To be published in *Journal of Memory and Language* (Elsevier)

Domain-General Auditory Processing as an Anchor of Post-Pubertal L2 Pronunciation Learning: Behavioural and Neurophysiological Investigations of Perceptual Acuity, Age, Experience, Development, and Attainment

Kazuya Saito[1]

Magdalena Kachlicka

Hui Sun

Adam Tierney

Abstract

In the cognitive psychology literature, auditory processing has been extensively researched and suggested as a foundation of first language acquisition in childhood. This study tests an emerging theoretical view that the same faculty underpins post-pubertal L2 pronunciation learning. A total of 100 late English-Polish bilinguals in the UK with diverse age and experience backgrounds were assessed for their ability to represent various characteristics of sounds via behavioural and neurophysiological measures. Subsequently, the participants' biographical backgrounds and auditory processing profiles were compared to various dimensions of their L2 pronunciation proficiency. According to the results of mixed-effects modeling analyses, individual differences in participants' L2 pronunciation proficiency were *equally* accounted for by age (age of arrival), experience (length of residence), and auditory processing (encoding, reproduction). Within the current dataset, the degree of auditory precision was negatively associated with participants' chronological age (19-45 years). The findings suggest that earlier age of onset may allow them to take advantage of more precise auditory processing, which in turn helps them to make the most of every input opportunity throughout extensive immersion experience, leading to more advanced L2 phonological skills in the long run.

Key words: auditory processing, second language, speech production, age, experience

---

[1] This study was funded by Leverhulme Trust Research Grant (RPG-2019-039) and Arnold Bentley New Initiatives Fund. We would like to thank Yui Suzukida, Shungo Suzuki, and the anonymous reviewers from the *Journal of Memory and Language* for their useful and constructive comments on the earlier versions of the manuscript.
Corresponding Author: Kazuya Saito (k.saito@ucl.ac.uk)

Highlights

- We recruited 100 late Polish-English bilinguals with varied proficiency levels.
- We compared the perceptual and biographical correlates of their L2 outcomes.
- Domain-general auditory processing was the primary determinant of L2 success.
- The findings suggest that audition helps drive language acquisition throughout life.

Introduction

The perceptual and cognitive foundations of language acquisition have been extensively investigated in the field of cognitive psychology. While the extent to which such underlying mechanisms are specifically devoted to learning language has remained open to debate (Campbell & Tyler, 2018), one influential view is that language-related processing involves neural networks which also underlie general-purpose learning (see Hamrick et al., 2018 for an overview). One such domain-general ability, which has received much scholarly attention, is *auditory processing*, defined as encoding, remembering, and proceduralizing time and frequency characteristics of sounds. Whereas different types and combinations of auditory information are processed on domain-specific levels during various learning behaviors (e.g., language, speech, music and emotion), they may essentially draw upon early auditory processing stages which are domain-general (i.e. precise representation of spectro-temporal details).

Under the framework of the auditory-processing-deficit theory (Goswami, 2015; Mueller et al., 2012; Tallal, 2004; Tierney & Kraus, 2014; Wright et al., 2000), individual differences in low-level auditory processing, speech perception, and language development are intricately interwoven. When learners have difficulty in integrating basic auditory information (e.g., frequency, duration, intensity, and amplitude rise time), such a deficit may prevent a fine-grained acoustic analysis of speech at a phonemic, phonological and syllabic level, resulting in a range of global language problems (e.g., dyslexia). Extending the audition-based account of language acquisition, there is emerging evidence that the same faculty explains success in post-pubertal second language (L2) speech *perception* performance among the somewhat limited number of relatively inexperienced learners (e.g., Omote et al., 2017; Kachlicka et al., 2019; Saito et al., 2019). The main objective of the current investigation is to examine the generalizability of the topic to the initial, mid, and final state of L2 speech *production* performance among a total of 100 late English-Polish bilinguals in the UK with different profiles of immersion experience, age of onset, and auditory ability levels.

Background

Domain-General Auditory Processing, Age, Experience and L1 Acquisition

Given that the auditory channel is the primary source of input for most language learners, the initial step of language learning involves converting acoustic input into linguistic information available for subsequent phonological, lexical and morphosyntactic encoding. For example, learners integrate a range of acoustic cues to identify phonemic and phonological units of speech (e.g., lower third and second formants, longer duration for English [r] in "present" rather than English [l] in "pleasant"; Espy-Wilson et al., 2000). Similarly, the presence of stress is associated with a variety of acoustic cues (e.g., pitch movements, longer duration, increased amplitude and spectral balance for English prominence "PREsent" rather than "preSENT"; Plag et al., 2011). More robust analyses of sounds could facilitate more accurate and faster word recognition. An appropriate lexical representation can be selected more easily from all competing lexical candidates thanks to the more rapid analyses of frequency, recency, and probability of input (e.g., "present" rather than "prescient,""prescind"; Norris & McQueen, 2008). As learners encounter more exemplars in diverse contexts, semantic, morphosyntactic, and pragmatic specifications of each lexical construction could be further updated, expanded, and entrenched (Tomasello, 2000).

During L1 acquisition, the development of basic perception ability is uniquely shaped by a range of age- and experience-related factors. Among normal hearing individuals, encoding of sound continues to improve with age up to 7-10 years, followed by a gradual decline for the rest of life (Clinard et al., 2010; Skoe et al., 2015). For other auditory processing abilities, such as the ability to align one's movements with auditory signals (audio-motor integration), the developmental trajectory reaches its peak around the mid-20s (Thomson et al., 2015).
So far, a number of studies have demonstrated correlations between auditory processing abilities and speed of acquisition of various L1 processing skills, such as phonological awareness (Casini et al., 2018; Goswami et al., 2011; Won et al., 2016), reading ability (Boets et al., 2011; Gibson, Hogben, & Fletcher, 2006), and literacy development (Boets et al. 2008; White-Schwoch et al., 2015). As a result, some scholars have suggested that auditory processing measures may serve as a diagnostic tool for certain kinds of language delay and impairment, such as autism spectrum

disorders and dyslexia (Russo et al., 2008; Hornickel & Kraus, 2013). Furthermore, training and interventions that focus on auditory processing might be an effective way of minimizing language deficits (Warrier et al., 2003).

Importantly, the *causal* relationship between auditory processing and L1 *attainment* has remained open to debate (Rosen & Manganari, 2001). It has been pointed out, for example, that not all individuals with dyslexia have auditory deficits (Gokula et al., 2019; Halliday et al., 2017). This could arguably be because some auditory deficits may eventually resolve (Rosen, 2003), and/or because auditory processing problems are not a core factor of language impairment, but a peripheral symptom of broader attentional difficulties (Snowling et al., 2018). It has also been suggested that toddlers with auditory problems may eventually develop normal linguistic proficiency through the use of compensatory mechanisms (see Jasmin et al., 2019 for dimension-selective attention).

In essence, the existing literature has shown a consensus that auditory processing is “the gateway to spoken language” (Mueller et al., 2012, p. 15953) with its impact on language acquisition being most clearly observed when learners begin to encounter, parse and process a new language (McArthur & Bishop, 2005). Of course, long-term language attainment could be affected by a number of cognitive and biographical factors at later stages of language learning. Since sound perception may be a bottleneck for language learning, individual differences in auditory processing could have a significant impact on every dimension and stage of L1 acquisition to some degree. In the current study, we aim to examine the relationship between individual differences in auditory processing and L2 speech production learning in adulthood.

Second Language Speech Acquisition in Adulthood

Second language speech is generally coloured with foreign-accents, especially when learners start learning a target language after puberty. Compared to other dimensions of language (e.g., lexicogrammar), the incidence of nativelike L2 pronunciation attainment (i.e., the main focus of this study) is extremely rare (Granena & Long, 2013). In many theoretical models in L2 phonetics (e.g., Best & Tyler, 2007 for Perceptual Assimilation Model; Flege, 2016 for Speech Learning Model), the level of learning difficulty has commonly been ascribed to L1 and L2 phonetic distance. Certain L2 learners are reported to achieve nativelike L2 speech proficiency, when their L1is phonetically similar to the target language (e.g., Dutch learners of English; Bongaerts et al,, 1997). When the distance between the L1 and L2 phonetic systems is relatively large, the rate and ultimate attainment of L2 speech learning could be mediated by the quantity and quality of practice. That is, L2 learners' speech proficiency can continue to be more targetlike, as long as they regularly use the target language for an extensive period of time (Flege et al., 1995). As for the relationship between perception and production, there is a consensus that L2 learners first become more capable of perceiving new phonetic and phonological contrasts at more fine-grained levels, whether they primarily attend to articulatory or acoustic features of speech signals (for the theoretical discussion on the nature of perception, see Perceptual Assimilation vs. Speech Learning Models). Subsequently, the improved perception skills are assumed to activate relevant articulatory configurations, leading to more intelligible and accurate L2 production performance (Sakai & Moorman, 2017).

More recently, scholars have begun to point out that post-pubertal L2 speech learning is characterized by a great deal of individual variation even within same L1-L2 pairings (for a comprehensive review, Trofimovich et al., 2015). Some learners are able to achieve a high level of L2 oral proficiency, while others show a tremendous amount of difficulty in their attempts to do so. Whereas experience is a necessary condition for both L2 speech learning, it is noteworthy that all experience-related factors together cannot fully explain the degree of success, accounting for a small-to-medium amount of variance in the outcomes of L2 speech learning in different contexts (e.g., R^2 = .10−.20 in Saito, 2015a). These differences in learning outcomes exist not only because of the amount of time spent practicing the target language, but also because some learners are more perceptually and cognitively adept at making the most of every opportunity for input and output. This consequently leads to greater gains from the same type of L2 experience, resulting in more advanced L2 proficiency in the long run (Doughty, 2019).

To date, scholars have extensively examined a set of explicit, intentional and analytic learning abilities among successful L2 learners in foreign language classroom settings, referred to as *foreign language aptitude*. Previous research has shown that such aptitude is instrumental to the acquisition of relatively difficult linguistic features within a short period of time, arguably because these abilities could help L2 learners better analyze, memorize and internalize what they have learned from explicit language instruction (see Skehan, 2019). Foreign language aptitude was originally conceptualized as a composite construct of cognition *specific* to the process and product of highly form-oriented foreign language education, which does not necessarily represent various types of L2 learning beyond classroom contexts (e.g., short, mid and long-term immersion). Additionally, such domain-specific aptitude conflates a range of abilities related to perception, awareness, memory, and attentional control. To this end, a growing number of scholars have emphasized the importance of establishing a more theoretically and empirically sound model of aptitude for various stages of adult L2 learning in both classroom *and* naturalistic settings (e.g., Linck et al., 2013 for their attempts to include domain-general short- and long-term memory as a predictor for the part of new L2 language aptitude).

Following the auditory deficit hypothesis in L1 acquisition (Goswami, 2015), we argue that domain-general auditory processing may serve as a component of aptitude for adult L2 speech learning. In fact, it might be even more critical than it is for L1 acquisition, arguably due to the *quantitative* and *qualitative* differences in L1 and L2 learning experience. In the former context, many L1 toddlers likely benefit from ample exposure to socially interactive and variable language for an extensive period of time. Such input-rich conditions may help those with auditory deficits overcome subsequent language problems using compensating strategies, such as taxing other cognitive skills to a greater degree (e.g., executive functions; Snowling et al., 2018) or/and relying to a greater degree on alternative perceptual channels (e.g., amusics using duration rather than pitch cues for identifying sentence stress in English; Jasmin et al., 2019). With respect to adult L2 learning, however, the amount of input and output opportunities is generally limited even under immersion conditions (relative to L1 acquisition) (Flege, 2016). As a result, very few learners have sufficient conversational experience for the development of successful compensatory mechanisms. As proposed in the aptitude framework in L2 learning (Doughty, 2019; Skehan, 2019), it is in such a demanding task (speaking the L2) that individual differences in language learning aptitude, including auditory processing skills, are particularly consequential.

Furthermore, we echo McAllister et al.'s (2002) theoretical discussion that noticing and encoding individual acoustic dimensions in order to map continuous variability along these dimensions onto discrete categories could be a more difficult and complex task, and thus a more critical skill, in L2 compared to L1 speech acquisition. Under this view (i.e., L1-L2 feature hypothesis), L2 speech acquisition takes place in a common linguistic space, wherein the L1 system has already been established. When encountering new sounds in an L2, learners filter, decode and analyze such information through the L1-based acoustic representations. Here, not only do learners need to restructure the L1-specific cue weightings (e.g., relying more on durational rather than pitch information for Chinese speakers' English prosody acquisition; Jasmin et al., 2020), but also detect new acoustic dimensions which are not regularly used as a primary cue for L1 phonological contrasts (e.g., F3 for Japanese speakers' English [r]-[l] acquisition; Saito, 2013).

Motivation for Current Study

Given that precise auditory processing serves as an anchor of L1 acquisition, this paper seeks to propose and provide evidence for this ability as an emerging framework of aptitude for the initial, mid and later stages of post-pubertal L2 speech learning. Surprisingly little is known about the role of domain-general auditory processing abilities in adult L2 speech learning. Under laboratory settings, some empirical studies have shown that individuals with greater auditory processing demonstrate more learning after brief training on sounds and words that they have never heard (e.g., Chandrasekaran et al., 2011; Kempe et al., 2012; Wong & Perrachione, 2007) and L2 sound contrasts which they have had some experience (e.g., Lengeris & Hazan, 2010 for Greek speakers' L2 English vowel acquisition).

Recently, our precursor work took the first step towards testing the role of auditory processing in naturalistic, conversational and meaning-oriented L2 speech learning among 25 Japanese residents with a mixed amount of immersion experience in the UK (Omote et al., 2017), 40 moderately experienced Polish residents (length of residence [LOR] < 5 years) (Kachlicka et al., 2019), and 48 Chinese international students (LOR = 1 year) (Saito et al., 2019). In essence, our precursor research has provided some preliminary evidence for the predictive power of auditory processing for L2 speech acquisition in adulthood. While the findings are promising, our precursor work led to more questions which future research needs to answer to obtain a better understanding of the intricate relationship between auditory processing, age, experience, and L2 speech acquisition.

First, the audition-acquisition link above was exclusively concerned with one comprehension dimension of L2 phonological skills (phoneme perception), and based on the results of a highly controlled task format (forced-choice identification), where adult L2 participants could carefully focus on monitoring their accurate perception of L2 vowels without any communicative pressure (Kachlicka et al., 2019; Omote et al., 2017). It has remained unclear whether and to what degree auditory processing relates to *production* ability on *spontaneous*levels, which many consider to be a barometer of how adult L2 learners actually use language in daily-life settings (Piske et al., 2011). Second, the previous literature was exclusively concerned with short- and mid-term L2 residents (LOR < 5 years) (Kachlicka et al.; Saito et al., 2019). We have yet to discover how auditory processing relates to different phases of L2 speech learning. In particular, it remains unclear the extent to which both biographical and auditory processing factors interact to determine the outcomes of highly experienced L2 speakers' pronunciation development and attainment (e.g., LOR > 5 years) (Abrahamsson & Hyltenstam, 2009).

Linking auditory processing, experience, speech perception, and production further generates ample theoretical implications. For example, Flege's speech learning model has stated that one's capacity used in L1 acquisition remains intact, and applies to L2 speech acquisition (Flege, 2016; McAllister et al., 2002). With sufficient exposure to new sounds, learners eitherrevise existing phonetic categories (assimilation) or create new categories (dissimilation), which will help first hear better and then produce better. However, the model has yet to address what perceptual and cognitive mechanisms underlie suchL2 speech assimilation/dissimilation, and how the nature of the mechanisms changes at different stages of L2 speech learning (short, mid vs. long-term immersion).

Whereas the prior work has hinted at a possibility that it is domain-general auditory precision ability that governs both L1 and L2 speech *perception* learning (e.g., Omote et al., 2017), examining the generalizability of the topic to L2 speech *production* would test our hypothesis that auditory category learning is characteristic of multiple dimensions (perception, production) and phases (rate of learning, ultimate attainment) of L2 speech learning. To extend this line of research, the current study aimed to compare the auditory, biographical, and linguistic profiles of 100 L1 Polish users of L2 English with different lengths of residence (0.1-20 years) and different ages of arrival in the UK (17-36 years). Specifically, two research questions and predictions were formulated.

1. To what degree do biographical factors relate to participants' auditory processing ability?

As reviewed earlier, the degree of auditory acuity can be subject to the influence of participants' age and experience. More specifically, we predict that certain L2 learners may demonstrate more precise auditory perception, not only because they are younger in chronological age (Skoe et al., 2015), but also because they have had more extensive bilingual experience (Roncaglia-Denissen et al., 2016) and longer music training (Tierney et al., 2015).

2. To what degree do auditory and biographical factors interact to determine success in L2 pronunciation proficiency?

In the previous literature, scholars have found the process and product of L2 speech learning to be influenced by a range of learner-extrinsic factors, such as the duration of immersion in an L2 speaking environment (Length of Residence; Trofimovich & Baker, 2006), the frequency of daily L2 use (Current L2 Use; Derwing & Munro, 2013), and the timing of the first intensive exposure to a target language (Age of Arrival; Flege et al., 1995). In the current investigation, our prediction is that learner-intrinsic factors—i.e., precise auditory perception—can further account for the remaining variance in the outcomes of adult L2 speech learning. As shown in the L1 acquisition literature (Mueller et al., 2012), greater auditory sensitivity facilitates phonological, lexical, and morphosyntactic encoding, leading to the attainment of more advanced global language skills. Such links between auditory processing and proficiency may be stronger for participants with earlier age of acquisition and longer length of residence, since when L2 learners have arrived in L2 speaking environment during earlier adulthood, they can access more precise auditory perception (Skoe et al., 2015). Moreover, learners who arrive in an L2 speaking environment earlier can spend more years immersed in L2 input, and so can make the most of every input and output opportunity in order to boost the rate and ultimate attainment of L2 pronunciation proficiency.

Method

Participants

Data collection took place at a university in London, UK. To access as many Polish residents as possible, the project was widely advertised via a range of social media outlets, community websites, bulletin boards in Polish stores and supermarkets, and group emails to many university-level schools across the city. While 200+ participants initially contacted us, two screening criteria were set up to narrow the scope of the participants eligible for the current project: (a) the participants needed to have arrived at English speaking countries after the age of 17 (i.e., late L2 learners); and (b) they needed to report their main language of communication either at work/school or home at the time of the project as English (indicating they were regular, constant and motivated L2 users).

According to the results of a learner questionnaire, all the participants started learning L2 English as a mandatory subject at school in Poland from Grades 1 to 7 (*M* = 9.9 years; *SD* = 3.4; *Range* = 0-15).[2] After years of foreign language education, they immigrated to the UK at different ages after puberty (*M* = 22.4 years; *Range* = 17-36 years). At the time of the project, the participants' chronological age widely varied (*M* = 30.2 years; *Range* = 19-45 years). Following the procedure adopted in the Language Contact Profile (Freed et al., 2004), we surveyed the quantity and quality of participants' *current* L2 use as per different types of interlocutors (fluent vs. non-fluent) and contexts (professional, social vs. family). In terms of what percentage they spent per day using L2 English especially with fluent users, the participants spent a varied amount of time using L2 English at work/school (*M* = 80.1%; *Range* = 10-100%), in social settings (*M* = 62.5%; *Range* = 10-100%), and at home (*M* =56.3%; *Range* = 0-100%). In terms of

[2] To reduce the number of predictors (minimizing multicollinearity problems), and following the suggestions in L2 education research, we did not include age of learning as a predictor in the subsequent analyses. The relationship between the onset of foreign language education and L2 speech learning has remained controversial (e.g., Muñoz, 2014). In fact, the participants' age of learning in Poland was not significantly associated with any dimensions of auditory precision and L2 pronunciation proficiency in the current dataset ($p > .05$). This could be arguably due to the fact that the type of input that learners receive in foreign language education varies widely across different school contexts in terms of source (there is large variability in teachers' oral fluency and general proficiency) and quantity (not all teachers use the target language as the language of communication in the classroom).

music experience, 37 reported that they had received more than six months of regular training on music instruments (*M* = 2.2 years; *Range* = 0-25 years).

The participants were further divided into two sub-groups to reflect two different stages of L2 speech learning—i.e., Interlanguage and Attainment. The former group comprised 50 participants whose length of residence was somewhat limited (*M* = 2.8 years; *Range* = 0.1-5 years) (see Trofimovich & Baker, 2006 for their operationalization of short- and mid-term immersion). The remaining 50 participants comprised the attainment group who had been using L2 English in the UK for an extensive period of time (*M* = 11.8 years; *Range* = 6-19 years) (for the same cut-off point of LOR beyond 6 years, see Birdsong & Molis, 2001). The interlanguage group (*M* = 24.8 years; *Range* = 19-39 years) was significantly younger than the attainment group (*M* = 35.7 years; *Range* = 24-45 years), $t = -11.336$, $p < .001$. However, their age of arrival profiles (*M* = 21.7 vs. 23.1 years; *Range* = 18-36 vs. 17-32 years) were comparable, $t = -1.885$, $p = .067$. As for their L2 use profile, there was no significant difference between the groups ($p < .05$). All the data collection, including biographic background interview, L2 speech tasks, and auditory processing measures (in this order), was individually administered with the researcher in a quiet room at the university. No participants reported any hearing problems.

Auditory Processing Measures

In the current study, three abilities relevant to speech perception and production were highlighted in order to look at participants' individual differences in auditory processing from multiple angles: (a) explicit acuity, (b) pre-conscious neural encoding, and (c) audio-motor integration.

Explicit Acuity

When learning to perceive new sound patterns, listeners first need to detect acoustic properties of these sounds in order to establish, revise and maintain robust perceptual representations. Such explicit encoding ability has traditionally been measured behaviourally through psychoacoustic tests, where learners are explicitly asked to discriminate and identify two different sounds on the basis of a number of auditory dimensions, including frequency, duration, and amplitude rise time—i.e., explicit acuity (Surprenant & Watson, 2001).

Stimuli. A total of four psychoacoustic A×B discrimination tests were used to assess the extent to which participants could perceive four acoustic dimensions of complex tones: formant, pitch, amplitude rise time, and duration. For the pitch, amplitude rise time, and duration discrimination tasks, a total of 100 four-harmonic complex tones were prepared via custom MATLAB scripts with a fundamental frequency set at 330 Hz and the amplitude of each harmonic set at 40 dB. The duration of the standard stimulus was 500 ms. To avoid the perception of transients (i.e. clicks), sound amplitude was ramped at the onset and endpoint of the stimulus (15 ms each). Throughout the 100 tokens, the target acoustic dimension for each test ranged with a step of 2.5 ms in duration (252.5-500 ms), 2.8 ms in amplitude rise time (17.8-300 ms) and 0.3 Hz in F0 (330.3-360 Hz), respectively. For the formant discrimination thresholds, a total of 100 complex tones were created. The duration of each token was 500 ms with a fundamental frequency of 100 Hz and harmonics up to 3000 Hz. Two points of 15ms rise time were inserted at the beginning and endpoint of the stimulus. Using the technique of a parallel formant filter bank (Smith, 2007), three formants were generated at 500 Hz, 1500 Hz and 3000 Hz. The target dimension of F2 varied between 1502 Hz and 1700 Hz with a step of 2 Hz. All the audio stimuli used in the discrimination tasks are deposited and available in IRIS (https://www.iris-database.org).

Procedure. After participants heard a sequence of three tones with an inter-stimulus interval of 0.5 s, they selected whether the first or the third tone differed from the second one by pressing the number “1” or “3.” Following Levitt’s (1971) adaptive threshold procedure, the level of difficulty changed from trial to trial according to participants’ performance. The tests started from Level 50 (out of 100). When their response was incorrect, the size of the difference became wider by a degree of 10 steps (making the discrimination task easier). When three consecutive correct responses were made, the size of the difference became smaller by a degree of 10 steps (making the discrimination task more difficult). The step size decreased when the direction of difficulty between trials reversed—i.e., when an increase in acoustic difference (easier) was followed by a decrease (more difficult), or vice versa. After the first reversal, the step size changed from 10 to 5, and then after the second reversal from 5 to 1. The tests stopped either after 70 trials or eight reversals. Participants' auditory processing score was determined by averaging the stimulus levels at which the reversals occurred after the third reversal.

Pre-Conscious Neural Encoding

Since the psychoacoustic test format inevitably entails some degree of attention and memory, the extent to which it actually captures perceptual (rather than cognitive) ability has been questioned (Snowling et al., 2018). More recently, scholars have begun to measure the degree of *pre-conscious* auditory encoding by using electrophysiological responses to variations and changes in sounds (e.g., frequency following response, Coffey et al., 2016; mismatch negativity, Näätänen et al., 2007). These auditory indices are considered to capture "pre-attentional" auditory processing, because they are automatically generated during passive listening tasks, and are relatively unaffected by cognitive state (Varghese, Bharadwaj, & Shinn-Cunningham, 2015)—i.e., pre-conscious encoding.

In the current study, following the procedure in Tierney and Kraus (2013), the participants' neural encoding of the fundamental frequency (100 Hz), first formant (720 Hz), and second formant (1240 Hz) of a synthesized speech token (/da/) was analyzed by examining the frequency-following response to sound, an electrophysiological measure which captures the spectral and temporal characteristics of the evoking stimulus (Skoe & Kraus, 2010).

Stimulus. The speech token /da/ (170 ms) was synthesized via a Klatt-based synthesizer. The first five ms of the sound contained an onset burst. The rest of the sound was voiced with a 100 Hz fundamental frequency. While the first, second and third formants shifted during the transitional period between 5 to 50 ms (400 to 720 Hz, 1700 to 1240 Hz, 2580 to 2500 Hz), these three formants stayed constant during the steady state between 50 and 170 ms (720 Hz, 1240 Hz, 2500 Hz). The fourth, fifth, and sixth formants stayed constant throughout the stimulus at 3300 Hz, 3750 Hz, and 4900 Hz, respectively. All the audio stimuli used in the Frequency Following Response tasks are deposited and available in IRIS (https://www.iris-database.org).

Procedure. The /da/ sound was presented repeatedly (6300 times over the course of 20 minutes) through insert earphones (ER-3; Etymotic Research) at 80 dB with 81 ms interstimulus intervals. Stimuli were presented at alternating polarities; in other words, every other stimulus was inverted, i.e. flipped upside down, so that positive samples became negative and vice versa. This procedure enables separate analysis of two different aspects of speech (Aiken & Picton, 2008): combining the brain's response to both polarities reveals its encoding of the lower-frequency amplitude envelope (i.e. the fundamental frequency), while inverting the response to one of the two polarities prior to analysis reveals the brain's encoding of the higher-frequency temporal fine structure of speech (i.e. the first and second formants; see the Data Analyses section for more details). During the task, the participants were encouraged to focus on reading a book of their choice in a relaxed environment, instead of paying special attention to sound properties. Continuous electrophysiological data were recorded using a BioSemi EEG system with a sample rate of 16384 Hz and open filters. A montage of five electrodes was used, with the left and right earlobes as unlinked references (linked offline during data analysis), two electrodes on the head serving as ground, and a single active electrode on the middle of the top of the head (i.e. at Cz).

Data Analyses. All neurophysiological analyses were conducted using custom-written software in MATLAB. Recordings were first filtered between 70 to 2000 Hz using a first-order Butterworth filter to remove the cortical evoked response to sound and isolate the frequency-following response. The recording was then segmented from -30 ms to 210 ms with respect to stimulus presentation. Trials containing amplitude spikes of >35 micro-volts were rejected as artifacts, and the first 2500 artifact-free responses to each stimulus polarity were selected for the main analysis. The degree of accurate neural sound encoding was calculated by inter-trial phase-locking. This analysis allows us to reveal each participant's phase consistency at a particular frequency level—100 Hz, 720 Hz, 1240 Hz in the current study. It is similar to spectral analysis of the average response, but provides a better signal-to-noise ratio, especially at higher frequencies, enabling more precise measurement of individual differences in neural encoding of speech (Zhu, Bharadwaj, Xia, & Shinn-Cunningham et al., 2013).

First, a Hanning windowed fast Fourier transform for each trial was calculated over two different response time windows. For analysis of encoding of the F0, a window from 10 to 180 ms after stimulus onset was used, while for analysis of encoding of the second formant, a window from 60 to 180 ms after stimulus onset was used. This procedure generates, for each frequency from 1 Hz to the Nyquist frequency (> 8000 Hz), a complex vector containing information about both amplitude and phase of the neural response. We then discarded amplitude information by converting each vector to a unit vector with a length of 1. These vectors were then averaged together. Vectors with similar phases, when added together, will lead to an average vector with greater length, while vectors with dissimilar phases will lead to a shorter average vector. The length of the average vector, therefore, was taken as the inter-trial phase consistency. This measure ranges from 0 (phases uniformly distributed) to 1 (phases identical across trials).

For analysis of the F0, the phase consistency calculation was conducted identically across all trials from both polarities. This process highlights the neural representation of the lower-frequency amplitude envelope of the sound. This is because half of the stimuli were inverted (i.e. the stimuli were presented in alternating polarities), and inner hair cell transduction of sound is half-wave rectified (in other words, hair cells discharge only in response to the rarefaction phase of the stimulus). As a consequence, when the responses to non-inverted and inverted stimuli are averaged together, the result emphasizes the response to the amplitude envelope but cancelling out the higher-frequency fine structure (see Figure 1, middle). On the other hand, when the responses to the non-inverted and inverted stimuli are subtracted, the result emphasizes the response to the higher-frequency temporal fine structure (the higher harmonics and speech formants), cancelling out the lower-frequency amplitude envelope (see Figure 1, bottom). (See Aiken & Picton, 2008 for a more detailed description of the rationale behind this procedure.) As a result, for analysis of F1 and F2, the phase consistency calculation was conducted after reversing the phases for the inverted polarity stimuli by 180 degrees.

Neural encoding of the fundamental frequency was quantified as the degree of inter-trial phase coherence between 80 and 120 Hz within the 10 to 180 ms time window. Neural encoding of the first formant was quantified as the degree of inter-trial phase coherence at the 7th harmonic, i.e. between 680 and 720 Hz within the 60 to 180 ms time window. Because the second formant was located roughly halfway between the 12th and 13th harmonics, neural encoding of the second formant was quantified as the average of the degree of inter-trial phase coherence between 1180 and 1220 Hz and between 1280 and 1320 Hz within the 60 to 180 ms time window. See Figure 1 for a display of the frequency content of the stimulus, and the robustness of neural encoding of speech across frequency in participants with good versus poor accuracy performance (median split).

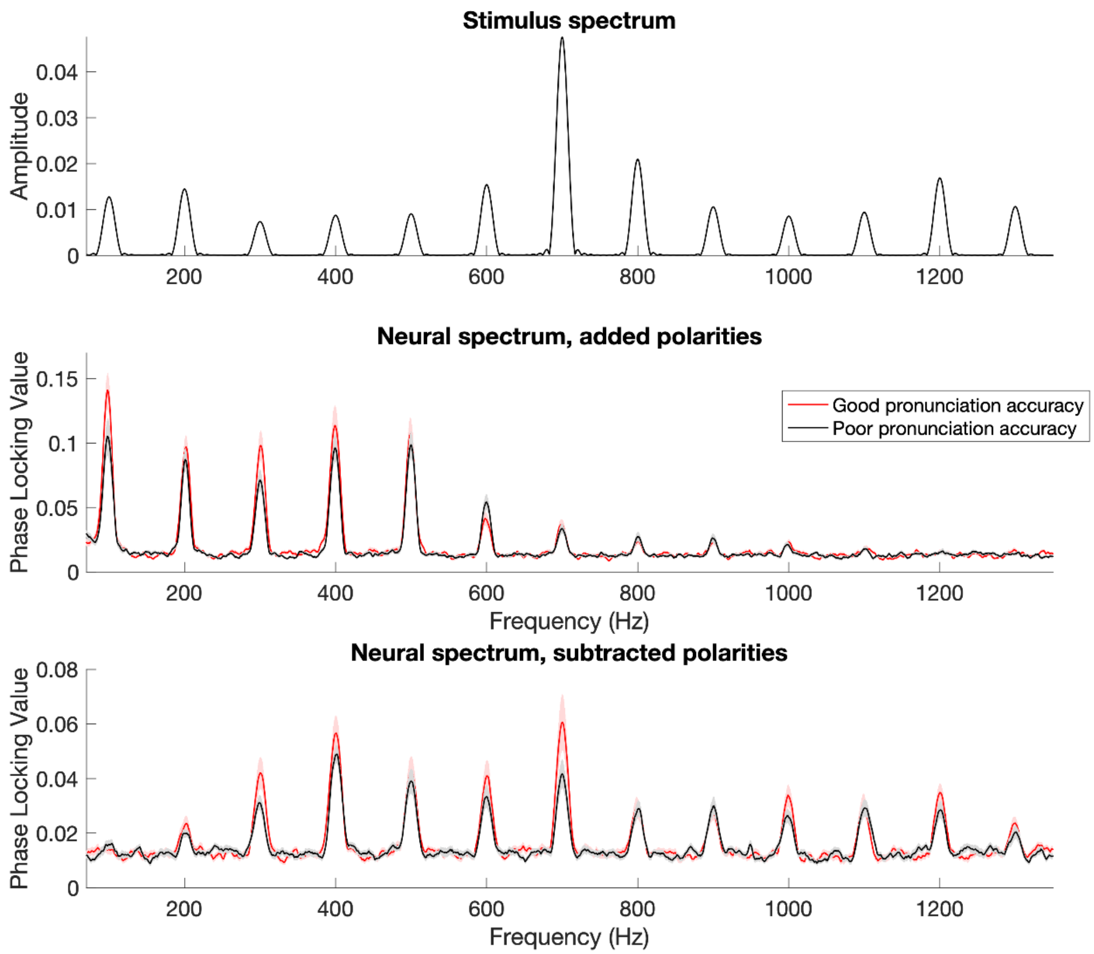

Figure 1
(Top) Frequency content of stimulus used to elicit electrophysiological responses. Neural data analysis focused on three frequencies: F0 (100 Hz), F1 (720 Hz), and F2 (1240 Hz). (Middle) Inter-trial phase locking across frequencies in Polish L1 participants with good and poor English L2 pronunciation accuracy. In this analysis, phases were combined across the inverted and non-inverted stimuli (i.e. polarities were "added"). Good (red lines) and poor (black lines) participants are the top and bottom quartiles (n = 25 for each). Phase locking is a unit-less measure that extends from 0 (no consistency in phase whatsoever between trials) to 1 (identical phase across trials). The shaded region indicates standard error. (Bottom) Inter-trial phase locking across frequencies in the subtracted polarities analysis, in which phases from inverted stimuli were reversed by 180 degrees prior to calculation of the phase locking value.

Auditory-Motor Integration

To produce new sounds, learners need to draw, remember, and consolidate connections between temporal and spectral patterns and their relevant motor movements—i.e., auditory-motor integration (Flaugnacco et al., 2014, 2015). One way to measure non-verbal auditory-motor integration is to ask learners to tap along in synchrony with a beat, or to remember and reproduce a sound pattern after a short time lapse (Tierney & Kraus, 2014; Tierney et al., 2017). To this end, two tests were implemented in the current study, both of which required participants to listen to sounds, assess sound timing, and produce appropriate responses by drumming on a conga drum.

In a metronome synchronization test, two blocks were included at each of three inter-onset-interval rates: 0.667, 0.5, and 0.333 ms (i.e., 1.5, 2, and 3 Hz). Each block of the synchronization test consisted of 40 presentations of the drum sound. After being presented with an isochronous interstimulus interval, participants were asked to synchronize to the beat as soon as they could, by tapping along on the drum such that their drum hits occurred as close as possible in time to the stimulus drum onsets. While the first 20 representations were used as practice, the last 20 were used to assess variability of synchronization. Drum hit onset times were marked via custom MATLAB scripts, for which two parameters could be adjusted for each participant: an amplitude threshold and a relaxation time. Any time point that exceeded the amplitude threshold was marked as a drum hit, as long as an amount of time greater than the relaxation time had elapsed since the last time point exceeding the amplitude threshold. These two parameters were adjusted manually by viewing the raw sound wave and the marked drum hits, to ensure that each drum hit was marked and that every marked hit corresponded to a real drum onset. For each participant, synchronization variability was measured via the standard deviation of the difference in time between the drum hit and the nearest stimulus onset, divided by the inter-onset-interval rate of the block (i.e., the coefficient of variation; Wagenmakers & Brown, 2007). Synchronization variability was then averaged across trials.

In a rhythm synchronization test, a total of 4 rhythm patterns of drum hits were prepared based on Povel and Essens (1985). Each rhythm pattern consisted of 16 segments, each 200 ms in duration, which could contain either a rest (7 segments) or a drum hit (9 segments). Each rhythmic pattern consisted of the same sequence of inter-drum-onset intervals, but the order of these intervals varied across patterns. These intervals were: five 200 ms, two 400 ms, one 600 ms, and one 800 ms. For each rhythm pattern (3.2 s in duration), participants were presented with eight repetitions, and asked to synchronize as soon as they could. For each segment of the target rhythm, rhythmic accuracy was calculated by determining whether the participant produced a drum hit or a rest in a 200 ms window centred on the onset of the segment. Subsequently, the participants' response was compared to the sequence of hits and rests of the target rhythm. The accuracy score was calculated by dividing the number of segments produced correctly by the total number of segments. Scoring was performed from the onset of the second repetition.

All the audio stimuli used in the metronome and rhythm synchronization tasks are deposited and available in IRIS (https://www.iris-database.org).

Reliability of Auditory Processing Tests

For the test-retest reliability of the acuity and reproduction tasks, see our Brief Report (Saito, Sun, & Tierney, 2020a), wherein we recruited and asked a total of 30 L1 and L2 English users to take the same tests twice. Their initial and second test scores demonstrated relatively strong associations for acuity ($r = .701$, $p < .001$) and reproduction ($r = .863$, $p < .001$). In another investigation (Sun, Saito, & Tierney, forthcoming), the test-retest reliability of FFR was tested and confirmed among a total of 46 Chinese learners of English ($r = .831$, $p < .001$). The findings here indicate that the tests can reliably tap into various dimensions of individuals' supposedly stable, trait-like auditory processing abilities.

L2 Speech Materials

To assess L2 pronunciation proficiency, many scholars have exclusively relied on controlled speech tasks (e.g., word, sentence and paragraph reading) so that they can guide participants to pronounce target sounds while carefully controlling for surrounding phonetic and lexical contexts (for a review, see Piske et al., 2011). It has been shown that adult L2 learners can demonstrate more target-like performance when tested through controlled rather than spontaneous tasks, arguably because the former format allows learners to consciously monitor their correct speech forms and draw on explicit metalinguistic knowledge (e.g., Lee et al., 2015). To tap into the present state of post-pubertal L2 proficiency, therefore, many scholars have emphasized the importance of adopting spontaneous speech tasks (e.g., picture description, oral interviews), wherein learners are pushed to equally attend to not only the phonological but also the temporal, lexical, grammatical, and discoursal domains of language with their primary focus on conveying their intended message in the most effective and efficient way (Spada & Tomita, 2010).

To this end, we adopted a cartoon narrative task, which has been widely used in previous L2 speech research (e.g., Derwing & Munro, 2013). The task included an eight-frame picture sequence where a man and a woman bumped into each other on a busy street corner in a cosmopolitan city, switched their similar-looking suitcases by mistake, and realized it later when they opened their suitcases at their different destinations (see **Supporting Information-A**). Following the procedure in L2 pronunciation research (Lee et al., 2015), the participants first had one minute for planning and then proceeded to describe the content of the event within two minutes. The first 30 seconds of each participant's speech were excised, normalized and stored in a WAV file. A total of 100 speech samples were prepared for subsequent speech analyses.

Expert Rater Judgements

In line with the framework of L2 pronunciation proficiency proposed by Saito and Plonsky (2019), and the training procedure that we developed and validated in our precursor research (Saito et al., 2017), four different dimensions of L2 pronunciation proficiency were assessed by expert judges—(a) segmentals (substitution, omission, or insertion of individual consonants or vowels); (b) word stress (misplaced or missing lexical stress); (c) intonation (appropriate and varied versus incorrect and monotonous use of pitch); and (d) optimal speed (speed of utterance delivery).

A total of five linguistically trained coders were recruited (3 males, 2 females). Four coders were native speakers of American and British English, and one coder was a highly proficient Polish speaker of English. All of them reported extensive experience in L2 speech analyses of this kind (by participating in our previous similar projects), and a high level of familiarity with British English and Polish-accented English.

All the speech samples were played to the coders in a randomized order via custom MATLAB scripts, and coders used a free moving slider on a computer screen to assess segmentals, word stress, intonation, and optimal speed. The slider was initially placed in the middle of each scale and raters were asked to use the whole scale continuum to indicate their judgement. When the slider was placed at the leftmost end of the continuum, labeled with a frowning face (indicating negative), it was recorded as “0.” When the slider was placed at the rightmost end of the continuum, labeled with a smiling face (indicating positive), it was recorded as “1000.” The coders were given a replay button to listen multiple times until they felt satisfied with their judgements.

Each rating session took place individually in the presence of the researcher. First, the coders received instructions on the goal of assessment (i.e., analyzing four different dimensions of L2 oral proficiency among 100 Polish users of English), and on the rating rubrics. To familiarize themselves with the procedure, the coders practiced with three speech samples which were not included in the main dataset. After rating each sample, they explained their decisions and received feedback from the researcher to check their understanding of the constructs. Finally, they proceeded to rate the main dataset of 100 audio samples with a 10-minute intermission halfway through. For training scripts and onscreen labels, see **Supporting Information-B**. The rater questionnaire and training scripts used in L2 speech judgements are deposited and available in IRIS (https://www.iris-database.org).

The coders’ inter-rater reliability was calculated via the Cronbach alpha analyses. The five raters demonstrated a varied degree of agreement for segmentals (α = .82), word stress (α =.78), intonation (α = .71) and optimal speed (α = .80). All the values can be considered satisfactory beyond Larson-Hall’s (2010) field-specific benchmark (α = .70). All the coders’ ratings were averaged across each sample for each speech category.

Acoustic Analyses

Following the notion of utterance fluency (Tavakoli & Skehan, 2005), the temporal aspects of L2 speech were analyzed for breakdown, speed, and repair fluency measures, which are assumed to correspond to four stages of L2 speech production—conceptualization, formulation, articulation, and monitoring. For breakdown fluency, the number of filled pauses (e.g., ah, oh, eh) and unfilled pauses (more than 250 ms of silence; see Bosker et al., 2013) were first calculated and divided by the total number of words. We calculated two different types of pause ratio (end- and mid-clause pauses) since it has been suggested that these different locations index different phases of speech production: conceptualization (what to say) and linguistic encoding (how to say it) (Lambert et al., 2017). For speed fluency, articulation rate was calculated by dividing the total phonation time (without all filled pauses) by the total number of syllables (i.e., articulation rate). For repair fluency, the number of repetitions and self-corrections was divided by the total number of words. Two linguistically trained coders conducted acoustic analyses on 10 speech samples (10% of the dataset). Since their inter-rater agreement was generally high ($\alpha > .90$), the first coder completed the rest of the fluency analyses.

Results

For the purpose of replication and reanalysis of the current dataset, the auditory processing, biographical backgrounds and L2 speech scores are summarized, deposited and available in IRIS (https://www.iris-database.org).

Constructs of Auditory Processing

In the current study, we used a total of nine measures of different dimensions of participants' auditory processing ability (explicit vs. pre-conscious; encoding vs. reproduction; spectral vs. temporal). According to the results of normality analyses (Kolmogorov-Smirnov test), all the auditory processing scores demonstrated a significant degree of skewness (D =.097-.231, $p < .05$), and thus underwent transformation via a log10 function. To make the directions of two temporal reproduction scores align with each other (less variability vs. more accuracy), audio-motor accuracy scores were inverted.

The first objective of our statistical analyses was to identify any broad patterns underlying these measures via an exploratory factor analysis. Direct Oblimin rotation was chosen, given that some of the variables may have been correlated with each other. The factorability of the entire dataset was considered adequate according to Bartlett's test of sphericity ($\chi^2 = 138.064$, $p < .001$) and the Kaiser-Meyer-Olkin measure of sampling adequacy

(.695). A “three-factor” solution was suggested with an eigenvalue beyond 1.0, accounting for 57.539% of the variance in the outcomes of the auditory processing measures.

All the factor loadings are summarized in Table 1. Factor 1 was labeled as “Explicit Acuity,” as the measures clustered here tapped into participants’ sensitivity to various dimensions of sounds (all discrimination task scores) and their ability to reproduce sound patterns accurately, as tested in behavioral tasks. Factor 2 was labeled as “Pre-Conscious Encoding,” as this factor covered two of the three FFR tasks which were designed to assess participants’ sensitivity to formant information. Factor 3 was labeled as “Temporal Reproduction,” as this factor corresponded to participants’ performance in the two audio-motor integration tasks (variability, accuracy).

Summary of Results. As conceptualized earlier, participants’ auditory processing was composed of three different abilities: (a) discrimination of subtle differences in acoustic signals (smaller values indicate greater sensitivity); (b) encoding of different acoustic dimensions at subcortical levels (greater values indicate more accurate encoding); and (c) reproduction of novel rhythmic patterns (greater values indicate more stable integration). The resulting factor scores (Explicit Acuity, Pre-Conscious Encoding, and Temporal Reproduction) were normally distributed (Kolmogorov-Smirnov test: D = .056 to .082, p > .05), and were used for the subsequent analyses.

Table 1

Summary of a Three-Factor Solution Based on a Factor Analysis of Auditory Acuity and Audio-Motor Integration Scores

	Factor 1: Explicit Acuity	Factor 2: Pre-Conscious Encoding	Factor 3: Temporal Reproduction
Cumulative%	28.953%	44.170%	57.539%
FFR at F0 (pre-conscious encoding)	.043	.489	.459
FFR at F1 (pre-conscious encoding)	-.066	**.704**	-.162
FFR at F2 (pre-conscious encoding)	.055	**.752**	.097
Duration (explicit acuity)	**.583**	-.263	.491
Formant (explicit acuity)	**.710**	.032	-.107
Pitch (explicit acuity)	**.772**	-.065	-.066
Rise time (explicit acuity)	**.789**	.076	.011
Audio-motor integration (variability)	.176	-.080	**-.732**
Audio-motor integration (accuracy)[a]	**.531**	.153	**-.538**

Note. All loadings > .5 were highlighted in bold; [a] for inversed

Constructs of L2 Pronunciation proficiency

The presence of underlying factors was investigated among a total of eight speech measures which were assumed to reflect segmental, prosodic and temporal aspects of participants' L2 pronunciation proficiency. The results of the Kolmogorov-Smirnov test revealed that the following four measures were significantly skewed ($p < .05$): perceived optimal speed, the ratio of pauses between clauses, the ratio of pauses within clauses, and the ratio of repair. We transformed the data via a log10 (1010-x) function for perceived optimal speed (because the data showed substantially negative skewness), and used the inverse transformation method for the pause and repair ratio (because the data showed severe positive skewness). For the transformed data, the directionality was reversed.

Next, participants' speech scores (rater judgements and acoustic analyses) were submitted to a factor analysis using Direct Oblimin rotation with an eigenvalue set to 1.0. The factorability of the entire dataset was confirmed via two tests: Bartlett's test of sphericity ($\chi^2 = 541.961$, $p < .001$) and the Kaiser-Meyer-Olkin measure of sampling adequacy (.819). A decision was made to identify a “three-factor” solution which accounted for 76.3% of the total variance in participants' L2 pronunciation proficiency. A value of 0.6 was used as the threshold coefficient for practically significant factor loadings.

As shown in Table 2, Factor 1 was labeled as “Accuracy,” as it featured all the variables which represented participants' composite ability to accurately pronounce individual sounds, correctly assign stress at both word and sentence levels and deliver speech at an adequate speed. Factor 2 was labeled as “Traits,” as this factor captured the ratio of clause-final pauses and self-repair (correction, repetition), both of which have been found to be tied to individual differences in attention and control, and personality traits rather than L2 proficiency (Zuniga & Simard, 2019). Factor 3 was labeled as “Fluency,” as it featured two temporal measures which have been suggested to relate to linguistic encoding (pauses within clauses) and automatization (articulation rate) specific to L2 learning (Lambert et al., 2017).

Summary of Results. Our comprehensive set of eight speech measures tapped into three different abilities—(a) pronouncing L2 sounds more accurately; (b) speaking with an optimal speed; and (c) making less repetition and self-corrections. In terms of directionality, all the factor scores are incremental, with greater values indicating better performance. The resulting factors (Accuracy, Fluency, Trait) were normally distributed (Kolmogorov-Smirnov test: D = .042 to .074, p > .05) and used for the subsequent analyses.

Table 2

Summary of a Three-Factor Solution Based on a Factor Analysis of the Rater Judgements and Acoustic Analyses

	Factor 1: Accuracy	Factor 2: Trait	Factor 3: Fluency
Cumulative%	51.837%	61.112%	78.761%
Segmentals	**.980**	.018	-.086
Word stress	**.967**	.040	-.012
Intonation	**.903**	-.048	.090
Perceived optimal speed[a]	**-.737**	.024	-.260
Articulation rate	.157	.121	**.786**
Mid-clause pause ratio[b]	-.030	-.070	**.934**
Clause-final pause ratio[b]	.197	**.663**	-.080
Repair ratio[b]	-.181	**.829**	.087

Note. All loadings > .5 were highlighted in bold; [a] for log10 transformed (directionality reversed); [b] for inverse transformed (directionality reversed)

Biographical Correlates of Auditory Processing

The next objective of our statistical analyses was to investigate the relationship between participants' biographical backgrounds (age, experience) and their auditory processing profiles. To this end, mixed-effects linear models were constructed using the lm functions from the lme package (Version 1.1-21; Bate et al., 2015) in the R statistical environment (R Core Team, 2018). Here, we aimed to identify an optimal combination of biographical factors that could explain the maximum amount of variance in participants' three different constructs of auditory processing scores—explicit acuity, pre-conscious encoding, and temporal reproduction. The fixed effects included six predictor variables—Chronological Age, Age of Arrival, Length of Residence, Current L2 Use, and Music

Experience. As for Current L2 Use, we calculated this composite variable by averaging the amount of L2 use (%) with fluent speakers at the time of the project in work, social, and home settings. In terms of Music Experience, a dummy code was given to categorize whether participants had received more than six months of music training (n = 63 for "no," n = 37 for "yes"). Finally, we added Experience Group as random effects. Note that LOR may not have captured the non-linear relationship between experience and L2 speech acquisition: whereas much L2 learning quickly takes place within the first few years of immersion, it is likely to plateau beyond the initial rate-of-learning stage (subject to a great deal of individual variation) (DeKeyser, 2013). Following the sub-group category in the current study, this learner group factor (Experience Group) consisted of 50 interlanguage learners (LOR < 5 years) and 50 attainers (LOR > 6 years).

For the purpose of comparability, all the predictor variables (measured on different scales) were converted into z scores. For the evaluation of the models, the Pairwise Likelihood Ratio Test was employed (Baayen, 2008). The selection of the predictors was based on whether they decreased Akaike's information criterion (AIC) of the model. A series of Chi-square tests were performed to check whether AIC values differed between two models in comparison. The deviance information criterion (DIC), and the Bayesian information criterion (BIC) were also presented for reference. Backward elimination was conducted such that all the possible model combinations of main and interaction effects were tested until the most optimal model accounting for the largest amount of the variance was identified. Since chronological age and age of arrival were highly correlated, r = .667 (resulting in high variance inflation factors in each analysis context > 5), we constructed separate models by including only one of them variables. Then, we determined the final model based on which variable (chronological age vs. age of arrival) led to lower AIC values. In the following subsections, a series of linear mixed-effects regression models were constructed to examine which biographical background variables could best explain three different types of participants' varied auditory processing profiles—Explicit Acuity, Pre-conscious Encoding, and Temporal Reproduction.

Explicit Acuity. The model fit significantly changed, resulting in the minimum AIC value (278.69), when we removed all the predictor variables (Chronological Age, AOA, LOR, Current L2 Use, Music, and Experience Group) except Music Experience (χ^2= 10.852, p< .001). As summarized in Table 3, the results indicated that explicit sound encoding ability was significantly (but weakly) related to the presence/absence of the participants' music training experience (R^2 = .06).

Table 3

Summary of the Final Model Explaining the Biographical Predictors of Explicit Acuity

Predictors	Estimate	*SE*	*t*-value	*p*	95% *CI*	
					Lower	Upper
(Intercept)	-.001	.098	-.0140	.989	-.220	.218
Music Experience	-.252	.099	-2.549	.012*	-.446	-.058
Random effect (intercepts)	Variance	*SD*				
Experience Group	< .001	< .001				
Information criterion	Estimate					
LogLikelihood	-135.34					
DIC	270.69					
AIC	278.69					
BIC	289.03					
R^2	Estimate					
Marginal	.06					
Conditional	.06					

Pre-Conscious Encoding. In general, removing predictor variables did not significantly improve the model fit, χ^2= 9.87, p =.079. A series of model comparisons suggested that the model featuring only chronological age as a predictor (AIC = 279.33, BIC = 289.67) had the lowest AIC and BIC values compared to others (e.g., AIC = 281.46, BIC = 307.31 for a model including all the variables). As summarized in Table 4, Chronological Age (β=-.46, t = -3.291, p =.002) accounted for 15% of the variance in the participants' pre-conscious encoding. The results here suggest that neural encoding of spectral characteristics of sound, especially the higher-frequency speech formants, may decrease as participants age (19-45 years) (Clinard et al., 2010; Skoe et al., 2015).

Table 4

Summary of the Final Model Explaining the Biographical Predictors of Pre-Conscious Encoding

Predictors	Estimate	*SE*	*t*-value	*p*	95% *CI*	
					Lower	Upper
(Intercept)	.003	.362	0.010	.994	-.808	.815
Chronological Age	-.460	.140	-3.291	.002*	-.712	-.051
Random effect (intercepts)	Variance	*SD*				
Experience Group	.243	.494				
Information criterion	Estimate					
LogLikelihood	-135.66					
DIC	271.33					
AIC	279.33					
BIC	289.67					
R^2	Estimate					
Marginal	.15					
Conditional	.33					

Temporal Reproduction. After a series of model comparisons, the model fit was considered best when all the predictor variables were kept (χ^2= 6.812, p =.009), compared to the minimal model (i.e., a random effect with no fixed effect). According to this model (summarized in Table 5), 15% of the variance in participants' temporal reproduction ability was significantly associated with a range of variables related to age, length of residence, current L2 use and music training experience. That is, the earlier adult L2 learners started naturalistic immersion in an L2 speaking environment, and the more bilingual and music experience they have, the better their temporal reproduction ability.

Table 5

Summary of the Final Model Explaining the Biographical Predictors of Temporal Reproduction

Predictors	Estimate	*SE*	*t*-value	*p*	95% *CI*	
					Lower	Upper
(Intercept)	-.087	.670	-0.130	.92	-1.776	1.601
Chronological Age	-.731	.339	-2.158	.033*	-1.379	-.085
LOR	.792	.270	2.934	.004*	.277	1.306
Current L2 Use	.207	1.025	2.024	.046*	.012	.403
Music Experience	.344	.091	3.794	<.001*	.171	.517
Random effect (intercepts)	Variance	*SD*				
Experience Group	.076	.276				
Information criterion	Estimate					
LogLikelihood	-130.36					
DIC	260.71					
AIC	276.71					
BIC	297.39					
R^2	Estimate					
Marginal	.14					
Conditional	.21					

Summary of Results. With respect to the biographical correlates of auditory perception, three overall patterns were found—(a) explicit acuity was weakly related to past music experience; (b) pre-conscious encoding was moderately associated with chronological age; and (c) temporal reproduction was accounted for by a range of experience and age factors.

Relationship Between Auditory Processing, Age, Experience and L2 Pronunciation proficiency

The final objective of the statistical analyses was to explore how biographical factors (Chronological Age, AOA, LOR, Current L2 Use, Music Experience) and auditory processing factors (Explicit Acuity, Pre-Conscious Encoding, Temporal Reproduction) relate to three different aspects of L2 pronunciation proficiency among 100 Polish residents in the UK: accuracy, fluency and traits.

Accuracy. A series of model comparisons identified the best fitted model (χ^2=1.063, p <.001) which featured four main effects (Explicit Acuity, Pre-Conscious Encoding, Temporal Reproduction, AOA, LOR) and one interaction effect (Explicit × AOA), accounting for 37% of the variance in participants' L2 segmental and prosodic accuracy performance (see Table 6). According to standardized beta values, the participants' L2 accuracy was equally associated with two biographical factors (β = -.402 for AOA, β = .271 for LOR) and three auditory processing factors (β = -.215 for Explicit, β = .225 for Pre-Conscious, β = .206 for Reproduction).

Table 6

Summary of the Final Model Explaining the Perceptual and Biographical Correlates of L2 Segmental and Prosodic Accuracy Performance

Predictors	Estimate	*SE*	*t*-value	*p*	95% *CI*	
					Lower	Upper
(Intercept)	.080	.085	0.946	.347	-.113	.259
Explicit Acuity	-.215	.084	-2.569	.012*	-.374	-.056
Pre-Conscious Encoding	.225	.084	2.673	.009*	.065	.385
Temporal Reproduction	.206	.084	2.441	.016*	.046	.365
AOA	-.402	.086	-4.671	<.001*	-.566	-.239
LOR	.271	.087	3.130	.002*	.097	.437
Explicit × AOA	-.273	.099	-2.760	.007*	-.461	-.086
Random effect (intercepts)	Variance	*SD*				
Experience Group	< .001	< .001				
Information criterion	Estimate					
LogLikelihood	-113.32					
DIC	226.64					
AIC	248.64					
BIC	277.08					
R^2	Estimate					
Marginal	.38					
Conditional	.37					

Interestingly, the interaction effect between Explicit Acuity and AOA reached statistical significance (p = .008), hinting at a possibility that explicit auditory processing may have even stronger predictive power among certain learners with unique AOA profiles. To disentangle this interaction effect, participants were divided into two sub-groups based on median value of AOA (22 years): *n* = 55 Early Adulthood Arrivals (AOA = 17-22 years) and *n* = 45 Late Adulthood Arrivals (AOA = 23-36 years). As visually summarized in Figure 2, the speech accuracy of late adulthood arrivals was significantly correlated with the degree of their explicit acuity (r = -.512, p< .001). However, this correlation was not significant among early adulthood arrivals (p> .05).

Taken together, the results suggest several interesting patterns. Adult L2 learners who demonstrate more accurate L2 segmental and prosodic proficiency are likely to have not only more extensive L2 experience (earlier AOA, longer LOR), but also more precise auditory processing abilities. When it comes to late adulthood arrivals, who may not benefit from the age advantage, it is explicit acuity that helps them to compensate and achieve a high level of L2 pronunciation proficiency.

A. Early Adulthood Arrivals B. Late Adulthood Arrivals

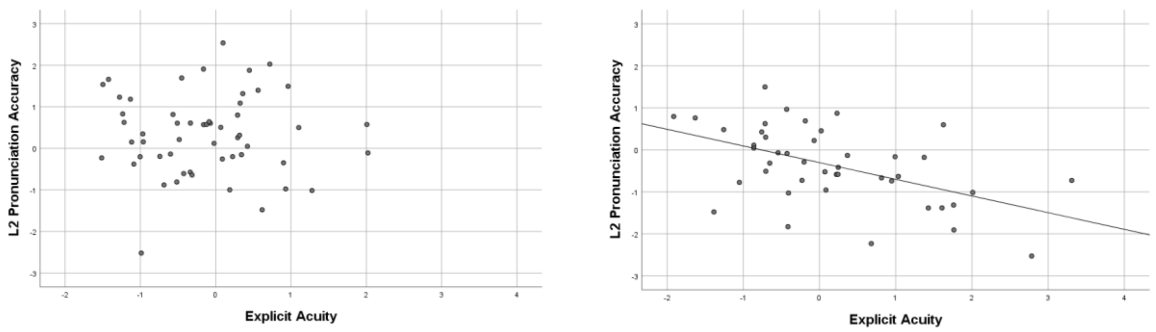

Figure 2

Relationship Between Accuracy and Explicit Acuity Among Early and Late Adulthood Arrivals

Fluency. According to the results of model comparisons, the most fitted model was identified (χ^2= 4.1337, p =.042), accounting for 18% of the variance in participants' L2 temporal fluency proficiency. As summarized in Table 7, the model consisted of one biographical factor (β = .330 for LOR) and one auditory processing variable (β = .254 for Temporal Reproduction). Interestingly, the interaction factor of LOR and AOA appeared to be the strongest predictor in the model (β = -.485). Using the AOA group distinction mentioned in the previous section (*n* = 55 for early and *n* = 45 for late adulthood arrivals), we conducted follow-up correlation analyses. As summarized in Figure 3, the L2 fluency performance of early adulthood arrivals was significantly correlated with LOR (r = .455, p < .001); but such correlation was not significant among late adulthood arrivals (r = .153, p = .316). In a nutshell, the results suggest that L2 temporal fluency is mainly tied to experience (longer immersion is better), especially when participants start immersion during early adulthood (< 22 years); auditory processing (temporal reproduction) in L2 fluency performance plays a secondary role.

Table 7

Summary of the Final Model Explaining the Perceptual and Biographical Correlates of L2 Temporal Fluency Performance

Predictors	Estimate	*SE*	*t*-value	*p*	95% *CI*	
					Lower	Upper
(Intercept)	.066	.096	0.681	.498	.146	.525
Temporal Reproduction	.209	.101	2.075	.040*	.015	.404
LOR	.331	.019	3.46	<.001*	.028	.102
LOR × AOA	-.215	.106	-2.026	.046*	-.420	-.010
Random effect (intercepts)	Variance	*SD*				
Experience Group	< .001	< .001				
Information criterion	Estimate					
LogLikelihood	-128.62					
DIC	257.23					
AIC	269.23					
BIC	284.74					
R^2						
Marginal	.23					
Conditional	.18					

A. Early Adulthood Arrivals B. Late Adulthood Arrivals

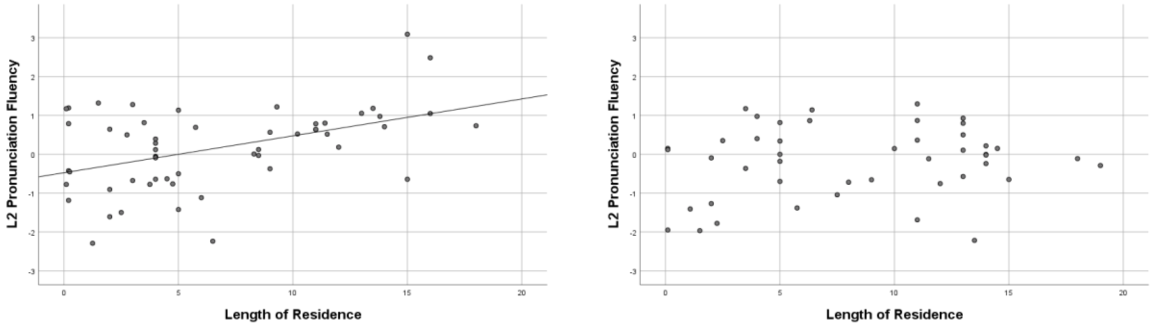

Figure 3

Relationship Between Fluency and Length of Residence Among Early and Late Adulthood Arrivals

Traits. Although all the possible combinations of main and interaction effects were tested, none of the models or predictors reached statistical significance (*p*> .05). Here, the results failed to support the link between age, experience, auditory processing, and the trait aspects of L2 speech performance. As suggested in the literature (e.g., Zuniga & Simard, 2019), both clause-final and repair ratio may be phenomena that are unrelated to cognitive and biographical backgrounds of L2 learning; these linguistic measures could mirror speakers' traits tied to individual differences in L1 rather than L2 behaviours.

Summary of Results. With respect to the perceptual and biographical correlates of L2 pronunciation proficiency, four overall patterns were found—(a) those who achieved greater L2 phonological *accuracy* had earlier age of arrival, longer immersion experience, and more precise auditory processing abilities; (b) those who achieved greater L2 phonological *fluency* had a longer length of residence in the UK and more robust auditory-motor integration abilities; (c) whereas early adulthood arrivals generally demonstrated more accurate and fluent L2 speech, variability in L2 speech production was greater in late adulthood arrivals, and more strongly related to explicit auditory perception skills; and (d) the frequency of speech repair (repetition, self-correction) may not be related to L2-related experience nor domain-general auditory perception factors.

Discussion

In the cognitive psychology literature, scholars have allocated an essential role to domain-general perception ability (i.e., auditory processing) in L1 development, L1 delay, and aging. The primary objective of the current study is to extend an emerging line of research which has hypothesized that the same faculty acts as a cornerstone of various dimensions of L2 speech learning in adulthood. Whereas the precursor work demonstrated that auditory processing facilitates the relatively early state of L2 speech *perception* proficiency (e.g., Kachlicka et al., 2019), we further examined the complex relationship between age, experience, auditory processing, and the rate and ultimate attainment of L2 speech *production* proficiency for 100 late Polish users of English in the UK.

According to the results of factor analyses, auditory processing comprised three broad components—encoding temporal and spectral details of sounds (explicit acuity), subcortical and preattentional representation of formants and fundamental frequency (pre-conscious encoding), and integrating temporal characteristics of sound for motor action (temporal reproduction). Participants' spontaneous L2 pronunciation proficiency, assessed via rater judgements and acoustic analyses, was broadly characterized as their ability to pronounce sounds, words, and sentences correctly (segmental and prosodic accuracy), and to deliver speech at an optimal speed (temporal fluency).

With respect to the biographical correlates of auditory processing (RQ1), our findings regarding late bilinguals concurred with those of L1 acquirers. We identified a small-to-medium inverse relationship between neural, pre-conscious encoding of sound and chronological age (R^2 = .15) and a positive medium relationship between temporal reproduction and various biographical variables on age, L2 experience, and music training (R^2 = .14). Similar to the L1 acquisition literature (Skoe et al., 2015), our results suggest that pre-conscious encoding ability declines with age in our dataset, n = 100 late Polish-English bilinguals whose chronological age varied between 19-45 years. Comparatively, temporal reproduction could be dynamically shaped not only by age (Thomson et al., 2015), but also by the breadth and depth of bilingual experience (Roncaglia-Denissen et al., 2016) and music experience (Tierney & Kraus, 2014). It is worth noting, however, that the relationship between music experience and temporal reproduction does not necessarily reflect a causal influence of musical training on auditory-motor skills. It is equally possible that individuals with better auditory skills are more likely to begin or to maintain musical training (Corrigall et al., 2013).

As for the relative weights of age, experience, and auditory processing in L2 speech learning (RQ2), the findings led us to speculate that the mechanisms of post-pubertal L2 speech learning may vary according to two different dimensions of L2 speech acquisition (fluency vs. accuracy) and age of acquisition (early vs. late adulthood arrivals). On the one hand, L2 temporal fluency was mainly associated with experience (length of residence), and to auditory processing to some degree (temporal reproduction). The findings indicated that many late L2 learners continue to enhance the temporal fluency aspects of L2 pronunciation proficiency, as long as they regularly use a target language over an extensive period of immersion (0.1-19 years in the current study).

This is in line with previous literature, wherein L2 fluency has been found to be susceptible to perceptible, rapid and continuous improvement as a function of increased length of residence in L2 speaking environments (e.g., Lahmann et al., 2017). The interaction between LOR and AOA showed that length of residence was more strongly related to L2 fluency among early adulthood arrivals. This could in part be because earlier AOA typically leads to longer immersion experience, which many studies (including the current investigation) have identified as a key determiner of successful L2 fluency development over time (Saito, 2015a; Trofimovich & Baker, 2006). Phonological memory, processing and reproduction ability may explain some additional variance in L2 fluency development (O’Brien et al., 2007).

On the other hand, L2 segmental and prosodic accuracy was *equally* tied to a range of factors related to age (age of arrival), experience (length of residence) and auditory processing (explicit acuity, pre-conscious encoding, temporal reproduction). To enhance L2 segmental and prosodic accuracy, which is characterized as a slow, gradual and extensive learning process due to inherent learning difficulty (Saito, 2015b; Trofimovich & Baker, 2006), and subject to individual differences to a great degree (Granena & Long, 2013), experience may be a necessary but not a sufficient condition. To efficiently and effectively acquire the relatively difficult aspects of L2 pronunciation proficiency (i.e., accuracy), our data suggest that the precision of auditory processing may serve as a key determiner of learning success.

The findings here (clearer audition effects in accuracy than fluency) concur with the developmental account of L2 aptitude and acquisition (Skehan, 2019). Under this view, aptitude is tied to high-level L2 proficiency attainment, because it facilitates the acquisition of relatively subtle, non-salient, and complex L2 features which many learners have difficulty acquiring (see also Doughty, 2019). Here, we argue that one form of aptitude, auditory processing, helps optimize L2 speech learning experience because individuals with greater auditory sensitivity can better represent, maintain, reproduce and acquire new sounds on segmental (Lengeris & Hazan, 2010) and suprasegmental levels (Chandrasekaran et al., 2011). On the whole, whereas all adult L2 learners can continue to become fluent as a function of increased bilingual experience, it may be that only perceptually acute L2 learners can ultimately achieve more advanced L2 phonological skills (fluent *and* accurate).

In terms of the relative weights of the affecting factors we identified age of acquisition as the strongest predictor (β = -.402) significantly explaining variances in L2 accuracy acquisition, compared to length of residence (β = .271) and auditory perception (β = .206-.225). The results suggest that those arriving in an L2 speaking environment during early adulthood (AOA = 17-22 years) may achieve better L2 accuracy performance through naturalistic immersion, arguably because earlier AOA allows participants to access more precise and flexible auditory processing skills (see also Saito, 2015b). As shown in the L1 literature (Roncaglia-Denissenet al., 2010; Skoe et al., 2015) and the current dataset (R^2 = .15-.24), both pre-conscious encoding and temporal reproduction aspects of auditory perception are negatively linked to chronological age from early through mid-adulthood.

Interestingly, significant interaction effects with participants' age of arrival indicate that although late adulthood arrivals (AOA = 23-36 years) may not benefit from such an age advantage, some may still demonstrate a high level of L2 proficiency when they have more precise explicit auditory processing ability as a compensatory strategy. This relationship could partially reflect the influence of cognitive factors, as explicit auditory processing of this kind, assessed via a behavioural A×B discrimination task, draws upon diverse higher-level executive functions, some of which are resistant to change and susceptible to individual differences beyond perceptual and cognitive aging (Park & Festini, 2017 for overviews on aging and executive functions; see Verhaeghen, 2011 for a meta-analysis).

Conclusion and Future Directions

Extending the cognitive psychology literature, where auditory processing has been extensively researched and suggested as a foundation of first language acquisition in childhood (i.e., auditory deficit hypothesis; Goswami, 2015), this study provides additional support to the emerging theoretical view that the same faculty underpins post-pubertal L2 speech acquisition (Mueller et al., 2012). Whereas the previous work evidenced the role of auditory processing in relatively inexperienced L2 learners' speech perception proficiency (e.g., Kachlicka et al., 2019; Omote et al., 2017), the current investigation has further extended the hypothesis that auditory processing is a crucial determinant of multiple dimensions (perception, production) and phases (rate of learning, ultimate attainment) of L2 speech learning. All in all, our conclusion is that auditory processing, operationalized as pre-conscious encoding of sound (Coffey et al., 2016), explicit discrimination of sound characteristics (Surprenant & Watson, 2001), and temporal reproduction ability (Tierney & Kraus, 2014), is a root of language acquisition throughout the lifespan, partially accounting for variability in both L1 and L2 acquisition.

Whereas many adult L2 learners appear to continue to enhance the temporal fluency aspects of L2 pronunciation proficiency as a function of increased immersion experience, the segmental and prosodic accuracy aspects of L2 pronunciation proficiency seem to be accounted for not only by factors related to biographical backgrounds (age of arrival, length of residence), but also by different types of auditory perception abilities (explicit acuity, pre-conscious encoding, temporal reproduction). Here, we argue that the effects of audition, age and experience may be somewhat inter-related. That is, individual differences in auditory perception skills (pre-conscious neural encoding of sound, temporal reproduction) are negatively associated with chronological age, as suggested in the previous literature (Skoe et al., 2013) and found in the current dataset (19-45 years). Explicit perceptual abilities (explicit acuity, reproduction) are positively influenced by the quantity and quality of L2 experience (Roncaglia-Denissen et al., 2016). If adult L2 learners start their naturalistic immersion in an L2 speaking environment at an earlier age, they can take advantage of more precise auditory neural encoding and integration, which in turn helps make the most of every input opportunity throughout extensive immersion experience, and then leads to more advanced L2 phonological skills (fluency *and* accuracy) in the long run (Saito, 2015b).

Overall, the findings of the current study agree with the primary principle of the L2 speech learning model that the mechanisms used in L1 speech acquisition are free of maturational changes, and germane to post-pubertal L2 speech acquisition (Flege, 2016; McAllister et al., 2002). To further explain in depth how the mechanisms function, the current study provides tentative suggestions on the following logical sequence. Similar to L1 acquisition, adult L2 learners may initially rely on domain-general auditory processing skills to analyze temporal and spectral properties of sounds, and then use a combination of these cues to form phonetic and phonological categories (i.e., auditory to speech category learning) (McAllister et al., 2002). Once existing categories are modified or new categories are formed, changes occur in perception, and translate to production dimensions (i.e., perception to production transfer) (Sakai & Moorman, 2017). Whereas the fundamental architecture of L1 and L2 speech acquisition is comparable (see also Saito, 2015a, 2015b), it could be biographical backgrounds that make the outcomes of the two events distinguishable. Different from L1 acquisition which is based on *blank* state, older L2 learners' speech performance could be more foreign-accented, because for every received L2 input, their processing of basic auditory information could be more susceptible to the influence of perceptual strategies appropriate for the L1 (Jasmin et al., 2020), and the degree of auditory processing precision slowly but gradually declines throughout the lifespan (Skoe et al., 2015).

What remains open to further investigation concerns the mechanisms underlying certain late adulthood arrivals who started immersion after their mid-20s (23-37 years). According to our findings, some individuals seem to overcome their relatively limited amount of perceptual resources, demonstrating advanced L2 pronunciation proficiency, thanks to their greater explicit auditory acuity, which we measured via an A×B discrimination task. We make a strong call for future research which will further explore the relationship between explicit acuity and the L2 speech acquisition of late adulthood arrivals. It is important to remember that the factor analyses provided evidence for the independence between explicit acuity (measured via the discrimination task) and pre-conscious encoding (measured via FFR) among late Polish-English bilinguals. Given that FFR is an index of preconscious sensitivity to sounds and thus closely approximates one's auditory processing without being confounded with other executive functions (Coffey et al., 2016), more research is needed to further scrutinize precisely what kinds of perceptual *and* cognitive abilities the explicit acuity task actually taps into (Snowling et al., 2018). Turning to the L2 speech literature, late L2 learners with a high level of L2 segmental proficiency likely have greater working memory, attention, and processing speed (Darcy et al., 2015), inhibitory control (Darcy et al., 2016), and phonological short-term memory (Silbert et al., 2015). In conjunction with similar research designs in the context of L1 acquisition (Snowling et al., 2018), it would be intriguing to probe both perceptual profiles (explicit acuity, pre-conscious encoding, and reproduction) and cognitive profiles (working memory, attention, and control) of post-pubertal L2 speakers' speech acquisition over the course of immersion from a longitudinal perspective (cf. Saito et al., 2020b; Sun et al., forthcoming).

In a similar vein, future longitudinal research can explore a *causal* link between different types of auditory processing and post-pubertal L2 speech acquisition by designing an intervention study with a pre-and-post-test design. There is evidence that providing specific auditory training leads to improvement in auditory processing abilities (see Hayeset al., 2003 for 35-40 hours of commercial auditory processing training programs; Carcagno & Plack, 2011 for 10 hours of pitch discrimination training; Strehlow et al., 2006 for four weeks of temporal processing training). The impact of such training gains appears to be robust (see Whiteford & Oxenham, 2018 for both behavioural and neurophysiological evidence for the maintenance of training gains for a one-year period). In terms of the transferability of auditory training to language-specific skills (reading and/or spelling abilities), however, research findings have been inconclusive (see McArthur et al., 2008). Interestingly, the transfer effect was observed when participants received both auditory and some form of language (e.g., reading) training at the same time. To test the role of auditory processing in post-pubertal L2 speech acquisition, therefore, future researchers are recommended to take a longitudinal look at whether and to what degree adding auditory training can further increase the acquisitional value of L2 speech instruction that previous research has already proven to be effective (see Barriuso, & Hayes-Harb, 2018 for high-variability phonetic training; Saito & Lyster, 2012 for focus-on-form; Mora & Levkina, 2017 for task-based pronunciation teaching).

Author Credit Statement

Kazuya Saito (Corresponding Author): Conceptualization, Methodology, Writing – Original Draft, Editing, & Revision, Validation, Formal Analysis, Investigation, Supervision, Project administration, Funding acquisition

Magdalena Kachlicka: Conceptualization, Formal Analysis, Validation, Data collection

Hui Sun: Data collection

Adam Tierney: Conceptualization, Methodology, Writing - Editing, Investigation, Supervision, Funding acquisition

References

Abrahamsson, N. & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning: A Journal of Research in Language Studies*, *59*(2), 249–306. https://doi.org/10.1111/j.1467-9922.2009.00507.x

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, *245*(1-2), 35–47. https://doi.org/10.1016/j.heares.2008.08.004

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge University Press.

Barriuso, T. A., & Hayes-Harb, R. (2018). High Variability Phonetic Training as a Bridge from Research to Practice. *CATESOL Journal*, *30*(1), 177–194.

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

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. Commonalities and complementaries. In: O. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning. In honor of James Emil Flege* (pp. 13-34). John Benjamins.

Birdsong, D., & Molis, M. (2001). On the Evidence for maturational constraints in second language acquisition. *Journal of Memory and Language*, *44*(2), 235–249. https://psycnet.apa.org/doi/10.1006/jmla.2000.2750

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

Boets, B., Wouters, J., Van Wieringen, A., De Smedt, B., &Ghesquiere, P. (2008). Modelling relations between sensory processing, speech perception, orthographic and phonological ability, and literacy achievement. *Brain and Language*, *106*(1), 29–40. https://doi.org/10.1016/j.bandl.2007.12.004

Bongaerts, T., Van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, *19*(4), 447–465. https://doi.org/10.1017/S0272263197004026

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*(2), 159–175. https://doi.org/10.1177%2F0265532212455394

Campbell, K. L., & Tyler, L. K. (2018). Language-related domain-specific and domain-general systems in the human brain. *Current Opinion in Behavioral Sciences*, *21*, 132–137. https://doi.org/10.1016/j.cobeha.2018.04.008

Carcagno, S., & Plack, C. J. (2011). Pitch discrimination learning: specificity for pitch and harmonic resolvability, and electrophysiological correlates. *Journal of the Association for Research in Otolaryngology*, *12*(4), 503. https://doi.org/10.1007/s10162-011-0266-3

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

Casini, L., Pech-Georgel, C., & Ziegler, J. C. (2018). It's about time: revisiting temporal processing deficits in dyslexia. *Developmental Science*, *21*(2), e12530. https://doi.org/10.1111/desc.12530

Chandrasekaran, B., Kraus, N., & Wong, P. C. (2011). Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology*, *107*(5), 1325–1336. https://doi.org/10.1152/jn.00923.2011

Clinard, C. G., Tremblay, K. L., & Krishnan, A. R. (2010). Aging alters the perception and physiological representation of frequency: evidence from human frequency-following response recordings. *Hearing Research*, *264*, 48–55. https://doi.org/10.1016/j.heares.2009.11.010

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*, 11070. https://doi.org/10.1038/ncomms11070

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in Psychology*, *4*, 222. https://doi.org/10.3389/fpsyg.2013.00222

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

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. https://doi.org/10.1016/j.lindif.2015.04.005

DeKeyser, R. M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning*, *63,* 52–67. https://doi.org/10.1111/j.1467-9922.2012.00737.x

Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*(2), 163–185. https://doi.org/10.1111/lang.12000

Doughty, C. J. (2019). Cognitive language aptitude. *Language Learning, 69*, 101–126. https://doi.org/10.1111/lang.12322

Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., &Alwan, A. (2000). Acoustic modeling of American English /r/. *The Journal of the Acoustical Society of America*, *108*(1), 343–356. https://doi.org/10.1121/1.429469

Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: a randomized control trial. *PLoS ONE*, *10*(9), e0138715. https://doi.org/10.1371/journal.pone.0138715

Flaugnacco, E., Lopez, L., Terribili, C., Zoia, S., Buda, S., Tilli, S., & Schön, D. (2014). Rhythm perception and production predict reading abilities in developmental dyslexia. *Frontiers in Human Neuroscience*, *8*, 392. https://doi.org/10.3389/fnhum.2014.00392

Flege, J. (2016, June). *The role of phonetic category formation in second language speech acquisition*. Plenary address delivered at New Sounds 2016, 8th International Conference on Second-Language Speech, Aarhus, Denmark.

Flege, J. E., Munro, M, & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America, 97*(5), 3125–3134. https://doi.org/10.1121/1.413041

Flege, J., Takagi, N., & Mann, V. (1995). Japanese adults learn to produce English /ɹ/ and /l/ accurately. *Language and Speech, 38*(1), 25–55. https://doi.org/10.1177%2F002383099503800102

Gibson, L. Y., Hogben, J. H., & Fletcher, J. (2006). Visual and auditory processing and component reading skills in developmental dyslexia. *Cognitive Neuropsychology*, *23*(4), 621–642. https://doi.org/10.1080/02643290500412545

Gokula, R., Sharma, M., Cupples, L., & Valderrama, J. T. (2019). Comorbidity of auditory processing, attention, and memory in children with word reading difficulties. *Frontiers in Psychology*, *10,* 2383. https://doi.org/10.3389/fpsyg.2019.02383

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

Goswami, U., Wang, H. L. S., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2011). Language-universal sensory deficits in developmental dyslexia: English, Spanish, and Chinese. *Journal of Cognitive Neuroscience*, *23*(2), 325–337. https://psycnet.apa.org/doi/10.1162/jocn.2010.21453

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

Halliday, L. F., Tuomainen, O., & Rosen, S. (2017). Auditory processing deficits are sometimes necessary and sometimes sufficient for language difficulties in children: Evidence from mild to moderate sensorineural hearing loss. *Cognition*, *166,* 139–151. https://doi.org/10.1016/j.cognition.2017.04.014

Hamrick, P., Lum, J. A., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences*, *115*(7), 1487–1492. https://doi.org/10.1073/pnas.1713975115

Hayes, E. A., Warrier, C. M., Nicol, T. G., Zecker, S. G., & Kraus, N. (2003). Neural plasticity following auditory training in children with learning problems. *Clinical Neurophysiology*, *114*(4), 673–684. https://doi.org/10.1016/S1388-2457(02)00414-5

Hornickel, J., & Kraus, N. (2013). Unstable representation of sound: a biological marker of dyslexia. *Journal of Neuroscience*, *33*(8), 3500–3504. https://doi.org/10.1523/JNEUROSCI.4205-12.2013

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

Jasmin, K, Sun, H., Tierney, A. (2020). Effects of language experience on domain-general perceptual strategies. *BioRxiv*. https://doi.org/10.1101/2020.01.02.892943

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

Kempe, V., Thoresen, J. C., Kirk, N. W., Schaeffler, F., & Brooks, P. J. (2012). Individual differences in the discrimination of novel speech sounds: Effects of sex, temporal processing, musical and cognitive abilities. *PLoS ONE, 7*(11), e48623. https://doi.org/10.1371/journal.pone.0048623

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

Lahmann, C., Steinkrauss, R., & Schmid, M. S. (2017). Speed, breakdown, and repair: An investigation of fluency in long-term second-language speakers of English. *International Journal of Bilingualism*, *21*(2), 228–242. https://doi.org/10.1177%2F1367006915613162

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, *39*(1), 167–196. https://doi.org/10.1017/S0272263116000085

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics, 36*(3), 345–366. https://doi.org/10.1093/applin/amu040

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

Levitt, H. C. C. H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467–477. https://doi.org/10.1121/1.1912375

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

McAllister, R., Flege, J., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics, 30*(2), 229–258. https://doi.org/10.1006/jpho.2002.0174

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

McArthur, G. M., Ellis, D., Atkinson, C. M., & Coltheart, M. (2008). Auditory processing deficits in children with reading and language impairments: Can they (and should they) be treated? *Cognition*, *107*(3), 946–977. https://doi.org/10.1016/j.cognition.2007.12.005

Mora, J. C., & Levkina, M. (2017). Task-based pronunciation teaching and research: Key issues and future directions. *Studies in Second Language Acquisition*, *39*(2), 381–399. https://doi.org/10.1017/S0272263117000183

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

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

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395. https://psycnet.apa.org/doi/10.1037/0033-295X.115.2.357

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, *29*(4), 557–581. https://doi.org/10.1017/S027226310707043X

Omote, A., Jasmin, K., & Tierney, A. (2017). Successful non-native speech perception is linked to frequency following response phase consistency. *Cortex*, *93*, 146–154. https://doi.org/10.1016/j.cortex.2017.05.005

Park, D. C., & Festini, S. B. (2017). Theories of memory and aging: A look at the past and a glimpse of the future. *The Journals of Gerontology: Series B*, *72*(1), 82–90. https://doi.org/10.1093/geronb/gbw066

Piske, T., Flege, J., MacKay, & Meador, D. (2011). Investigating native and non-native vowels produced in converational 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). Peter Lang.

Plag, I., Kunter, G., & Schramm, M. (2011). Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics*, *39*(3), 362–374. https://doi.org/10.1016/j.wocn.2011.03.004

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception: An Interdisciplinary Journal*, *2*(4), 411–440. https://doi.org/10.2307/40285311

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012. https://www.R-project.org

Roncaglia-Denissen, M. P., Roor, D. A., Chen, A., &Sadakata, M. (2016). The enhanced musical rhythmic perception in second language learners. *Frontiers in Human Neuroscience*, *10*, 288. https://doi.org/10.3389/fnhum.2016.00288

Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything? *Journal of Phonetics*, *31*(3-4), 509–527. https://doi.org/10.1016/S0095-4470(03)00046-9

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

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*(8), 1720–1731. https://doi.org/10.1016/j.clinph.2008.01.108

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187–224. https://doi.org/10.1017/S0142716417000418

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*(4), 546–562. https://doi.org/10.1016/j.jml.2013.07.003

Saito, K. (2015a). Experience effects on the development of late second language learners' oral proficiency. *Language Learning*, *65*, 563–595. https://doi.org/10.1111/lang.12120

Saito, K. (2015b). The role of age of acquisition in late second language oral proficiency attainment. *Studies in Second Language Acquisition, 37*, 713-743. https://doi.org/10.1017/S0272263115000248

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

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*(3), 652-708. https://doi.org/10.1111/lang.12345

Saito, K., Sun, H., & 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*(5), 1123–1140. https://doi.org/10.1017/S1366728918000895

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

Saito, K., Sun, H., & Tierney, A. (2020b). Domain-general auditory processing determines success in L2 pronunciation learning in adulthood: A longitudinal study. *Applied Psycholinguistics.*

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

Silbert, N. H., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., & Tare, M. (2015). Non-native phonemic discrimination, phonological short term memory, and word learning. *Journal of Phonetics*, *50*, 99–119. https://doi.org/10.1016/j.wocn.2015.03.001

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

Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear and Hearing*, *31*(3), 302–324. https://doi.org/10.1097/AUD.0b013e3181cdb272

Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2015). Stability and plasticity of auditory brainstem function across the lifespan. *Cerebral Cortex*, *25*(6), 1415–1426. https://doi.org/10.1093/cercor/bht311

Smith, J. O. (2007). *Introduction to digital filters with audio applications*. Retrieved from: http://ccrma.stanford.edu/~jos/filters/

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

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

Strehlow, U., Haffner, J., Bischof, J., Gratzka, V., Parzer, P., & Resch, F. (2006). Does successful training of temporal processing of sound and phoneme stimuli improve reading and spelling? *European Child & Adolescent Psychiatry*, *15*(1), 19–29. https://doi.org/10.1007/s00787-006-0500-4

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

Surprenant, A. M., & Watson, C. S. (2001). Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *The Journal of the Acoustical Society of America*, *110*(4), 2085–2095. https://doi.org/10.1121/1.1404973

Tallal, P. (2004). Improving language and literacy is a matter of time. *Nature Reviews Neuroscience*, *5*(9), 721–728. https://doi.org/10.1038/nrn1499

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In: R. Ellis (Ed). *Planning and Task Performance in a Second Language* (pp. 239–273). John Benjamins Publishing Company.

Thompson, E. C., White-Schwoch, T., Tierney, A., & Kraus, N. (2015). Beat synchronization across the lifespan: Intersection of development and musical experience. *PLoS ONE*, *10*(6), e0128839. https://doi.org/10.1371/journal.pone.0128839

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*(38), 14981–14988. https://doi.org/10.1523/JNEUROSCI.0612-13.2013

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

Tierney, A. T., Krizman, J., & Kraus, N. (2015). Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences*, *112*(32), 10062–10067. https://doi.org/10.1073/pnas.1505114112

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*(5), 855–868. https://doi.org/10.1162/jocn_a_01092

Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, *11*, 61–82. https://doi.org/10.1515/cogl.2001.012

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

Trofimovich, P., Kennedy, S., & Foote, J. A. (2015). Variables affecting L2 pronunciation development. In: M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 353–373). Wiley.

Varghese, L., Bharadwaj, H. M., & Shinn-Cunningham, B. G. (2015). Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. *Brain Research*, *1626*, 146–164. https://doi.org/10.1016/j.brainres.2015.06.038

Verhaeghen, P. (2011). Aging and executive control: Reports of a demise greatly exaggerated. *Current Directions in Psychological Science*, *20*(3), 174–180. https://doi.org/10.1177/0963721411408772

Wagenmakers, E. J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*(3), 830–841. https://doi.org/10.1037/0033-295X.114.3.830

White-Schwoch, T., Woodruff Carr, K., Thompson, E. C., Anderson, S., Nicol, T., Bradlow, A. R., Zecker, S. G., Kraus, N. (2015). Auditory processing in noise: A preschool biomarker for literacy. *PLoS Biology*, *13*(7), 1–17. https://doi.org/10.1371/journal.pbio.1002196

Whiteford, K. L., &Oxenham, A. J. (2018). Learning for pitch and melody discrimination in congenital amusia. *Cortex*, *103*, 164–178. https://doi.org/10.1016/j.cortex.2018.03.012

Won, J. H., Tremblay, K., Clinard, C. G., Wright, R. A., Sagi, E., & Svirsky, M. (2016). The neural encoding of formant frequencies contributing to vowel identification in normal-hearing listeners. *The Journal of the Acoustical Society of America*, *139*(1), 1–11. https://doi.org/10.1121/1.4931909

Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(4), 565–585. https://doi.org/10.1017/S0142716407070312

Wright, B. A., Bowen, R. W., & Zecker, S. G. (2000). Nonlinguistic perceptual deficits associated with reading and language disorders. *Current Opinion in Neurobiology*, *10*, 482–486. https://doi.org/10.1016/S0959-4388(00)00119-7

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*(1), 384–395. https://doi.org/10.1121/1.4807498

Zuniga, M., & Simard, D. (2019). Factors influencing L2 self-repair behavior: The role of L2 proficiency, attentional control and L1 self-repair behavior. *Journal of Psycholinguistic Research*, *48*(1), 43–59. https://doi.org/10.1007/s10936-018-9587-2