To be published in *Bilingualism: Language and Cognition* (Cambridge University Press)

Explicit and Implicit Aptitude Effects on Second Language Speech Learning: Scrutinizing Segmental and Suprasegmental Sensitivity and Performance via Behavioural and Neurophysiological Measures[1]

Kazuya Saito[a], Hui Sun[a] & Adam Tierney[b]

Birkbeck, University of London
[a] Department of Applied Linguistics
[b] Department of Psychological Sciences

Abstract

The current study examines the role of cognitive and perceptual individual differences (i.e., aptitude) in second language (L2) pronunciation learning, when L2 learners' varied experience background is controlled for. A total of 48 Chinese learners of English in the UK were assessed for their sensitivity to segmental and suprasegmental aspects of speech on explicit and implicit modes via behavioural (language/music aptitude tests) and neurophysiological (electroencephalography) measures. Subsequently, the participants' aptitude profiles were compared to the segmental and suprasegmental dimensions of their L2 pronunciation proficiency analyzed through rater judgements and acoustic measurements. According to the results, the participants' segmental attainment was associated not only with explicit aptitude (phonemic coding), but also with implicit aptitude (enhanced neural encoding of spectral peaks). Whereas the participants' suprasegmental attainment was linked to explicit aptitude (rhythmic imagery) to some degree, it was primarily influenced by the quality and quantity of their most recent L2 learning experience.

Key words: Cognitive individual differences, second language speech, pronunciation, explicit aptitude, implicit aptitude

[1] Acknowledgement: This study was funded by the Birkbeck College Additional Research Fund.

Understanding the process and product of second language acquisition (SLA) is complex, as it can be explained not only by factors related to experience (i.e., the extent to which second language [L2] learners practice the target language), but also by those related to aptitude (i.e., the cognitive and perceptual factors which determine the extent to which L2 learners can make the most of relevant L2 experience). Whereas the previous literature has examined aptitude in reference to L2 lexicogrammar development (for reviews, see Li, 2016; Skehan, 2016), surprisingly little is known about the role of aptitude in L2 pronunciation learning. The present study aims to fill this gap by proposing a new framework of cognitive abilities relevant to the degree of success after years of explicit and implicit pronunciation learning under various L2 learning conditions. To achieve this main objective, we assessed the segmental and suprasegmental sensitivity of 48 Chinese learners of English in the UK by using a range of behavioural (language and music aptitude tests) and neurophysiological (electroencephalography) measures. Subsequently, we explored which pronunciation learning aptitude variables were linked to the segmental and suprasegmental aspects of the learners' L2 pronunciation performance, controlling for their L2 learning backgrounds (i.e., their past and recent L2 use).

Background

Second Language Pronunciation Development

Second language pronunciation proficiency is a composite skill which comprises the capacity to (a) pronounce new consonantal and vocalic sounds in a L2 without deleting or substituting them for L1 counterparts (segmental accuracy); (b) use adequate prosody at the word (correct assignment of word stress) and sentence (appropriate use of intonation for declarative and interrogative intensions) levels; and (c) deliver speech at an optimal tempo (speed fluency) without making too many pauses (breakdown fluency) nor self-repetitions or corrections (repair fluency). According to general L2 speech theories (e.g., Kormos, 2014), comprehension processes primarily draw on the decoding of phonological information. When speech includes mispronunciations or unclear pronunciation, listeners may activate inappropriate lexical items, which in turn may hinder their prompt, timely and successful understanding of speakers (Broesma, 2012). Relative to other domains of language (vocabulary, grammar), therefore, the accurate and fluent use of pronunciation is considered to be a particularly fundamental component of L2 oral proficiency (Derwing & Munro, 2009).

A common feature of theoretical models of SLA is that L2 learners continue to improve their pronunciation proficiency with increased input and output of the target language (e.g., Flege, 2016 for Speech Learning Model). More specifically, usage-based accounts of language development explain in depth how experience uniquely facilitates SLA according to how often (frequency), where (contexts) and when (recency) L2 learners practice the target language (e.g., Ellis, 2006). Similar to first language acquisition, early L2 learners (e.g., age of acquisition < 6 years) are likely to achieve high-level pronunciation proficiency, given ample opportunities for language exposure (Abrahamsson & Hyltenstam, 2009). When it comes to adult L2

pronunciation learning, strong experience effects (i.e., more practice is better) are observed at the initial stage of L2 pronunciation learning (e.g., Munro & Derwing, 2008 for first three months of immersion). Yet, a great deal of individual variability is present in the final outcome of late L2 pronunciation learning. Even if any two given L2 learners have similar kinds of L2 experience, the extent to which they notice, understand and learn to produce L2 features can greatly vary. One possible source of these individual differences in L2 learning outcome is proficiency in a variety of cognitive and perceptual skills, which together make up *second language learning aptitude*.

Second Language Learning Aptitude

One of the most extensively-researched topics in the field of SLA has been the explanatory power of individuals' aptitude for the rate and ultimate attainment of L2 learning (for reviews, see Li, 2016; Skehan, 2016). As originally conceived, aptitude referred to the explicit and intentional learning abilities necessary for successful foreign language learning through formal instruction. In Carroll and Sapon's (1959) influential aptitude model, such abilities constitute phonemic coding, grammatical sensitivity, inductive learning ability and associative memory. According to previous validation studies (e.g., Carroll, 1962), L2 learners' different levels of aptitude, measured by the Modern Language Aptitude Test battery, demonstrated significant associations with their achievements in various classroom settings, such as course grades and SAT scores.

More recently, a growing number of scholars (e.g., Linck et al., 2013; Skehan, 2016) have proposed new theoretical frameworks for conceiving aptitude in terms of implicit and incidental learning (i.e., learning without awareness)—a type of learning which may be crucial for high-level L2 acquisition in naturalistic settings. Different from explicit learning aptitude, which is measured through tasks comprising both practice and testing phases, implicit and incidental learning aptitude is measured while participants complete tasks without any practice nor awareness of what is being learned. Developing a composite test battery of 11 domain-general cognitive measures (Hi-LAB), for example, Linck et al. (2013) examined the aptitude profiles of advanced L2 learners who obtained high reading and listening scores on Defense Language Proficiency Tests. These learners demonstrated not only greater associative (paired associations) and phonological short-term memory (letter span), but also higher implicit language aptitude (serial reaction time).

In order to analyze the influence of aptitude in various L2 learning contexts, the LLAMA aptitude test battery has been widely adopted in the field of SLA. Building on Carroll's aptitude model, LLAMA features not only explicit learning aptitude—associative memory, phonemic coding and grammatical inferencing, but also incidental learning aptitude—sound sequence recognition. According to previous investigations, explicit LLAMA test scores appeared to predict the extent to which L2 learners can benefit from explicit (rather than implicit) instruction within a short amount of time under laboratory (e.g., Yilmaz & Granena, 2016) and classroom (e.g., Yalçın & Spada, 2016) conditions. In contrast, L2 learners with high-level incidental

aptitude (sound sequence recognition) tend to attain advanced proficiency in L2 morphosyntax, especially when they have had regular access to naturalistic language input since an early age (e.g., Granena, 2013).

Whereas an extensive body of literature has scrutinized the complex relationship between various kinds of aptitude (explicit and implicit), L2 proficiency (beginner, intermediate, advanced) and context (naturalistic vs. classroom settings), it is noteworthy that most of the relevant research evidence has nearly exclusively considered the effects of aptitude on the learning of the acquisition of listening/reading skills, measured via general proficiency tests (e.g., Linck et al., 2013 for the Defense Language Proficiency Tests), and the learning of L2 morphosyntax (Granena, 2013; Yalçın & Spada, 2016). Very few studies have examined the impact of these factors on the acquisition of L2 adult learners' phonological skills while speaking spontaneously via a comprehensive set of aptitude and speech measures. (cf. Saito, Suzukida, & Sun, 2018) Furthermore, very few studies have used a combination of both behavioural and neurophysiological metrics.

Developing a New Aptitude Framework for L2 Pronunciation Learning

In this study, L2 pronunciation learning aptitude is defined as comprising the cognitive abilities related to the explicit and implicit processing of acoustic information, which is crucial for perceiving various phonetic dimensions of L2 speech. We propose that learners who have greater aptitude in tracking and retaining acoustic information are better able to attend to the primary acoustic correlates of segmentals (high-frequency spectral information), prosody (fundamental frequency height and contour), and fluency (relative ratio of speaking/silent time). We consider this kind of aptitude to be receptive rather than productive in nature, as we follow the predominant theoretical assumption that L2 speech learning is perception-driven (i.e., changes in perception lead to production development) (Flege, 2016). To develop the aptitude framework, we excluded cognitive tasks using non-speech materials, such as letter- and non-word span for phonological short-term memory (Linck et al., 2013), retrieved-induced inhibition for inhibition control (Darcy, Mora & Daidone, 2016), and speeded naming for processing speed (Darcy, Park, & Yang, 2015). This decision is motivated by recent research evidence that both child and adult L2 speech learning is closely tied to human sensitivity to complex speech signals such as language (Diaz, Mitterer, Broersma, Escera, & Sebastian-Galles, 2016) and music (Milovanov, Pietilä, Tervaniemi, & Esquef, 2010).

Based on a synthesis of extant studies on the cognitive predictors of L1 and L2 speech learning, we identified a total of four measures that differentially reflect aptitude in the explicit and implicit processing of L2 phonological information at the segmental and suprasegmental levels (as summarized in Table 1). Our framework is novel as all the tasks correspond to cognitive/perceptual abilities which are thought to be directly relevant to L2 learners' potentially different processing of segmental, prosodic and temporal information in both explicit (phonemic coding, tonal/rhythmic imagery) and implicit (auditory encoding precision) modes.

Table 1. *Constructs and Measures of Pronunciation-Specific Aptitude*

Aspects of L2 pronunciation proficiency	Construct of aptitude	Measures
Segmentals (consonants, vowels)	Explicit	Phonemic coding
	Implicit	FFR at F1 and F2
Suprasegmentals (prosody, fluency)	Explicit	Tonal and rhythmic imagery
	Implicit	FFR at F0

Note. FFR for frequency following response

One explicit speech-specific component of aptitude is phonemic coding, defined as L2 learners' ability to analyze, categorize and remember new segmental sounds in relation to corresponding symbols. Individual differences in children's explicit knowledge of the phonology of their L1 have been linked to differences in speech perception (Rvachew & Grawburg, 2006) and auditory processing (Anvari, Trainor, Woodside, & Levy, 2002). In Li's (2016) meta-analysis, this form of aptitude seems to be only weakly associated with the development of global listening and speaking skills. More recently, however, four studies have explored and confirmed the moderate-to-strong predictive power of phonemic coding for adult L2 learners' pronunciation attainment, especially at a segmental level after years of classroom (Saito, 2017, in press; Saito et al., 2018) and naturalistic (Granena & Long, 2013) L2 learning. Given that the cognitive underpinnings of experienced L2 learners' segmental learning and attainment remains open to debate (e.g., Darcy et al., 2015, 2016), the acquisition-aptitude link needs to be further examined.

Recent work has shown connections between shared processes in language and music comprehension on several levels, including between harmony and syntax (Patel, Gibson, Ratner, Besson, & Holcomb, 1998), rhythm and stress (Cason, Astésano, & Schön, 2015), melody and intonation (Liu, Patel, Fourcin, & Stewart, 2010) and semantics (Daltrozzo & Schön 2009). These connections between music and language suggest that aptitude for acquiring *suprasegmental* aspects of new languages (prosody, fluency) and aptitude for learning to perceive and produce music may be overlapping constructs as well. In music aptitude tests (e.g., Gordon, 1995), learners are tested for their abilities to hear differences in music in pitch/intensity (i.e., tonal imagery) and speed/timing (i.e., rhythmic imagery), when listening to two musical notes. This test is considered as one form of explicit aptitude test, since participants are explicitly guided to pay conscious attention towards analyzing the tone/rhythm of the notes during the test taking session.

Several empirical studies have pointed out that those with higher music aptitude (e.g., musicians) can better recognize and produce sounds not only in a familiar L2 (English) (e.g., Milovanov et al., 2010; Slevc & Miyake, 2006), but also in an unfamiliar tonal language that they have never learned (Mandarin) (e.g., Gottfried, 2007; Wong, Skoe, Russo, Dees, & Kraus, 2007). In an intervention study with a pre- and post-test design, Li and DeKeyser (2017) recently provided longitudinal evidence that music aptitude could mediate the effects of explicit

instruction on American learners' acquisition of L2 Mandarin lexical tones. Specifically, the authors hypothesized that more musically endowed learners may be more sensitive to and capable of capturing acoustic information in speech related to F0 height and contour. Therefore, it is possible that L2 learners' music aptitude for perceiving tonal and rhythmic imagery could contribute L2 pronunciation proficiency especially at prosodic and temporal levels—an assumption that the current study was designed to test.

Departing from previous aptitude studies predominantly concerned with explicit aptitude, the current study measures implicit pronunciation-specific aptitude in terms of L2 learners' neural encoding of speech, which we measure using an electrophysiological response known as the frequency following response (FFR), a response with origins within the cortical and subcortical auditory system (Coffey, Herholz, Chepesiuk, Baillet, & Zatorre, 2016). The FFR reproduces the temporal and spectral content of the evoking stimulus, and so can be used to assess the stability and precision of the auditory system's encoding of spectral, pitch, and durational information, acoustic features which convey segmental and prosodic information in speech. Attention is not necessary for the elicitation of the FFR; during recording, therefore, participants can engage in absorbing tasks which draw attention away from the sounds (e.g., reading books, watching silent movies). As such, the method is an ideal way to assess auditory processing without the contaminating influence of cognitive and affective state.

In the neurophysiology literature, the degree of auditory precision, estimated through FFR, continues to develop up until around 7-10 years of age (Skoe, Krizman, Anderson, & Kraus, 2013) before reaching a relatively stable state (Hornickel, Knowles, & Kraus, 2012). Individual differences in the FFR have been found to exhibit strong correlations with language skills such as reading (White-Schwoch et al., 2015) and speech in noise perception (Anderson, Skoe, Chandrasekaran, & Kraus, 2010), suggesting that the auditory skills indexed by the FFR are vital for language processing and acquisition. Nevertheless, there has been only a single previous study of the relationship between neural encoding of speech as measured by the FFR and success in learning a second language in adulthood. Omote, Jasmin, & Tierney (2017) examined perception of English phonology and FFR phase-locking in native Japanese adults who moved to the United Kingdom in adulthood. Robust neural encoding of the F0 of speech was linked to successful English speech perception. Interestingly, it was also shown that the participants' FFR predicted their performance even more strongly than their experience backgrounds did (their length of residence in the UK). These results are in line with previous findings that bilingual experience enhances the neural representation of speech F0 (Krizman, Slater, Skoe, Marian, & Kraus, 2015). Here we built upon these previous findings by asking for the first time whether neural encoding of speech in the frequency-following response is linked to proficiency in second-language production.

Neural encoding of pitch was assessed by measuring the robustness of neural phase-locking to the fundamental frequency (100 Hz) of a synthesized speech syllable (/da/). We hypothesized that neural encoding of pitch would be linked to participants' ability to produce suprasegmental features of second language speech. Neural encoding of higher-frequency

spectral information was assessed by measuring neural phase-locking at the frequencies of the first (720 Hz, measured at 700 Hz) and second (1240 Hz, measured by averaging responses at 1200 and 1300 Hz) formants. We hypothesized that neural encoding of speech formants would be linked to participants' ability to produce segmental features of second language speech.

Current Study

Adopting the proposed L2 pronunciation learning aptitude framework (phonemic coding, music aptitude, auditory encoding precision), the main objective of the current study was to scrutinize the cognitive correlates of successful L2 pronunciation proficiency attainment among 48 Chinese learners of English in the UK. First, we carefully checked how the participants differed in terms of their sensitivities to segmental and suprasegmental pronunciation learning—i.e., aptitude factors—and their past and recent use of the L2 in classroom and naturalistic settings—i.e., experience factors. To examine the relative effects of experience and aptitude factors on L2 pronunciation attainment, we examined how the aptitude and experience factors differentially contributed to their segmental and suprasegmental aspects of L2 pronunciation proficiency at the time of the project (after 11-17 years of L2 learning in both classroom and naturalistic settings). The hypothesized relationships between independent variables (Aptitude, Experience) and dependent variables (Pronunciation) was summarized in Figure 1.

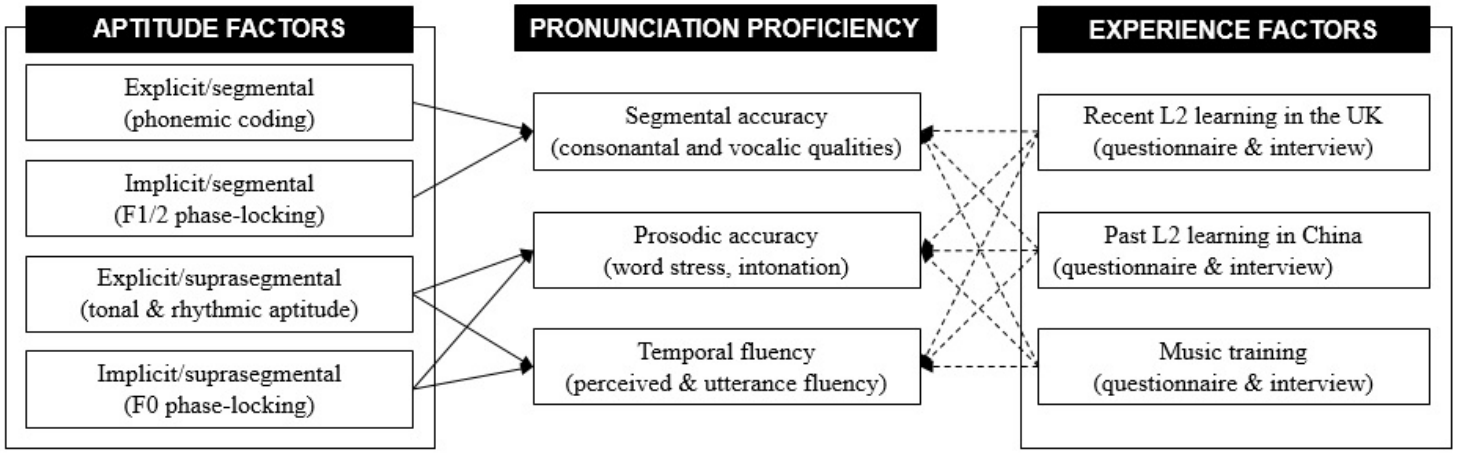

Figure 1

Summary of the Hypothesized Relationships between Independent Variables (Aptitude, Experience) and Dependent Variables (Pronunciation)

Participants

As a part of a larger project designed to survey English oral proficiency among international students in the UK, a total of 48 native speakers of Mandarin Chinese were recruited (6 males, 42 females) for the study. All the participants were students enrolled in various postgraduate programmes (but one who was at undergraduate level) in London (M_{age} = 23.8 years, *Range* = 21-27) with similar length of residence in the UK (i.e., eight to nine months). During the academic programme, they took a different number of content-based courses in various subjects in sciences (e.g., engineering, mathematics, chemistry) and social sciences (e.g., economics, linguistics, law), while none of them attended any English-as-a-second-language classes. Some participants had many opportunities to speak L2 in the class through group discussions and presentations, whereas others had less chances due to a different course focus. On the other hand, their L2 use (in terms of speaking, listening, reading and writing) outside the class also varied to a great degree (as reported in the Result section). Prior to coming to the UK, they had studied L2 English only in China for 10-16 years without any study-abroad experience in an English-speaking environment, albeit with different ages of learning onset ($M_{age\ of\ learning}$ = 8.4 years: *Range* = 6-13 years). Their self-reported IETLS scores widely varied from 6 to 8 out of 9 (M = 7.1, SD = 0.4). According to CEFR bands, this signals that their general proficiency could be considered from B2 (Independent users) to C1 (Proficient users).

All participants had audiometric thresholds ≤ 25 dB HL for octaves from 500 Hz to 4000 Hz, confirming their normal hearing. The data collection was conducted in a soundproof booth. Each session lasted for approximately 90 minutes per participant with the three main tasks being administered in the following order: aptitude test, pronunciation test and experience interview. To avoid any misunderstandings of the procedure, all instruction was delivered in Chinese by an L1 Mandarin speaking researcher.

Measures of Pronunciation Proficiency

Speaking Task. In the field of L2 speech research, controlled speech tasks (e.g., delayed sentence repetition) have been typically used to elicit participants' production of certain segmental and suprasegmental features. Yet, some scholars have continuously emphasized the importance of adopting more spontaneous speech tasks, especially for adult L2 learners, who can carefully monitor their correct pronunciation forms when they are allowed to draw on their explicit phonetic knowledge without any attention to the meaningful use of language (Piske, Flege, MacKay, & Meador, 2011). Indeed, it has been shown that adult L2 learners' speech behaviours are different when elicited via controlled and free speech tasks with their former performance being more targetlike and accurate than the latter performance (Major, 2008). In extemporaneous speech tasks, L2 learners are guided to produce language with a primary focus on conveying their intended message under time pressure (for a review, see Skehan, 2016). Similar to previous L2 speech studies (e.g., Lambert, Kormos, & Minn, 2017), a timed picture narration task was adapted from the Pre-Grade 1 Level of the EIKEN English Test (EIKEN, 2016).

Procedure. Since types of topics could affect the participants' L2 performance (Gass & Varonis, 1984), two different versions of the narration task were prepared (Versions A and B). A total of 25 participants were randomly assigned to Version A, and the remaining 23 participants to Version B. For each version (A, B), the participants had one minute to prepare how to describe a four-frame cartoon, and two minutes to narrate the story. To avoid false starts, the participants were given the first sentence that they had to use (for materials, see Appendix A). All the speech samples were recorded with a Roland-05 audio recorder, set at 44.1 kHz sampling rate and 16-bit quantization, and a unidirectional condenser microphone. In line with L2 speech research standards (e.g., Derwing & Munro, 1997), and to reduce any fatigue effects on listeners in the subsequent rating sessions (see below), the first 30 sec of the speech samples was excised and normalized for peak amplitude for subsequent L2 pronunciation analyses.

Data Analyses: Subjective Judgements. In the analysis of segmental and prosodic qualities of spontaneous speech, objective measures (e.g., acoustic analyses) are not commonly used in the L2 speech literature, due to variability in phonetic context (e.g., following and preceding vowels) and talker characteristics (e.g., anatomical difference in vocal tract). Rather, many scholars have relied on linguistically trained raters' subjective scalar judgements (e.g., Piske et al., 2011 for segmentals; Derwing & Munro, 1997 for prosody). In our precursor research (Saito, Trofimovich, & Isaacs, 2017), a training procedure was elaborated for experienced native-speaking raters to assess four different categories of L2 pronunciation proficiency—segmentals, word stress, intonation and speech rate.[1]

To this end, five expert raters with ample linguistic and pedagogical backgrounds (3 females, 2 male) were recruited in London (M_{age} = 35.4 years). Whereas three out of five raters were originally from North America, they had resided in the UK more than 10 years, reporting high-level familiarity with Received Pronunciation. All of them held MA degrees in applied linguistics and reported extensive experience in teaching ($M_{\text{years of teaching}}$ = 7.8 years) and speech analyses of this kind through participating in rating sessions as research assistants and/or enrolling in rater training for high-stakes L2 speaking tests. None of them reported any hearing problems. Their familiarity with Chinese-accented speech was relatively high ($M_{familiarity}$ = 5.3, range = 5-6) on a 6-point scale (*1 = not at all, 6 = very much*).

Each rating session took place individually in a quiet room at a university in London with a researcher who had provided similar training in our previous studies. The raters listened to speech samples played in a randomized order via custom software (which was developed via MATLAB), and then used a moving slider to rate them on a 1000-point scale for segmental errors (*0 = frequent, 1000 = infrequent or absent*); word stress errors (*0 = frequent, 1000 = infrequent or absent*), intonation accuracy (*0 = unnatural, 1000 = natural*), and perceived tempo (*0 = too slow or too fast, 1000 = optimal speed*). Each end of the continuum was signalled with a frowny (for "0") face and a smiley (for "1000") face (for onscreen labels, see Appendix C). To

ensure the precision of ratings, the raters were allowed to listen to each speech sample as many times as they wanted to.

To familiarize the raters with the procedure, the researcher first gave brief instructions on the definition of each pronunciation category (for training materials, Appendix B). Second, they practiced the procedure by rating three speech samples which were not included in the main dataset. For each sample, they explained their decisions and received feedback from the researcher to check their understanding of the constructs. Finally, they moved onto analyzing a total of 48 speech samples with a 5-minute intermission halfway through. The entire session took approximately 90 minutes.

In terms of the inter-rater reliability, the results of the Cronbach alpha analyses identified medium agreement for the five raters' judgements of segmentals ($\alpha = .72$) and perceived tempo ($\alpha = .75$), both of which are in line with the standard in L2 research (i.e., $\alpha > .70$) (Larson-Hall, 2010). Pronunciation scores were averaged across all raters to generate a single score per participant according to segmentals and speech rate. Their word stress and intonation ratings yielded relatively low Cronbach alpha values ($\alpha = .56, .67$). As a remedy, two raters who demonstrated the strongest agreement ($\alpha = .85, 87$) were identified. Their averaged scores were used for the following word stress and intonation analyses.

Data Analyses: Acoustic Judgements. As operationalized in previous L2 suprasegmental studies (for a review, Lambert et al., 2017), the temporal aspects of the participants' spontaneous speech were acoustically examined according to three key constructs of fluency—breakdown, repair and speed. From a theoretical perspective (e.g., Kormos, 2014), these constructs are believed to correspond to L2 learners' cognitive operations at three different stages of L2 speech production—breakdown for conceptualization and linguistic formulation (searching what and how to say), repair for monitoring (correcting already-produced utterances) and speed for automatization (optimizing the entire production processes).

Breakdown fluency was calculated by dividing the number of filled (lexical fillers such as eh, um) and unfilled (silence) pauses by the total number of words. Whereas filled pauses were counted based on raw transcripts, unfilled silent pauses were automatically identified via a script programmed in *Praat* (Boersma & Weenink, 2017) with minimum silence duration set to 250 milliseconds. Repair fluency was calculated by dividing the total number of self corrections and repetitions (based on raw transcripts) by the total number of words. Speed fluency was measured via the articulation rate, which was calculated by dividing the total number of syllables by phonation time (i.e., total length of each audio file minus all silent, unfilled pauses). For similar fluency analysis methodology, see Bosker, Pinget, Quené, Sanders, & De Jong (2013).

To investigate inter-coder reliability, two researchers (both of whom had extensive experience on L2 fluency analyses) separately analyzed the breakdown, repair and speed fluency of 10 samples from the entire dataset. The results of Cronbach alpha analyses found relatively high agreement between the coders for breakdown ($\alpha = .92$), repair ($\alpha = .93$) and speed ($\alpha = .98$).

Where disagreement was found, they discussed to find a consensus. One of the coders thus proceeded to analyze the rest of the data (n = 38).

Behavioural Measures of Explicit Aptitude

Phonemic Coding. The participants' phonemic coding ability—the ability to associate unfamiliar sounds to symbols—was assessed via one component of the LLAMA test (Meara, 2005). In this subtest (LLAMA-E), the participants were first asked to remember the relationship between 24 recorded syllables (consonant-vowel) and their corresponding phonetic symbols within two minutes. The sound stimuli were created based on an indigenous language in Canada. After the practice session, their recollection was tested, specifically whether they could correctly identify symbols corresponding to two syllable words (a total of 20 items). The participants' phonemic coding aptitude scores were calculated out of 100 based on the tailored scoring rubrics in LLAMA.

Music Aptitude (Melody, Rhythm). Two subsections of the Musical Aptitude Profiles for Japanese (MAP-J) (Ogawa, 2009) were used to assess the participants' abilities to perceive tonal and temporal aspects of musical phrases. Building on Gordon's (1995) oft-used, validated music aptitude test (Music Aptitude Profile), the MAP-J was developed to evaluate, in particular, the aptitude of young and adolescent students in Japan and other east-Asian countries who are exposed to both western (violin, piano) and oriental (Japanese/Chinese drums, harp) musical instruments. Both melody and rhythm subtests required participants to make same/different judgments of pairs of short musical phrases. Participants assessed whether they were identical or different in pitch contour (for the melody subtest) and in the number/patterns of beats (for the rhythm subtest).

At the beginning of each subtest, the participants completed a practice session (listening to an "identical" and a "different" pair), followed by a main session which comprised 20 sets of musical phrases. The melody and rhythm scores were calculated out of 20.

Neurophysiological Measures of Implicit Aptitude

Stimulus. The speech token /da/ (170ms) was synthesized via a Klatt-based synthesizer. The first five ms of the sound was the onset burst, and the rest of the sound was voiced with a steady 100 Hz fundamental frequency throughout. While the first, second and third formants shift during the transitional period between 5 to 50ms (400 to 720Hz, 1700 to 1240Hz, 2580 to 2500Hz), all formants stayed constant during the steady state between 50 and 170ms (720Hz, 1240Hz, 2500Hz).

Procedure. The /da/ sound was presented repeatedly (6000 times over the course of 20 minutes) in alternating polarities through insert earphones (ER-3; Etymotic Research) at 80dB with 81ms interstimulus intervals. Presenting stimuli in alternating polarities (i.e. with half of the stimuli inverted) affords the opportunity to separately examine the envelope and the temporal

fine structure of speech by adding and subtracting responses of opposite polarities, respectively (Aiken & Picton, 2008). During the task, the participants were encouraged to focus on reading their favorite books in a relaxed environment, instead of paying special attention to sound properties. The electrophysiological responses to sound stimuli (/da/) were collected from the participants using a BioSemi EEG system with open filters and a sample rate of 16384 Hz. A single active electrode was located at the centre of the top of the head (i.e. at Cz), the reference electrodes were located at the earlobes, and ground electrodes were placed on the forehead.

Data Analyses. All neurophysiological analysis was conducted using custom-written software in MATLAB. First, the recording was epoched between -40 ms and 210 ms, relative to stimulus presentation. Trials containing amplitude spikes of >100 micro-volts were rejected as artifacts, and the first 5000 artifact-free responses to each stimulus polarity were selected for the main analysis.

Precision of neural sound encoding was measured using inter-trial phase-locking. Inter-trial phase-locking measures the degree of jitter at a particular frequency within a particular time window. This provides a frequency-specific metric of neural sound encoding that benefits from a relatively robust signal-to-noise ratio compared to analyses of the spectrum of cross-trial average waveforms (Zhu, Bharadwaj, & Shinn-Cunningham, 2013). A sliding-window technique was used to assess phase-locking at each time-frequency point. For each trial, a Hanning windowed fast Fourier transform was calculated on 40-ms segments centered at time points between 0 and 170 ms, with 1 ms intervals between time points. The resulting complex vector was then normalized to have a magnitude of 1. For calculation of the phase-locking of the envelope response, vectors were averaged across trials, while for calculation of the phase-locking of the temporal fine structure response, vectors for one of the two stimulus polarities were shifted 180 degrees before averaging. The length of the resulting average vector formed the inter-trial phase-locking value for that time-frequency point. Inter-trial phase locking can vary from zero (no phase consistency whatsoever) to one (perfect phase consistency across trials).

The envelope response contains a robust representation of the F0, and so was used for measurement of fundamental frequency encoding, while the temporal fine structure response contains a robust representation of the higher harmonics, and so was used for measurement of the speech formants. See Figure 2 for an illustration of the relationship between phase-locking in envelope and temporal fine structure responses and the spectro-temporal characteristics of the stimulus.

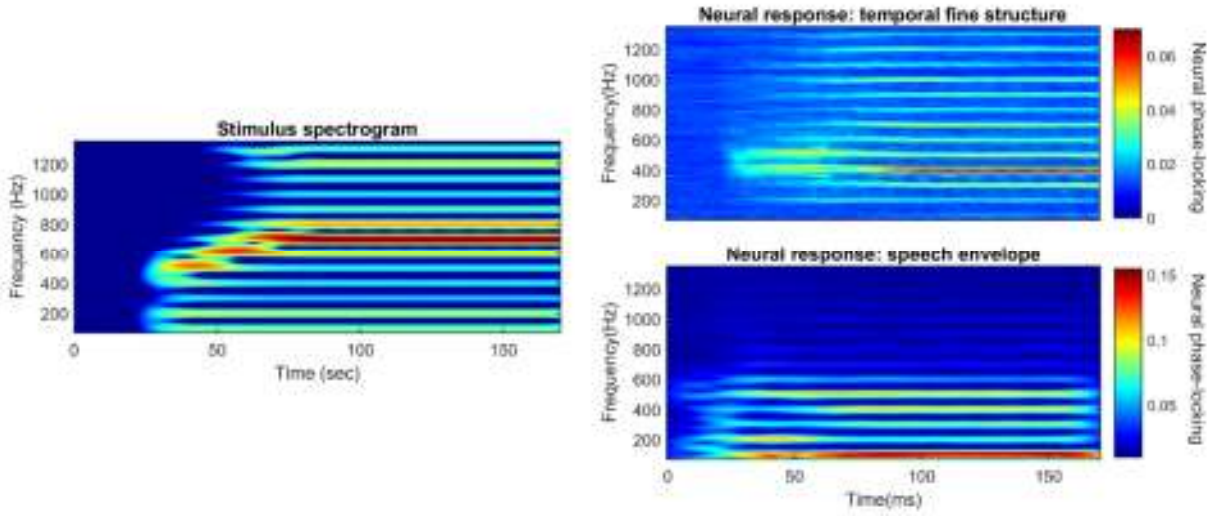

Figure 2

(Left) Spectrogram of stimulus used to evoke frequency-following responses. (Right) Inter-trial phase locking across time and frequency for the temporal fine structure response (top) and the envelope response (bottom).

F0 and formant encoding were quantified in the following manner. Since the fundamental frequency was steady at 100 Hz across the response, F0 phase-locking was quantified as the average phase-locking between 81 and 120 Hz from 10 to 170 ms in the envelope response. However, encoding of the first and second formants was assessed only during the portion of the response in which they were unchanging, i.e. from 60 to 170 ms. F1 was calculated as average phase-locking between 680 and 720 Hz (i.e. as the amplitude of the 7th harmonic), while F2 was calculated as average phase-locking between 1180 and 1220 Hz and between 1280 and 1320 Hz (i.e. as the mean amplitude of the 12th and 13th harmonics) in the temporal fine structure response.

Measures of Experience

Although the participants' length of residence in the UK was identical (i.e., eight-to-nine months), the quantity and quality of their L2 learning experience prior to and during their study-abroad differed to a great degree. The participants were individually interviewed to uncover their past and recent L2 learning backgrounds in a retrospective manner, using a similar interview scheme as that used in Muñoz (2014). As such, the participants self-reported the extent to which they had practiced L2 English inside and outside classrooms according to elementary, secondary and university-level schools in China as well as university-level schools in the UK. Finally, they also reported whether and for how long they had engaged in music training (e.g., experience of playing instruments), which has been linked to various aspects of L1 and L2 development (Slevc & Miyake, 2006; Tierney, Krizman, Kraus, & Tallal, 2015).

Results

Constructs of L2 Pronunciation Proficiency

Table 2 summarizes the results of the participants' segmental and suprasegmental dimensions of L2 pronunciation proficiency, measured by both the expert raters' judgements and acoustic analyses. As summarized in Table 3, the inter-relationships between seven pronunciation measures were assessed via a set of Pearson correlation analyses. An alpha value was corrected via Bonferroni corrections. Strong associations were observed particularly among the expert raters' segmental and prosodic (word stress, intonation) scores; the intonation and perceived tempo scores; and perceived and objective fluency scores (perceived tempo, articulation rate, pause ratio). Similar to the author's previous research (e.g., Saito et al., 2017), the pronunciation measures adopted in the study appeared to reveal the participants' four different pronunciation abilities to (a) pronounce individual sounds/words accurately (segmentals, word stress); (b) access adequate prosody (intonation, perceived tempo); (c) produce optimal fluency (perceived tempo, articulation rate, pause ratio); and (d) avoid too much self-monitoring (repair ratio).

Table 2. *Descriptive Statistics of Participants' Segmental and Suprasegmental Proficiency*

	M	*SD*	95% CI	
			Lower	Upper
A. Expert rater judgements				
Segmentals (0-1000 points)	490	123	454	525
Word stress (0-1000 points)	484	78	461	506
Intonation (0-1000 points)	488	94	461	516
Perceived tempo (0-1000 points)	625	103	593	653
B. Acoustic analyses				
Articulation rate (no. of syllables per second)	3.48	0.42	3.35	3.60
Pause ratio (%)	25.4	8.9	22.7	28.0
Repair ratio (%)	6.6	5.0	5.2	8.1

Given that L2 speech performance is likely influenced by task type, we further probed whether participants' pronunciation proficiency differed in two different task prompts. A set of paired samples t-tests were performed to compare the segmental and suprasegmental scores of the participants who used Versions A ($n = 25$) and B ($n = 23$). The results did not find any significant difference in any pronunciation measures ($p > .05$), suggesting that task effects could be considered minimal in this study.

Table 3. *Interrelationships between Segmental and Suprasegmental Proficiency Scores*

	Word stress	Intonation	Perceived tempo	Articulation rate	Pause ratio	Repair ratio
A. Expert rater judgements						
Segmentals (0-1000 points)	.47*	.39*	.27	.16	-.17	-.28
Word stress (0-1000 points)		.70*	.29	.21	-.01	-.01
Intonation (0-1000 points)			.46*	.29	-.07	-.09
Perceived tempo (0-1000 points)				.44*	-.32†	.01
B. Acoustic analyses						
Articulation rate (no. of syllables per second)					-.38*	.11
Total pause ratio (%)						-.07

Note. *indicates statistical significance at $p < .008$; † indicates marginal significance at $p < .01$ (Bonferroni corrected).

Constructs of Explicit and Implicit Aptitude Measures

As summarized in Table 4, descriptive statistics demonstrated a great deal of variation in participants' explicit and implicit aptitude scores. According to normality analyses (the Kolmogorov-Smirnov goodness-of-fit test), positive and negative skewness was observed for phonemic coding and FFR phase-locking at F1 ($p < .05$); therefore, these aptitude scores were transformed using the Log10 function for subsequent analyses.

Table 4. *Descriptive Statistics of Learner Aptitude Profiles*

	M	*SD*	95% CI	
			Lower	Upper
A. Explicit aptitude				
Phonemic coding (0-100)	78.1	22.1	71.6	84.4
Tonal imagery (0-20)	14.0	2.5	13.3	14.8
Rhythmic imagery (0-20)	15.1	2.0	14.5	15.7
B. Implicit aptitude				
FFR at F0 (0-1)	.118	.042	.105	.130
FFR at F1 (0-1)	.028	.016	.023	.033
FFR at F2 (0-1)	.024	.009	.021	.027

Table 5 summarizes the results of the Pearson correlations among the participants' explicit and implicit aptitude scores. With a view of multiple comparisons (across explicit vs implicit aptitude constructs), an alpha level was set at .025 after Bonferroni corrections. Among a total of 48 Chinese learners of English in the current study, their aptitude scores did not demonstrate significant associations, indicating that the explicit and implicit aptitude contrasts seemed to be independent of each other ($p > .025$). As predicted earlier (in Table 1), the results presented here support our assumption that the six aptitude scores used in the study could tap into the following constructs of pronunciation learning aptitude: (a) explicit/segmental (phonemic coding), (b) explicit/prosody (melodic discrimination), (c) explicit/fluency (rhythmic discrimination), (d) implicit/segmental (FFR phase-locking at F1 and F2), and (e) implicit/prosody (FFR phase-locking at F0).

Table 5. *Interrelationships between Explicit and Implicit Aptitude Scores*

	Tonal imagery	Rhythmic imagery	FFR at F0	FFR at F1[a]	FFR at F2
A. Explicit aptitude					
Phonemic coding[a]	-.16	-.05	.10	-.14	.03
Tonal imagery		.28	.12	.15	-.16
Rhythmic imagery			-.22	-.06	-.12
B. Implicit aptitude					
FFR at F0				-.04	-.07
FFR at F1[a]					.24

Note. *indicates statistical significance at $p < .025$; † indicates marginal significance at $p < .05$. [a]The data transformed via the log10 function.

Experience Profiles of Participants

Table 6 reveals that the participants' past L2 learning experience widely differed in terms of the number of hours they had practiced L2 English inside classrooms at elementary-, secondary-, and university-level schools in China (920-6840 hours). To further increase their L2 use outside of classrooms, many chose to go to cram and language conversation schools outside of their regular school curriculums (350-6080 hours). As for their more recent L2 experience during the eight to nine months of study-abroad in the UK, all of them were enrolled in a range of sciences and social sciences classes at university in London (e.g., engineering, economics, law) (360-3000 hours). At the same time, some of the participants actively sought opportunities to use L2 English at non-academic settings during their study-abroad in the UK (e.g., conversing with English-speaking friends) (0-1680 hours). Finally, the participants reported the presence and length of music training.

Table 6. *Descriptive Statistics of 48 Chinese Learners' Past/Recent L2 and Music Training Experience*

A. Past L2 experience in China	*M*	*SD*	*Range*
• Total hours of L2 use inside classroom in China	3054hr	1259	920-6840
• Total hours of L2 use outside classroom (e.g., cramming school) in China	2508hr	1286	350-6080
B. Recent L2 experience in the UK	***M***	***SD***	***Range***
• Total hours of all L2 practice inside classroom during study-abroad	906.7hr	488.6	360-3000
• Total hours of all L2 practice outside classroom during study-abroad	632.5hr	407.1	0-1680
C. Music training experience	***M***	***SD***	***Range***
• Prior music training	Yes (n = 30) No (n = 18)		
• Length of music training	2.9 yr	4.3	0-19

Aptitude, Experience and Pronunciation

To provide a general picture on how the participants' pronunciation proficiency attainment was individually related to their aptitude and experience factors, a set of Pearson correlation analyses were performed (see Table 7). To adjust for two conceptual comparisons (proficiency vs. experience; proficiency vs. aptitude), the alpha level was set at .025 via the Bonferroni correction. The results identified a moderate relationship between the participants' segmental scores and their explicit (phonemic coding) and implicit (FFR at F1) segmental sensitivity. See Figure 3 for a depiction of the difference in neural F1 encoding between participants with high and low L2 segmental proficiency.

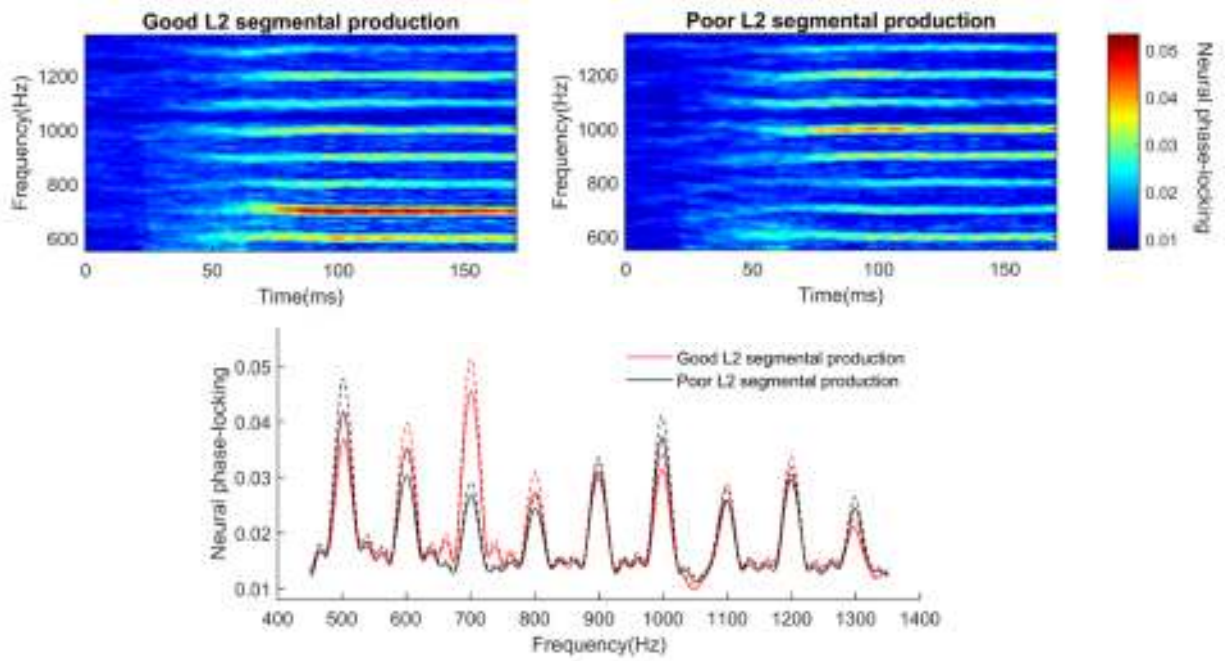

Figure 3
(Top) Inter-trial phase locking across time and frequency for participants with good (left) and poor (right) L2 segmental production scores (median split, n = 24 participants in each group). (Bottom) Inter-trial phase locking collapsed across time within a window from 60 to 170 ms in participants with good (red) and poor (black) L2 segmental production. Dotted lines indicate +1 standard error of the mean. Only the participants with good L2 segmental production showed a spectral peak around the first formant (700 Hz).

As for suprasegmental attainment, whereas the participants' perceived tempo was significantly tied to rhythmic discrimination, their prosodic (intonation, perceived tempo) and fluency (articulation rate, pause ratio) performance was significantly correlated with their recent experience inside and outside classrooms rather than any aptitude factors. No statistically significant correlations were found between the participants' pronunciation proficiency and their musical training experience.

Table 7. *Interrelationships between Aptitude, Experience and Pronunciation Attainment Scores*

	A. Aptitude Factors						B. Experience Factors				
	Phonemic coding[a]	Melodic discrimination	Rhythmic discrimination	F0 phase-locking	F1 phase-locking[a]	F2 phase-locking	Past use (inside)	Past use (outside)	Recent use (inside)	Recent use (outside)	Music training (length)
Segmentals	-.41*	.25	.18	-.14	.40*	-.03	.16	.07	.01	.27	.08
Word stress	-.09	.04	.08	-.03	.32†	-.16	-.09	.11	.24	.21	.07
Intonation	-.15	.08	.18	-.02	.09	-.08	.11	.16	.32†	.30†	.02
Perceived tempo	-.17	.27	.47*	-.17	.17	-.11	.07	.11	.37*	.21	.12
Articulation	.07	.07	.06	-.13	.01	-.03	-.02	-.11	.51*	.21	.26
Pause ratio	-.26	-.01	.02	-.01	-.20	-.10	.02	.08	-.20	-.09	-.17
Repair ratio	-.01	.12	.04	.01	-.21	-.24	-.14	-.01	.12	-.10	.20

Note. *indicates statistical significance at $p < .025$; †indicates marginal significance at $p < .05$ (Bonferroni corrected); [a]The data transformed via the log10 function.

Relative Weights of Aptitude and Experience Effects

As shown above, the participants' aptitude and experience factors were uniquely correlated with the quality of their pronunciation proficiency attainment, indicating a complex relationship between aptitude, experience and L2 pronunciation learning. In order to examine in more depth the extent to which the aptitude factor alone could predict successful L2 pronunciation learning, the participants' varied experience backgrounds need to be statistically controlled for.

A set of stepwise multiple regression analyses were performed with their pronunciation scores as dependent variables and with aptitude and experience scores as independent variables. To ensure a reliable interpretation of the regression model, a decision was made to select only aptitude and experience variables which showed significant or marginally significant correlations with any aspects of pronunciation proficiency (*Variance Inflation Factor* [VIF] < 1.02) (see Table 6). Such variables include phonemic coding, tonal and rhythmic imagery, FFR phase-locking at F1, and recent L2 use inside and outside classrooms.

Table 8. *Significant Results of Stepwise Multiple Regression Analyses Using Explicit and Implicit Aptitude and Experience as Predictors of L2 Pronunciation Attainment*

Predicted variable	Predictor variables	*Adjusted* R^2	R^2 *change*	*F*	*p*
Segmentals	Phonemic coding	.170	.170	9.398	.004
	Phase-locking at F1	.289	.119	9.163	<.001
Word stress	Phase-locking at F1	.083	.083	5.276	.028
Intonation	Recent use inside	.103	.103	5.306	.026
Perceived tempo	Rhythmic imagery	.228	.228	13.556	.001
	Recent use inside	.320	.092	10.593	< .001
Articulation rate	Recent use inside	.262	.262	16.305	< .001
Pause ratio		*n.s.*			
Repair ratio		*n.s.*			

Note. The variables entered into the regression equations included phonemic coding, rhythmic imagery, FFR and F1 and F2, and recent L2 use inside and outside classrooms

According to the results summarized in Table 8, the regression models explained 8-30% of variance in the participants' segmental (segmentals, word stress), prosodic (word stress, intonation, perceived tempo) and fluency (perceived tempo and articulation rate) proficiency. Significant aptitude-acquisition links were found between phonemic coding and segmental proficiency, FFR phase-locking at F1 and segmental/word stress proficiency, and rhythmic discrimination and perceived tempo proficiency. In contrast, the recent experience (rather than aptitude) factor appeared to play a key role in accounting for variance in the participants' intonation, perceived tempo and articulation rate proficiency.

Discussion

In the context of 48 Chinese learners of English in the UK with varied L2 learning experiences, the current study examined whether, to what degree and how the proposed framework of cognitive factors—i.e., L2 pronunciation learning aptitude—could roughly explain four dimensions of pronunciation proficiency attainment—correct pronunciation of individual sounds/words (segmentals, word stress), adequate prosody (intonation, perceived tempo); optimal fluency (perceived tempo, articulation rate, pause ratio) and self-monitoring (repair ratio). Unlike earlier aptitude studies which were exclusively concerned with explicit language learning cognition (e.g., Saito, 2017, in press), we measured L2 learners' explicit *and* implicit sensitivity to the segmental (phonemic coding, FFR at F1/F2), prosodic/intonational (tonal imagery, FFR at F0) and temporal (rhythmic imagery, FFR at F0) aspects of speech by adopting a range of behavioural (language and music aptitude tests) and neurophysiological (electroencephalography) measures.

Overall, the results of the descriptive analyses showed that approximately 11 years of English learning experience in China and the UK (i.e., 7000+ hours of L2 use inside and outside classrooms) imparted a positive influence on all dimensions of their L2 pronunciation proficiency (Flege, 2016). At the same time, the results of the correlation analyses supported our earlier prediction that the extent to which the learners ultimately improved their segmental and suprasegmental proficiency was uniquely driven by the interaction of different types of experience (past vs. recent) and aptitude (explicit vs. implicit) factors.

Segmental Sensitivity and Performance

With respect to L2 segmental proficiency, the multiple regression models revealed that the participants' correct pronunciation was primarily linked to their explicit segmental sensitivity (17.0%: phonemic coding), and secondarily associated with their implicit segmental sensitivity (11.9%: FFR at F1). Comparatively, the final quality of the participants' L2 segmental performance was not significantly related to their past nor recent experience factors. The findings here successfully replicate previous aptitude studies which identified the presence of significant aptitude (but not experience) effects on experienced L2 learners' attained segmental accuracy (Granena & Long, 2013; Saito, 2017, in press; Saito et al., 2018).

One potential reason for the relatively greater weight of the aptitude factor over the experience factor for L2 segmental attainment is difficulty of this specific L2 speech learning instance. According to previous cross-sectional (e.g., Flege, Bohn, & Jang, 1997; Saito & Brajot, 2013) and longitudinal (e.g., Munro & Derwing, 2008) investigations, L2 English learners' segmental pronunciation forms quickly become intelligible within the first year of immersion in an English-speaking country, but followed by a levelling-off. Whereas most continue to show detectable L1-related accents despite years of practice, the mastery of high-level, more nativelike segmental accuracy is limited to very few individuals; (Abrahamsson & Hyltenstam, 2009; Saito, 2013). As L2 aptitude researchers have recently emphasized, it is in the acquisition of these

relatively challenging L2 features that aptitude plays the most prominent role (Li, 2013; Skehan, 2016; see also Saito, 2017, in press).

It is likely more important to stress that our findings identified not only explicit (phonemic coding) but also implicit (FFR at F1) aptitude as a significant predictor for the participants' L2 segmental attainment. In the neurophysiology literature, the implicit sensitivity of humans to spectral and temporal features of complex speech signals—as measured using the frequency following response—serves as an anchor for various developmental phenomena in L1 acquisition related to literacy (e.g., White-Schwoch et al., 2015), normal hearing (e.g., Russo et al., 2009) and musicality (e.g., Tierney et al., 2015). Building on our precursor research (Omote et al., 2017), the electrophysiological results of the current investigation suggest that the robustness of auditory processing may be an important foundation for post-pubertal L2 speech learning as well. More specifically, our study demonstrated that neural speech encoding appeared to be independent of adult L2 learners' explicit phonetic analysis/memory (phonemic coding), and tied to relatively difficult aspects of adult L2 speech learning (segmentals, word stress).

On the one hand, the findings do agree with the dominant view in the field that post-pubertal SLA is mainly driven by explicit language learning cognition (e.g., Suzuki & DeKeyser, 2017). Our participants had exclusively practiced L2 English through a number of form-focused classes in Chinese EFL classrooms before they arrived in the UK. Given that explicit aptitude (phonemic coding) could facilitate L2 pronunciation learning to a great degree in such foreign language contexts (Saito, 2017, in press), it is not surprising to find relatively strong effects of explicit aptitude on these participants' L2 segmental attainment.

On the other hand, our study identified a significant relationship between adult learners' implicit sensitivity to speech signals (FFR) and their L2 pronunciation performance. Our findings echo the strong FFR-acquisition link clearly observed in L1 literature (e.g., White-Schwoch et al., 2015). In this regard, our study adds empirical support to the competing theoretical stance that the same cognitive factors underlying L1 acquisition—notably implicit language learning cognition—remains intact throughout the lifetime, and are therefore active in post-pubertal L2 speech learning as well (Birdsong & Molis, 2001; Bundgaard-Nielsen, Best, & Tyler, 2011; Flege, 2016; Saito, 2013, 2015).

Despite the participants' extensive form-oriented L2 experience prior to their study-abroad in the UK, all of them had been residing in the UK for eight to nine months at the time of the project. As shown in the results (see Table 6), the participants frequently accessed L2 English for meaning rather than form with various interlocutors in diverse conversational contexts. Thus, in this study, certain learners with higher implicit aptitude could have benefited more from this period of naturalistic L2 learning by processing incoming input not only explicitly (with awareness) but also implicitly (without awareness). Our argument here is harmonious with recent theoretical discussion in the L2 aptitude literature on the importance of a combination of explicit and implicit learning. Such multifaceted cognition can help L2 learners make the most of any given input/output opportunities, which is believed to be a necessary condition for the attainment

of high-level L2 proficiency (Doughty et al., 2010; Linck et al., 2013; Saito et al., 2018; Skehan, 2016).

Suprasegmental Sensitivity and Performance

When it comes to L2 suprasegmental proficiency, the participants' explicit sensitivity (rhythmic imagery) significantly accounted for 22.8% of the variance in their perceived tempo, confirming the relationship between music aptitude and L2 pronunciation learning (Li & DeKeyser, 2017). Unlike their L2 segmental attainment (closely linked to explicit and implicit aptitude), however, most of the participants' L2 suprasegmental attainment was generally predicted by their *recent* L2 use inside and outside classrooms during their study-abroad in the UK, regardless of their *past* English-as-a-Foreign-Language experience in China (9.2-26.2 %). The findings here concur with the theoretical claims (e.g., Ellis, 2006) and empirical evidence (e.g., Saito & Hanzawa, 2016) that SLA is adaptively sensitive to the quantity, quality and recency of input, as form and meaning connections become stronger in accordance with how often certain linguistic items are practiced in the most immediate contexts.

The findings also echo previous observations on the relatively salient effects of experience on L2 suprasegmental (rather than segmental) learning. Whereas L2 segmental learning is a slow, gradual process especially beyond the initial rate of learning stage (Flege, 2016), L2 learners' suprasegmental accuracy and fluency improve substantially and continuously for an extensive period of time, as long as they use the target language on a daily basis (Mora & Valls-Ferrer, 2012; Trofimovich & Baker, 2006; Saito, 2015). This strong relationship between experience and L2 suprasegmental learning could be arguably linked to the fact that the suprasegmental quality of L2 speech more directly affects listeners' successful comprehension than the segmental quality does (Isaacs & Trofimovich, 2012), and that L2 learners are assumed to intentionally or intuitively prioritize the acquisition of L2 suprasegmentals (rather than segmentals) as a function of increased experience (Derwing, Munro, Thomson, & Rossiter, 2009).

Interestingly, no significant associations were found between FFR fundamental frequency encoding and L2 suprasegmental attainment. These results suggest that the process and product of adult L2 suprasegmental learning may derive from explicit rather than implicit cognition. However, the findings need to be interpreted with much caution, since F0 phase-locking in the study may not have fully captured natural prosody in English. To this end, different FFR metrics, such as Gamma phase-locking (80 Hz), may be needed for the analysis of participants' sensitivity to lower frequency (Omote et al., 2017). Additionally, the relationship between individual differences in characteristics of the FFR and performance in various auditory tasks remains imperfectly understood. Prior research has found that FFR phase-locking at the F0 is linked to the ability to consistently synchronize to a metronome (Tierney & Kraus, 2013) and rapidly adapt to stimulus perturbations while synchronizing (Tierney & Kraus, 2016). This suggests that FFR phase-locking at lower frequencies could be an index of the precision with

which the auditory system represents the timing of sound, a skill which may be useful for extracting temporal cues to prosodic features such as phrase boundaries.

However, FFR phase-locking has not been found to relate to the ability to remember and reproduce temporal patterns, a skill which instead correlates with inter-trial consistency in slower cortical response to sound (Tierney, White-Schwoch, MacLean, & Kraus, 2017), suggesting that integration of rhythmic information across time relies more upon cortical than subcortical processing. Cross-trial consistency in cortical responses such as the passive auditory ERPs and cortical tracking of slow changes in amplitude envelope and pitch contour, therefore, may be more promising measures of implicit suprasegmental proficiency. Future studies need to conceptualize, elaborate and validate more reliable aptitude measures by which to measure L2 learners' capacities to process a wide range of low frequencies to produce word stress, intonation and fluency with adequate rhythmic timings.

Future Directions

To conclude, we would like to emphasize a strong call for more L2 speech research of this kind in order to further examine the cognitive and perceptual correlates of successful L2 pronunciation learning with a larger number of participants with varied levels of proficiency and experience, and different pairings of L1 and L2 backgrounds. Given the exploratory nature of the project, several methodological limitations need to be acknowledged with an eye towards future replication studies. First, the current study was a cross-sectional investigation of the aptitude profiles of intermediate-to-advanced level participants with varied L2 learning experience backgrounds. To unravel the relative impacts of explicit and implicit aptitude on L2 pronunciation learning, future studies can adopt longitudinal, pre-and-posttest-designs (cf. Saito et al., 2018). Such studies will shed light on whether and to what degree high explicit and implicit aptitude learners can differentially benefit from two essentially different L2 learning conditions—(a) naturalistic immersion with ample opportunities to use the L2 meaningfully with native and non-native speakers on a regular basis; vs. (b) form-focused lessons in foreign language settings without many conversational opportunities outside of the classroom.

Relatedly, such future studies should also longitudinally examine the intricate link between aptitude and experience. In the field of SLA and music education, there is empirical evidence that learners' aptitude test scores (e.g., phonemic coding, music aptitude) are unlikely to change dramatically over time, suggesting that such explicit aptitude can be a relatively stable trait (e.g., Carroll, 1962; Gordon, 1995). As shown in the current study, participants' aptitude and experience profiles *independently* related to L2 speech performance ($VIF < 1.02$), indicating they are essentially different factors of SLA. When it comes to FFR measures, however, the neurophysiology literature has shown individual variability when researchers compare participants with *substantially* different backgrounds (Bidelman, Gandour, & Krishnan, 2011 for tonal vs. non-tonal language users; Krizman et al., 2015 for simultaneous vs. sequential bilinguals). These studies suggest that FFR can be modulated by long-term experience to a certain degree. To our knowledge, however, no empirical studies have probed whether and to

what degree FFR can change when learners engage in a short-term, but intensive exposure to foreign language (e.g., study abroad); examining this topic is crucial, as it will allow us to further understand the extent to which FFR measures can serve as predictors rather than the result of L2 phonological attainment.

Another interesting direction concerns the role of aptitude in the acquisition of advanced L2 phonology, especially among more experienced L2 learners at the later stage of SLA. It is important to remember that the length of study-abroad among the participants in the current study was only eight months, and that their pronunciation performance was far below the nativelike norm (e.g., see Table 2 for their pronunciation ratings of around 500-600 out of 1000). This indicates that these participants had much room for improvement. In the previous nativelikeness literature in SLA, certain adult L2 learners have been identified as demonstrating high-level L2 pronunciation proficiency, which native listeners cannot perceptibly distinguish from other native samples. Whereas these participants typically have processed an extensive amount of L2 experience (> 10 years) (DeKeyser, 2013; Saito, 2013) together with strong professional and integrative motivation (Moyer, 2014) and some form of explicit language learning cognition (Granena & Long, 2013), it has remained unclear the extent to which implicit language learning cognition—the driving force for successful L1 and early L2 acquisition—can still explain the incidence of exceptional L2 speech learning after puberty (cf. Linck et al., 2013).

Third, although the current study exclusively drew on production measures, it is notable that any change in a learner's representational system first impacts the perception phase prior to the production phase in both L1 and L2 acquisition (Flege, 2016). It would thus be intriguing for future studies to elucidate the role of explicit and implicit aptitude in L2 learners' perception performance, especially when they are exposed to natural and synthetic tokens varying in the F1 × F2 × F3 domain (Flege et al., 1997) and duration of F1 transition (Underbakke, Polka, Gottfried, & Strange, 1988), under various lexical conditions (i.e., target sounds in frequent words vs. infrequent words: Flege, Takagi, & Mann, 1996), and with reaction time instruments (Ingvalson, McClelland, & Holt, 2011). In terms of analyses, such future studies should also highlight not only the global dimensions of L2 speech, but also specific segmental, prosodic and temporal features difficult for a particular group of L2 learners. For instance, one of the most well-researched topics in L2 speech learning is the acquisition of the English /ɹ/ and /l/ contrast by Japanese learners (for a review, Bradlow, 2008). Few Japanese learners have been reported to attain nativelike performance in perceiving and producing English /ɹ/ and /l/ due to their significant lack of sensitivity to highly complex speech signals in F2 and F3 and articulatory configurations (simultaneous constrictions in labial, alveolar and pharyngeal areas of vocal tract) (e.g., Flege et al., 1996; Ingvalson et al., 2011; Saito, 2013; Saito & Brajot, 2013). To provide a full-fledged picture of this specific aptitude-acquisition link, it would be intriguing for future studies to explore the extent to which Japanese learners' cognitive and perceptual individual differences could explain the attainment of high-level English /ɹ/-/l/ performance. As a result, such follow-up studies will allow us to evaluate the replicability and robustness of our aptitude framework at a fine-grained level.

Finally, our finding that neural encoding of spectral peaks correlates with L2 segmental production suggests that neural sound processing, as measured using EEG, provides an alternate method of measuring implicit language learning aptitude, complementing behavioural approaches. Future research into neural correlates of language learning aptitude could investigate other EEG metrics that could be similarly useful. Slow (< 8 Hz) rhythms in the EEG signal, for example, entrain to the amplitude envelope (Doelling, Arnal, Ghitza, & Poeppel, 2014) and pitch contour (Meyer, Henry, Gaston, Schmuck, & Friederici, 2017) of speech, and the fidelity of this entrainment is related to L1 language abilities in children (Power, Colling, Mead, Barnes, & Goswami, 2016). This measure, therefore, is a promising candidate for a neural foundation of implicit sensitivity to segmental and suprasegmental aspects of speech.

Endnote

[1]Our validation study (Saito et al., 2017) showed that expert raters' segmental, prosodic and temporal scores significantly corresponded to the actual number of segmental errors (deletion/substitution of L2 vowels/consonants) and of prosodic errors (misplacement and absence word stress and intonation), and articulation rate and pause frequency in L2 speech, respectively (for similar results, see Bosker et al., 2013).

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language learning, 59*, 249-306.

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing research*, *245*, 35-47.

Anderson, S., Skoe, E., Chandrasekaran, B., & Kraus, N. (2010). Neural timing is linked to speech perception in noise. *Journal of Neuroscience*, *30*, 4922-4926.

Anvari, S., Trainor, L., Woodside, J., & Levy, B. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology, 83*, 111-130.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second language acquisition. *Journal of Memory and Language, 44*, 235-249.

Boersma, D., & Weenink, P. (2017). *Praat: Doing phonetics by computer version 6.0.29.* Retrieved from http://www.praat.org

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.

Bradlow, A. R. (2008). Training non-native language sound patterns. In J. Hansen & M. Zampini (Eds.), *Phonology and second language acquisition* (pp. 287-308). Philadelphia, PA: John Benjamins.

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and cognitive processes*, *27*(7-8), 1205-1224.

Bundgaard-Nielsen, R., Best, C., & Tyler, M. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition, 33,* 433-461.

Carroll, J.B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training, research, and education* (pp. 131-151). New York: Wiley.

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

Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm: rhythmic priming and audio-motor training affect speech perception. *Acta Psychologica, 155*, 43-50.

Coffey, E. B., Herholz, S. C., Chepesiuk, A. M., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature communications*, *7*.

Daltrozzo, J., & Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience, 21,* 1882-1892.

Darcy, I., Mora, J., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*. Advanced online publication. doi: 10.1111/lang.12161

Darcy, I., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, *40*, 63-72.

DeKeyser, R. M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning*, *63*, 52-67.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *12*, 1-16.

Derwing, T. M. & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*, 476-490.

Derwing, T. M., Munro, M. M., 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.

Doelling, K., Arnal, L., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage, 85*, 761-768.

Doughty, C. J., Campbell, S. G., Mislevy, M. A., Bunting, M. F., Bowles, A. R., & Koeth, J. T. (2010). Predicting near-native ability: The factor structure and reliability of Hi-LAB. In *Selected proceedings of the 2008 Second Language Research Forum* (pp. 10-31). Cascadilla Press.

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

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics. 27*, 1-24.

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., Bohn, O-S., & Jang, S. (1997). The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics*, 25, 437-470.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *Journal of Acoustical Society of America, 99,* 1161-1173.

Gass, S., & Varonis. (1984). The effect of familiarity on the comprehensibility of non-native speech. *Language Learning, 34*, 65–89.

Gordon, E. E. (1995). *Manual: Musical Aptitude Profile*. Chicago: GIA Publications.

Gottfried, T. L. (2007). Music and language learning: Effect of musical training on learning L2 speech contrasts. In O.-S. Bohn and M. J. Munro (Eds.) *Language Experience in Second Language Speech Learning*: In honour of James Emil Flege (pp. 221-237). Amsterdam: John Benjamins.

Granena, G. (2013). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning, 63*, 665-703.

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.

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.

Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of phonetics*, *39*, 571-584.

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

Kormos, J. (2014). *Speech production and second language acquisition*. Routledge.

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.

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

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

Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, *97*, 634-654.

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition.* doi: 10.1017/S027226311500042X

Li, M., & DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Learning*. Advanced online publication. doi: 10.1017/S0272263116000358

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.

Liu, F., Patel, A., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain, 133*, 1682-1693.

Major, R. (2008). Transfer in second language phonology: A review. In J. Hansen Edwards & M. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 63–94). Amsterdam: John Benjamins.

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

Meyer, L., Henry, M,. Gaston, P., Schmuck, N., & Friederici, A. (2017) Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cerebral Cortex, 27*, 4293-4302.

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

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.

Moyer, A. (2014). Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics*, *35*, 418-440.

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics, 35*, 463-482.

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

Ogawa, Y. (2009). *Developing a music aptitude test for school children in Asia*. CD-ROM

Omote, A., Jasmin, K., & Tierney, A. (2017). Successful non-native speech perception is linked to frequency following response phase consistency. *Cortex*. Advanced online publication. doi: 10.1016/j.cortex.2017.05.005

Patel., A., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. (1998). Processing syntactic relations in language and music: an event-related potential study. *Journal of Cognitive Neuroscience, 10,* 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.

Power, A., Colling, L., Mead, N., Barnes, L., & Goswami, U. (2016). Neural encoding of the speech envelope by children with developmental dyslexia. *Brain and Language, 160*, 1-10.

Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., & Kraus, N. (2008). Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clinical Neurophysiology*, *119*, 1720-1731.

Rvachew, S., & Grawburg, M. (2006). Correlates of phonological awareness in preschoolers with speech sound disorders. *Journal of Speech, Language, and Hearing Research*, *49*, 74-87.

Saito, K. (2013). Age effects on late bilingualism: The production development of /ɹ/ by high-proficiency Japanese learners of English. *Journal of Memory and Language, 69*, 546-562.

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

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. (in press). The role of aptitude in second language segmental learning: The case of Japanese learners' English /ɹ/ pronunciation attainment in classroom settings. *Applied Psycholinguistics.*

Saito, K., Suzukida, Y., & Sun, H. (2018). Aptitude, experience and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition*. DOI: 10.1017/S0272263117000432

Saito, K., & Brajot, F. (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*, 847-863.

Saito, K., & Hanzawa, K. (2016). Developing second language oral ability in foreign language classrooms: The role of the length and focus of instruction and individual differences. *Applied Psycholinguistics, 37*, 813-840.

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.

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

Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2013). Stability and plasticity of auditory brainstem function across the lifespan. *Cerebral Cortex*, *25*, 1415-1426.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, *17*, 675-681.

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.

Tierney, A., & Kraus, N. (2013). The ability to move to a beat is linked to the consistency of neural responses to sound. *Journal of Neuroscience, 33*, 14981-14988.

Tierney, A., & Kraus, N. (2016). Getting back on the beat: links between auditory-motor integration and precise auditory processing at fast time scales. *European Journal of Neuroscience, 43*, 782-791.

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.

Tierney, A., White-Schwoch, T., MacLean, J., & Kraus, N, (2017). Individual differences in rhythm skills: links with neural consistency and linguistic ability. *Journal of Cognitive Neuroscience, 29*, 855-868.

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.

Underbakke, M., Polka, L., Gottfried, T. L., & Strange, W. (1988). Trading relations in the perception of /ɹ/ and /l/ by Japanese learners of English. *Journal of the Acoustical Society of America, 84*, 90-100.

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

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience*, *10*, 420.

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

Yilmaz, Y., & Grañena, G. (2016). The role of cognitive aptitudes for explicit language learning in the relative effects of explicit and implicit feedback. *Bilingualism: Language and Cognition*, *19*, 147-161.

Zhu, L., Bharadwaj, H., Xia, J., & Shinn-Cunningham, B. (2013). A comparison of spectral magnitude and phase-locking value analyses of the frequency-following response to complex tones. *The Journal of the Acoustical Society of America*, *134*, 384-395.

Appendix A: Speaking Tasks (Versions A and B)

Version A

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 B

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.*

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

Appendix B: 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.

Appendix C: Onscreen Labels for Pronunciation Measures