayana S, and Wicha NY (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 HLS, Cruz A, et al. (2011) Language-universal sensory deficits in developmental dyslexia: English, Spanish, and Chinese. *Journal of Cognitive Neuroscience* 23: 325–37.

Hu X, Ackermann H, Martin JA, et al. (2013) Language aptitude for pronunciation in advanced second language (L2) learners: Behavioural predictors and neural substrates. *Brain and Language* 127: 366–76.

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

Kachlicka M, Saito K, and 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, Bublitz D, and Brooks PJ (2015) Musical ability and non-native speech–sound processing are linked through sensitivity to pitch and spectral information. *British Journal of Psychology* 106: 349–66.

Kraus N and Chandrasekaran B (2010) Music training for the development of auditory skills. *Nature Reviews Neuroscience* 11: 599–605.

Larson-Hall J (2010) *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

Lengeris A and 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: 3757–68.

Levitt HCCH (1971) Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America* 49: 467–77.

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

Li S (2016) The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition* 38: 801–42.

Milovanov R, Pietilä P, Tervaniemi M, and Esquef PA (2010) Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education. *Learning and Individual Differences* 20: 56–60.

Munro MJ and Derwing TM (2001) Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition* 23: 451–68.

Munro MJ, Derwing TM, and Saito K (2013) *English L2 vowel acquisition over seven years*. In: Levis J and LeVelle K (eds) *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching conference 2012*. Ames, IA: Iowa State University, pp. 112–19.

Patel A (2003) Language, music, syntax and the brain. *Nature Neuroscience* 6: 674–81.

Pei Z and Ting R (2013) An empirical study on the effects of musical aptitude on English phonological skills of Chinese university students. *Foreign Language World* 1: 36–44.

Pei Z, Wu Y, Xiang X, and Qian H (2016) The effects of musical aptitude and musical training on phonological production in foreign languages. *English Language Teaching* 9: 19–29.

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

Plonsky L and Ghanbar H (2018) Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal* 102: 713–31.

Reiterer SM, Hu X, Erb M, et al. (2011) Individual differences in audio-vocal speech imitation aptitude in late bilinguals: Functional neuro-imaging and brain morphology. *Frontiers in Psychology* 2: 271.

Saito K (2015) Experience effects on the development of late second language learners' oral proficiency. *Language Learning* 65: 563–95.

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

Saito K, Trofimovich P, and Isaacs T (2017) Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics* 38: 439–62.

Saito K, Sun H, and Tierney A (2019) Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental and suprasegmental sensitivity and performance via behavioural and neurophysiological measures. *Bilingualism: Language and Cognition* 22: 1123–40.

Saito K, Sun H, and Tierney A (2020a) Domain-general auditory processing determines success in second language pronunciation learning in adulthood: A longitudinal study. *Applied Psycholinguistics* 41: 1083–1112.

Saito K., Sun H., and Tierney A. (2020b). Brief report: Test–retest reliability of explicit auditory processing measures. *bioRxiv*. Epub ahead of peer review 17 June 2020. DOI: 10.1101/2020.06.12.149484.

Saito K, Kachlicka M, Sun H, and Tierney A (2020c) 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* 115: 104168.

Saito K, Sun H, Kachlicka M, and Carvajal Alayo JR (2020d). Domain-general auditory processing explains multiple dimensions of L2 acquisition in adulthood. *Studies in Second Language Acquisition*. Epub ahead of print 9 November 2020. DOI: 10.1017/S0272263120000467.

Saito K, Suzukida Y, Tran M, and Tierney A (2020e) Auditory processing partially explains L2 speech learning in classroom settings: A review and generalization study. *Language Learning*. Article accepted 7 October 2020.

Schellenberg EG (2015) Music training and speech perception: A gene-environment interaction. *Annals of the New York Academy of Sciences* 1337: 170–77.

Segalowitz N (2016) Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching* 54: 79–95.

Slevc LR and Miyake A (2006) Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science* 17: 675–81.

Strait DL, Hornickel J, and Kraus N (2011) Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behavioral and Brain Functions* 7: 44.

Surprenant AM and Watson CS (2001) Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *The Journal of the Acoustical Society of America* 110: 2085–95.

Tierney A and Kraus N (2014) Auditory-motor entrainment and phonological skills: Precise auditory timing hypothesis (PATH). *Frontiers in Human Neuroscience* 8: 949.

Tierney AT, Krizman J, and Kraus N (2015) Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences* 112: 10062–67.

Trofimovich P and 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.

Trofimovich P, Kennedy S, and Foote JA (2015) Variables affecting L2 pronunciation development. In: Reed M and Levis J (eds) *The handbook of English pronunciation*. Malden, MA: Wiley, pp. 353–73.

Wing H (1968) *Tests of musical ability and appreciation: An investigation into the measurement, distribution, and development of musical capacity*. Cambridge: Cambridge University Press.

Zhang JD, Susino M, McPherson GE, and Schubert E (2020) The definition of a musician in music psychology: A literature review and the six-year rule. *Psychology of Music* 48: 389–409.

Appendix 1. 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.

Source. Adapted from Saito et al. (2017).