Title: Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound

Authors: Magdalena Kachlicka[a,c], Kazuya Saito[b], Adam Tierney[a†]

Affiliations

[a]Department of Psychological Sciences
Birkbeck, University of London
Malet Street, London, WC1E 7HX, United Kingdom

[b]Institute of Education
University College London
Gower Street, London, WC1E 6BT, United Kingdom

[c]Institute for Environmental Design and Engineering
University College London
Gower Street, London, WC1E 6BT, United Kingdom

[†]Corresponding author
Adam Tierney, Ph.D.
Birkbeck, University of London
Malet Street, London, WC1E 7HX
United Kingdom
Email: a.tierney@bbk.ac.uk
Phone: +44-020-7631-6368

Declarations of interest: none

Abstract

There is a great deal of individual variability in outcome in second language learning, the sources of which are still poorly understood. We hypothesized that individual differences in auditory processing may account for some variability in second language learning. We tested this hypothesis by examining psychoacoustic thresholds, auditory-motor temporal integration, and auditory neural encoding in adult native Polish speakers living in the UK. We found that precise English vowel perception and accurate English grammatical judgment were linked to lower psychoacoustic thresholds, better auditory-motor integration, and more consistent frequency-following responses to sound. Psychoacoustic thresholds and neural sound encoding explained independent variance in vowel perception, suggesting that they are dissociable indexes of sound processing. These results suggest that individual differences in second language acquisition success stem at least in part from domain-general difficulties with auditory perception, and that auditory training could help facilitate language learning in some individuals with specific auditory impairments.

Keywords: Auditory; bilingualism; FFR; language; rhythm

1. Introduction

In a globalized world, a growing number of people are moving to a new country and attempting to learn a second language (L2) in adulthood. However, L2 acquisition is characterized by large individual differences, with some people achieving near-native performance with ease while others produce heavily accented speech, struggle to comprehend speech, read at a rudimentary level, and display less grammatical and lexical knowledge (Li, 2016). Understanding the underlying mechanisms of L2 learning could lead to remedial approaches designed to boost L2 skills in struggling learners (DeKeyser, 2012).

Prior research on individual differences in L2 learning has found that greater success is linked to a variety of characteristics of the language input received. L2 learners are more successful, for example, when they are immersed in an L2 environment at an early age (Flege, Yeni-Komshian, & Liu, 1999; Birdsong & Molis, 2001; Abrahamsson & Hyltenstam, 2009), have been resident in an L2 environment for a greater amount of time (Flege, Bohn, & Jang, 1997; Trofimovich & Baker, 2006), and use their L2 more often on a daily basis (Flege & Liu, 2001; Derwing & Munro, 2013). Nonetheless, even after accounting for these characteristics of the input a substantial amount of variability in L2 learning success remains unaccounted for. This suggests that other factors are at play, some of which may be biological in nature, and may predispose some individuals to make better use of every learning opportunity, develop their second language system more efficiently and effectively, and attain more advanced proficiency in the long run (Foster, Bolibaugh, & Kotula, 2014).

The auditory channel is the primary source of language input for most people. Learning a language, therefore, requires complex auditory analysis: patterns of timing, pitch, and spectral shape must be tracked across multiple timescales. For example, listeners must be able to discriminate spectral and temporal patterns in order to distinguish between speech sounds and build an inventory of a language's component sounds. Listeners must also be able to track patterns of pitch and duration in order to extract prosodic features such as accent, stress, and phrase boundaries (de Pijper & Sanderman, 1994; Fear, Cutler, & Butterfield, 1995; Turk & White, 1999). These prosodic features highlight relevant portions of the discourse (Wang, Li, & Yang, 2014) and provide cues to word boundaries (Cutler & Butterfield, 1992) and grammatical structure (Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992). There are, therefore, links between auditory patterns and linguistic structure at every level, including phonetic, prosodic, lexical, and grammatical features.

One possible source of difficulties with L2 learning, therefore, may be impairments in the perception of auditory patterns. These may impede or delay the acquisition of phonological, semantic, and syntactic knowledge. The hypothesis that auditory processing may be a bottleneck for second language learning is supported by short-term training studies which have assessed auditory perception prior to asking participants to briefly learn a speech sound contrast from an unfamiliar language. These studies have found that rapid speech sound learning is linked to behavioural measures of auditory processing such as spectral (Wong & Perrachione, 2007; Lengeris & Hazan, 2010) and temporal (Kempe, Thoresen, Kirk, Schaeffler, & Brooks, 2012; Kempe, Bublitz, & Brooks, 2015) discrimination acuity, as well as greater white matter density and volume in left Heschl's gyrus (Golestani, Molko, Dehaene, LeBihan, & Pallier, 2007; Golestani & Pallier, 2007) and the robustness of neural encoding of speech (Chandrasekaran, Kraus, & Wong, 2011). Recently, we have shown that the robustness of neural encoding of speech is linked to English speech perception ability in native Japanese speakers (Omote, Jasmin, & Tierney, 2017) and to English speech production ability in native Mandarin Chinese speakers (Saito, Sun, & Tierney, 2018) living in the UK, suggesting that individual differences in auditory encoding may relate to L2 learning outside of the laboratory as well.

Prior investigation of relationships between auditory processing and L2 learning have focused entirely on speech perception and production. However, as mentioned above, detection of auditory cues can facilitate perception of prosodic features (Goswami et al., 2013), which communicate information about language structure at multiple levels, including syntax, pragmatics, and semantics. Indeed, research on first language acquisition in children has suggested that auditory processing may be related to a wide range of language skills. For example, children who struggle to acquire language skills such as the perception and use of syntax are also more likely to display a variety of auditory processing difficulties, including impaired frequency discrimination (McArthur & Bishop, 2004), duration discrimination, and amplitude rise-time discrimination (Richards & Goswami, 2015), worse discrimination of musical rhythms (Gordon et al., 2015), more variable synchronization to a metronome (Corriveau & Goswami, 2009), and diminished robustness of auditory brainstem responses to sound (Basu, Krishnan, & Weber-Fox, 2010). These findings suggest that difficulties with auditory processing may have consequences for second language acquisition that extend beyond speech perception, potentially also affecting the acquisition of linguistic knowledge at other levels, including syntax. However, to our knowledge this has not previously been investigated.

Here, we investigated the extent to which behavioral measurements of the precision of auditory processing and neural measurements of the robustness of sound encoding were linked to individual differences in the English-language skills of native Polish speakers living in London. Psychoacoustics, musical rhythm perception, and synchronization tests were used to measure auditory processing behaviourally. The frequency-following response (FFR), an electrophysiological response which reproduces the frequencies present in the evoking sound and reflects activity in the brainstem and cerebral cortex (Coffey, Herholz, Chepesiuk, Baillet, & Zatorre, 2016), was used as a measurement of the neural encoding of sound. On the one hand, we predicted that English vowel perception would be linked to measures of spectral processing, given that this is the primary cue distinguishing English vowels (Peterson & Barney, 1952); these measures included frequency and formant discrimination as well as inter-trial phase locking in the neural representation of high-frequency speech formants. On the other hand, we predicted that English sentence grammaticality judgments would be linked to both spectral and temporal processing, as both of these dimensions are relevant to the perception of speech prosody. Moreover, given prior work showing that the consistency of the frequency-following response is linked to language skills in children (Hornickel & Kraus, 2013), we predicted that grammatical knowledge would also be linked to inter-trial phase-locking of the neural representation of the fundamental frequency of speech.

Finally, we investigated the extent to which variability in neural encoding of sound explained individual differences in auditory perceptual ability. Prior work has revealed links between the robustness of the frequency-following response and a variety of auditory skills, including the precision of auditory-motor synchronization (Tierney & Kraus, 2013, 2016; Tierney, White-Schwoch, MacLean, & Kraus, 2017), amplitude modulation detection (Purcell, John, Schneider, & Picton, 2004; Bharadwaj, Masud, Mehraei, Gerhulst, & Shinn-Cunningham, 2015), and frequency discrimination (Krishnan, Bidelman, & Gandour, 2010; Krishnan, Bidelman, Gmalt, Ananthakrishnan, & Gandour, 2012; Marmel et al., 2013). However, as each of these studies examined relationships between the frequency-following response and either a single test or a narrow range of tests designed to measure similar skills (i.e. temporal processing), the specificity of these relationships remains unclear. Here we investigated correlations between behavioural and neural measures of auditory processing, to investigate whether different aspects of the neural encoding of sound reflect different dissociable auditory processing factors. In particular, we predicted that lower-frequency phase-locking (i.e. at the fundamental frequency) would be linked to temporal processing, while higher-frequency phase-locking (at the speech formants) would be linked to spectral processing.

2. Methods

2.1 Participants

All the participants provided their written consent to participate in the study, before the beginning of the testing session. All procedures were approved by the Research Ethics Committee of the Department of Psychological Sciences at Birkbeck. Forty native speakers of Polish (29 female, age M=25.63, SD=4.76, range 19 to 39) speaking English as a second language who lived in the UK at least one year and not more than six years were recruited for this study. On average participants arrived in the UK at 21.9 (SD = 4.2, range 18 to 36) years of age, had been in the country for 3.6 (1.3, range 1.08 to 5.75) years, underwent 9.4 (4.4, range 0.5 to 20) years of in-class training in English prior to coming to the UK, and used English 65.7% of the time (19.8%, range 18.3% to 96.7%). 14 people reported having previously engaged in some musical training (M = 9.46 years, SD = 5.32, range 1 to 16). Participants were students enrolled in various undergraduate and graduate programmes or working professionals living in London. All participants reported no prior diagnosis of a hearing impairment or neurological disorder that affects hearing. One participant requested not to complete the electrophysiological battery, and so was only included in analyses examining behavioral variables.

2.2 Behavioural measures

2.2.1 Measures of experience.

Demographic data and measures of experience were collected online via a custom-made questionnaire of language experience and proficiency, which we distributed to the potential participants to establish their eligibility for the study before inviting them for testing. There is a consensus among all relevant theories that experience is a necessary condition for any dimensions of successful second language learning (e.g., Ellis, 2006). Not surprisingly, it has often been reported that little learning takes place when second language learners choose to use their first language without many opportunities to interact with native and other non-native speakers in the target language (e.g., Jia & Aaronson, 2003). In the current study, however, our main focus lied in investigating the sources of individual variability among regular, active and motivated second language users (Doughty, 2018). Our hypothesis was that the attainment of high-level second language proficiency could be tied to participants' auditory processing abilities rather than experience-related factors. To this end, a decision was made to recruit only those who used English as a main language of communication at work or home. All participants had arrived in the UK after the age of 16 years. None reported prior diagnosis of a hearing impairment.

By tailoring two questionnaire instruments designed to capture demographic variables relevant to successful second language learning in naturalistic (Language Contact Profile: Freed, Dewey, Segalowitz, & Halter, 2004) and classroom settings (Foreign Language Experience Questionnaire: Saito & Hanzawa, 2016), we surveyed the following information: age of acquisition, length of residence, length of L2 English learning in classroom settings, frequency of use of English, and whether participants had previously received any musical training. Given that the degree of success in L2 learning is claimed to relate not only to quantity but also to quality (Flege, 2016), the nature of participants' use of English was assessed as a percentage of language use, averaged across professional, social, and home settings. In addition, to gain a rough estimate of global language proficiency, participants were asked to self-assess their L2 English proficiency on a scale from 1 (heavily accented) to 9 (nativelike). Participants generally reported high proficiency (M = 6.80, SD = 1.54, range 2.5 to 9).

2.2.2 Auditory processing battery.

All participants completed a battery of psychophysical assessments measuring thresholds for discrimination of pitch, amplitude rise time, duration, and formant frequency. Stimuli were complex tones, constructed using custom MATLAB (The MathWorks, Inc., Natick, MA) scripts, and modified as necessary for each test. An adaptive three-alternative forced-choice procedure was used, modified from the transformed up-down procedure described by Levitt (1971). That is, the difficulty of the task decreased after every incorrect response and increased after every third correct response. For all tests, a continuum of 100 stimuli was created. The test presentation began at stimulus level 50 with a starting step size of 10 (i.e., the task became easier by 10 steps after an incorrect response and more difficult by 10 steps after every third correct response). After a first reversal, the step size changed to five, after a second reversal to two, and after a third reversal to one and remained at this level until the end of the test. The program stopped either after 70 trials or eight reversals. Eight reversals were reached by 31 participants for the duration test, 20 participants for the frequency test, 25 participants for the formant test, and 25 participants for the rise time test. The score was calculated as the levels of each reversal from the second onward. In each test, three tones were presented with a constant inter-stimulus interval of 0.5 s, with either the first or the third sound different from the other two. Participants' task was to indicate which sound was different by either pressing the number '1' or '3' on a keyboard.

The stimuli for the pitch, duration, and rise time discrimination tests were constructed by modifying a standard four-harmonic complex tone (equal amplitudes across harmonics) with a duration of 0.5 s, a 0.015 s linear ramp at the beginning and end, and an F0 of 330 Hz. For the pitch discrimination test, while the standard stimulus was always presented at a fundamental frequency (F0) of 330 Hz, the target stimulus continuum ranged from an F0 of 330.3-360 Hz. For the duration discrimination test, while the standard stimulus always had a duration of 0.25 s, the target stimulus duration continuum ranged from 0.2525-0.5ms. For the rise time discrimination test, while the standard stimulus always had a rise time of 0.015 s, the target stimulus continuum ranged from 0.0178 to 0.3 s. For the formant discrimination test, stimuli were complex tones with an F0 of 100 Hz and harmonics up to 3000 Hz onto which three formants were imposed using a parallel formant filter bank (Smith, 2007). Stimuli were 0.5 s in duration with a 0.015 s linear ramp at the beginning and the end. The first formant (F1) was kept constant at 500 Hz and third formant (F3) at 2500 Hz. The second formant (F2) of the standard stimulus was always 1500 Hz. The target stimulus continuum ranged from an F2 of 1502-1700 Hz. To form a composite measure of *spectral processing*, pitch and formant discrimination thresholds were converted to z-scores and averaged. To form a composite measure of *temporal processing*, rise time and duration discrimination thresholds were converted to z-scores and averaged.

2.2.3 Auditory-motor temporal integration.

Two auditory-motor temporal integration tests were used to establish participants' ability to detect and reproduce temporal patterns in non-verbal auditory stimuli. The base component stimulus for both tests consisted of a recording of a 150-ms conga drum hit acquired at freesound.org. The stimuli were presented using MATLAB through Etymotic-3A audiometric insert earphones (Etymotic, Elk Grove Village, IL) at 80-dB sound pressure level (SPL). Participants were asked to drum along to the stimuli by hitting a hand drum with their dominant hand such that their drum hits occurred at the same time as the stimulus onset. Drum hits were received by a microphone, with the alignment of audio presentation and participant drumming accomplished using an RTBox (Li, 2010) and custom MATLAB scripts.

In a metronome synchronization test, participants were asked to maintain a steady beat while synchronising to an isochronous stimulus at multiple rates. In each block of the synchronisation test, participants heard 40 presentations of the drum sound, with an isochronous inter-stimulus interval. Participants were asked to synchronize as soon as they were able, but only their synchronization to the final 20 presentations was analysed, ensuring that the measurement was of variability of synchronization rather than speed of synchronization. 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). In a rhythm synchronization test, participants were asked to rapidly perceive and synchronize with a complex metrical rhythmic sequence. In each block of the rhythm synchronisation test, participants were presented with eight repetitions of a rhythm pattern 3.2 s in duration. Rhythm patterns were taken from Povel and Essens (1985) and consisted of 16 segments, each 200 ms, containing either a rest or a drum hit.

Drum hit onset times were marked offline by setting amplitude thresholds and relaxation times manually for each participant, such that any time point with amplitude exceeding the threshold was marked as a drum hit unless an amount of time less than the relaxation time had elapsed since the last hit. For the metronome synchronization test, performance was calculated as *synchronization variability*, measured using the coefficient of variation, i.e. the standard deviation of the interval between each drum hit and the closest stimulus onset, divided by the inter-onset interval, then averaged across trials. For the rhythm synchronization test, *rhythmic accuracy* was calculated by determining, for each segment of the target rhythm, whether the participant produced a rest or a drum hit in a 200 ms window centred on the onset of the segment, and then comparing the participant's response to the content of the target rhythm. The accuracy score for a given trial consisted of the number of segments produced correctly divided by the total number of segments. Scoring began at the onset of the second repetition.

2.2.4. Grammatical Judgement Test.

A Grammatical Judgment Test (GJT; Godfroid, Loewen, Jung, & Park, 2015) was used to measure participants' ability to indicate the syntactical acceptability of written sentences. Subjects were presented with 68 sentences written in the English language, 34 of which were grammatical and 34 ungrammatical. They were asked to rate their grammatical acceptability by pressing the 'u' key if a sentence was ungrammatical and a 'g' key if the sentence was grammatical. Participants were given only a few seconds to respond; this time limit varied from item to item, depending on stimulus length, following the time limits used by Godfroid et al. (2015). The grammatical forms manipulated covered a wide range of grammatical structures which second language learners of English have difficulty learning (e.g., plurals -s, possessives, indefinite articles, past tense). The final score was calculated as a sum of correct identifications of grammatical sentences and correct rejections of ungrammatical sentences. Participants were instructed to read each sentence silently, and in no case did a participant read the stimuli out loud.

2.2.5. Language aptitude test.

A Language Aptitude Test was used to assess participants' general predispositions to learn a second language. LLAMA is a set of language-neutral tests (Meara, 2005) based on the components of the Modern Language Aptitude Test (Carroll & Sapon, 1959). Being practically constrained by the amount of time available with each participant we could not administer the whole battery and so we focused on the measure of phonemic coding (LLAMA E), which has been shown to be a predictor of L2 pronunciation (e.g., Hu et al., 2013) and morphosyntax (e.g., Saito, 2017). LLAMA E is a test of the formation of sound-symbol associations measuring the ability to associate unfamiliar symbols with sounds and to dissociate sounds from the way that they are typically written in English. During the one-minute practice session, participants were asked to remember the 24 recorded syllables (consonant-vowel pairs) and their corresponding phonetic symbols. Then, their task was to identify which of the two presented spellings accurately represents the two-syllable word they heard (a total of 20 items). They had unlimited time to complete the task. The outcome measure was a portion of correct responses.

2.2.6. Speech perception test.

A speech perception test was used to assess participants' ability to perceive English speech contrasts which are known to be problematic for Polish learners of the English language due to cross-linguistic differences between English and Polish phonetic systems. Stimuli presented included vowel contrasts /æ/ versus /ʌ/ (Rojczyk, 2010) and /æ/ versus /e/ (Schwartz, Aperlinski, Jekiel, & Malarski, 2016), consonant contrasts /g/ versus /k/ and /t/ versus /d/ (Rojczyk, 2012), and lexical stress contrasts (Peperkamp, Gendelin, & Dupoux, 2010). All contrasts were presented in a word context. There were 20 tokens for each contrast. All stimuli for the speech perception test were recorded by a native speaker of Southern British English and presented using custom MATLAB programmes. Participants were asked to listen to a spoken word and then to indicate the correct spelling from the two options displayed on the screen by either pressing the number '1' or '2' on the keyboard. The outcome measure was a portion of correct responses. Performance was at ceiling on the consonant items (M=97.4%, SD = 4.2%) and at floor for the stress items (M=53%, SD=13.4%), and so only data from the vowel items was analysed further.

2.3. Electrophysiology.

2.3.1. Stimulus.

The stimulus used to evoke electrophysiological responses was the consonant-vowel syllable /da/ (170 ms in duration) synthesised with a Klatt-based synthesiser. The stimulus began with a 5 ms onset burst. Between 5 and 50 ms F1 rose from 400 to 720 Hz, F2 fell from 1700 to 1240 Hz, and F3 fell from 2580 to 2500 Hz. Between 50 and 170 F1, F2, and F3 were stable at 720 Hz, 1240 Hz, and

2500 Hz, respectively. F4, F5, and F6 were constant between 5 and 170 ms at 3300 Hz, 3750 Hz, and 4900 Hz, respectively. The F0 was constant throughout the stimulus at 100 Hz.

2.3.2. Data collection.

The stimulus was presented diotically through Etymotic 3A insert earphones at 80 dB SPL. The stimulus was presented repeatedly (6300 times over the course of 25 minutes) at alternating polarities at a rate of 4.35 Hz. Presentation of alternating polarities enables separate examination of the amplitude envelope and temporal fine structure of speech (Aiken & Picton, 2008; see the Data Analyses section for details). During the recording, participants read a magazine or a book of their preference and were asked to relax and restrain from extraneous body movement and not to pay attention to the sound. Continuous electrophysiological data were recorded using a BioSemi ActiveTwo EEG system at a 16384 Hz sample rate and with open filters in ActiView (BioSemi) acquisition software. A montage of five electrodes with a sintered Ag-AgCl pallet was used. One active electrode was placed on the top of the head (i.e., at Cz), two on the left and right earlobes as reference points and two on the forehead as ground electrodes.

2.3.3. Data processing.

Neuropsychological data processing was conducted using custom MATLAB scripts. First, recordings were bandpass filtered from 70 to 2000 Hz using a first-order Butterworth filter. Next, each trial was epoched between -30 and 210 ms with respect to stimulus onset. Trials containing amplitude spikes of >35 µV were rejected as artifacts, and the first 2500 artifact-free responses to each stimulus polarity were selected for analysis, for a total of 5000 sweeps. Finally, inter-trial phase locking analysis was used to measure the precision of neural sound encoding across trials on a frequency-by-frequency basis. For each trial, the time frequency spectrum was calculated using a Hanning windowed fast Fourier transform. This procedure generates, for each trial, an amplitude value and a phase value. The resulting vectors were then transformed into unit vectors, which retains the phase value but discards the amplitude, and averaged. The length of the resulting vector was calculated as a measure of inter-trial phase locking, which varies from zero (no consistency) to one (perfect consistency).

2.3.4. Data Analyses.

For analysis of neural encoding of the F0, inter-trial phase locking was calculated across all 5000 trials using the procedure described above for a response time window between 10 and 180 ms. F0 phase-locking was then calculated as maximum inter-trial phase coherence between 80-120 Hz. For analysis of neural encoding at F1 and F2, inter-trial phase locking was calculated only during the steady state (60-170 ms), i.e. during the portion of the response corresponding to the part of the stimulus in which the formants were constant. Moreover, before calculating phase-locking, the neural phase for trials corresponding to one polarity were flipped 180° relative to the other. This procedure emphasizes the temporal fine structure in the response, enabling investigation of the neural representation of the higher-frequency speech formants (Aiken & Picton, 2008). F1 phase-locking was then computed as the maximum inter-trial coherence between 680-720 Hz, while F2 phase-locking was computed as the mean of the maximum inter-trial coherence between 1180-1220 Hz and the maximum between 1280-1320 Hz.

Figure 1 displays the stimulus waveform and spectrum (top two panels), as well as the response waveform and phase-locking across the spectrum for the added (middle two panels) and subtracted (bottom two panels) polarities analyses. This figure illustrates the close resemblance between the stimulus and response waveform, as well as the fact that the subtracted polarities analysis can be used to measure encoding of the formants but not the F0, while the added polarities analysis can be used to measure encoding of the F0 but not the formants.

2.4. Procedure

Data collection was conducted at the Department of Psychological Sciences at Birkbeck, University of London. Each testing session lasted approximately 150 minutes. Tasks were administered in the following order: Auditory Processing Battery, GJT, LLAMA-E, Speech Perception Task, and auditory-motor temporal integration tasks. As the last step, the FFR was recorded. All instructions was delivered in English with Polish translation by an L1 Polish speaking researcher where necessary to avoid any misunderstandings of the procedure. Participants were reimbursed 20£ in cash for their time upon completion of the testing session.

2.5. Analysis

Several variables were non-normally distributed, and so underwent transformation prior to analysis. A 1/x transformation was used for age of acquisition, a rau transform was used for the rhythm memory and language aptitude tests, and a log transform was used for the psychophysical thresholds and F0 phase locking. Two outliers in synchronization variability were excluded (>2 SD from the mean). False Discovery Rate was used to correct for multiple comparisons when conducting multiple correlations (Benjamoni-Hochberg procedure, Benjamini & Hochberg, 1995). Processed (but untransformed) data can be found on Open Science Framework at https://osf.io/gwxkb/.

3. Results

3.1. Language experience and auditory processing as predictors of L2 learning. To examine the extent to which language experience versus auditory processing explained variance in L2 learning success, we performed two multiple linear regressions with backward elimination, with grammatical judgment and speech perception as the predicted variables (Table 1). Auditory processing (temporal processing, spectral processing, synchronization variability, rhythm memory, and neural encoding of F0, F1, and F2), linguistic aptitude (LLAMA performance) and demographic measures (age, age of acquisition, length of residence, years of in-class training, musical training, and frequency of English use) were entered as potential predictors. Musical training was entered as a categorical predictor, with participants given a value of 0 if they had undergone no musical training, and 1 if they had experienced at least one year of musical training.

For grammatical judgment, age and length of residence emerged as significant demographic predictors, indicating that participants who displayed better grammatical knowledge had been resident in English-speaking countries for longer amounts of time but were younger overall. Several auditory processing variables were also significant predictors, with more successful participants also showing more precise temporal processing, less variable auditory-motor temporal synchronization, and less robust neural encoding of F1. For speech perception, age of acquisition and length of in-class training emerged as significant demographic predictors, with better speech perception linked to earlier age of acquisition and shorter length of residence. Several auditory processing variables were also significant predictors, with more successful participants showing more precise spectral processing, less variable auditory-motor temporal synchronization, less robust neural encoding of F1, and more robust neural encoding of F2. For both grammatical judgment and speech perception, auditory processing variables were associated with the highest standardized betas, compared to language experience measures.

Given the large number of predictors included in the regression analyses and our moderate number of participants, some of the predictors outlined in the previous analysis could have been the result of overfitting. To provide a more stringent test of our hypothesis that auditory processing explains independent variance in L2 learning success even once demographic variables have been accounted for, we ran follow-up correlational analysis with False Discovery Rate correction for multiple comparisons. Partial correlations were conducted between behavioural and neural auditory processing measures and performance on the speech perception and grammatical judgment tests, with age, age of acquisition, length of residence, years of in-class training, and amount of daily English use as covariates (Table 2). As predicted, speech perception correlated with spectral processing ($r = -0.63$) and neural encoding of F2 ($r = 0.44$), such that more precise sound perception and more stable neural encoding were linked to more accurate vowel perception. Speech perception correlated with synchronization variability as well ($r = -0.46$), indicating that participants who were better able to perceive English vowels could more consistently synchronize to a metronome. Grammatical judgment correlated with spectral and temporal processing, indicating that participants who had more successfully acquired knowledge of English grammar could more precisely discriminate between sounds on the basis of both frequency and duration. Although in the regression analysis neural encoding of F1 emerged as a significant predictor of speech perception and grammatical knowledge, it did not significantly correlate with either outcome measure, making interpretation of the role of this predictor in the regression models difficult.

See Figure 2 for scatterplots displaying the relationship between psychoacoustic thresholds, neural encoding of F2, and language skills. See Figure 3 for a comparison of subtracted polarities phase-locking in participants with good versus poor English vowel perception (median split). This comparison demonstrates that the enhanced phase-locking in participants with good English vowel perception is largely limited to the frequency region surrounding F2 (1240 Hz).

3.2. Relationships between neural encoding of speech and behavioural measures of auditory processing. In an attempt to provide a biological framework explaining individual differences in auditory processing, we investigated partial correlations between frequency-following-response encoding of F0, F1, and F2 (as measured using inter-trial phase-locking) and behavioural measures of auditory processing (Table 3), covarying for age, age of acquisition, length of residence, years of in-class training, and amount of daily English use. As predicted, participants who displayed more robust neural encoding of the fundamental frequency were also better able to synchronize consistently to a metronome ($r = -0.54$). No other correlations between neural encoding and behavioural measures of auditory processing approached significance ($p > 0.1$).

3.3. Relationships between language experience and auditory processing. To test the hypothesis that language experience can enhance auditory processing, age of acquisition, length of residence, and years of L2 class training were correlated with the neural and behavioural measures of auditory processing listed above with age as a covariate. No correlations survived correction for multiple comparisons ($p > 0.1$). Nevertheless, this null result should be interpreted with caution, given that several correlations reached significance prior to correction.

4. Discussion

Although age and experience-related variables strongly predict the extent to which early bilinguals can ultimately attain second language proficiency (younger and more practice is better), it has been shown that post-pubertal second language learning (age of acquisition > 16 years) is subject to a great deal of individual variability. Even if two adults practice a target language for the same amount of time in an identical manner, their outcomes may most likely differ in various dimensions of language (Doughty, 2018). Building on first language acquisition literature (e.g., Bishop & McArthur, 2005), and following our preliminary investigations (Omote et al., 2017; Saito et al., 2018), we asked whether and to what degree individual differences in auditory processing play a role in determining the degree of success among late second language learners. In the current investigation, we tested native Polish speakers living in London and found that proficient English grammatical knowledge and speech perception abilities were both linked to more precise auditory discrimination while speech perception was additionally linked to the robustness of neural responses to sound. Auditory processing and measures of language experience such as age of acquisition and length of residence explained independent variance in speech perception and grammatical judgment. Finally, we found that behavioural and neural auditory measures were largely uncorrelated, with the exception of a relationship between fundamental frequency neural phase-locking and synchronization variability.

Auditory processing was the strongest predictor of L2 learning success, exceeding the predictive power of language input characteristics such as age of acquisition and length of residence, as well as a test of language learning aptitude. The spectral psychoacoustic measures alone, for example, could explain 40% of the variance in speech perception. This suggests that some individuals who struggle to learn a second language may do so because they lack the auditory precision to detect phonetic and prosodic structure. It is possible that these individuals could benefit from auditory training programs designed to remediate these deficits, potentially increasing the efficacy of existing methods of language instruction. Spectral processing, for example, can be boosted by as little as a few hours of training (Micheyl, Delhommeau, Perrot, & Oxenham, 2006), even in individuals who initially have very severe deficits (Whiteford & Oxenham, 2018), and the robustness of neural encoding of the fundamental frequency can also be enhanced by short-term pitch discrimination training (Carcagno & Plack, 2011).

Given that these auditory processing tests are easily automated and quick to run, they could be a useful addition to test batteries assessing language learning aptitude. Historically, language aptitude tests have limited themselves to assessing short-term explicit learning of linguistic structure (Carroll & Sapon, 1959), and have had only modest success in explaining variance in L2 learning. A recent meta-analysis, for example, revealed that language aptitude explained an average of only 9.6% of the variance in L2 grammar learning (Li, 2014). There has been recent interest in expanding tests of language aptitude to include cognitive assessments (Linck et al., 2013); our finding that auditory processing measurements strongly predict various aspects of second language learning performance suggests that it would be worthwhile to include auditory processing measurements in language aptitude test batteries as well.

We find that individuals who encode the neural representation of the second formant of speech more robustly are also better able to perceive English vowels. This is in accordance with previous findings that vowel perception is linked to the extent to which the frequency-following response distinguishes between different vowels (Won et al., 2016), and that formant encoding in the frequency-following response tracks with the accuracy of segmental L2 production (Saito et al., 2018). Here we extend these results by showing that neural formant encoding and behavioural measurements of spectral processing explain independent variance in speech perception. This suggests that they reflect two different aspects of spectral processing. One possibility is that psychoacoustic measurements reflect the ability to make explicit auditory judgments and therefore draw upon attention and short-term memory, while the frequency-following response, which is relatively unaffected by cognitive state (Varghese, Bharadwaj, & Shinn-Cunningham, 2015), is an implicit measurement which primarily reflects bottom-up perceptual resolution and fidelity. Together, these metrics may form a more comprehensive picture of an individual's capacity for second language learning.

We find that the link between robust auditory processing and successful L2 acquisition extends beyond speech perception to other language skills, as better L2 grammatical judgment tracked with more precise auditory perception. This suggests that precise representation of the acoustic characteristics of speech is foundational for the acquisition of syntactical knowledge. For example, imprecise auditory perception may interfere with the perception of prosodic features such as phrase boundaries, which are conveyed by brief changes in pitch and duration (de Pijper & Sanderman, 1994). Difficulties with prosody perception, in turn, could delay the acquisition of syntax, given that listeners can use prosody to detect hierarchical structure in language (Langus, Marchetto, Bion, & Nespor, 2012; Marslen-Wilson et al., 1992).

Extensive prior work has shown that rhythm skills—i.e. the ability to perceive and produce patterns in time—are linked to language skills in children (McGivern, Berka, Languis, & Chapman, 1991; Douglas & Willatts, 1994; David, Wade-Wolley, Kirby, & Smithrim, 2007; Thomson & Goswami, 2008; Corriveau & Goswami, 2009; Dellatolas, Watier, Le Normand, Lubart, & Chevrie-Muller, 2009; Huss, Verney, Fosker, Mead, & Goswami, 2011; Strait, Hornickel, & Kraus, 2011; Tierney & Kraus, 2013; Flaugnacco et al., 2014; González-Trujillo, Defior, & Gutiérrez-Palma, 2014; Gordon et al., 2015; Woodruff Carr, White-Schwoch, Tierney, Strait, & Kraus, 2014; Tierney et al., 2017). Here we show for the first time that rhythmic skill, in particular synchronization, is linked to second language acquisition in adulthood as well. Temporal patterns across multiple time scales convey a wealth of information about the structure of language, including phase boundaries (de Pijper & Sanderman, 1994), word boundaries (Smith, Cutler, Butterfield, & Nimmo-Smith, 1989), lexical stress (Liberman & Prince, 1977), and phonetic distinctions such as voicing (Lisker, 1957). Adults learning a second language who can also more easily detect and reproduce patterns in time may benefit more from the information hidden in temporal patterns of speech, potentially facilitating the acquisition of phonetic, semantic, and syntactic knowledge.

Our finding of a link between rhythm perception and production and second language acquisition is in line with our precursor research which demonstrated that adult second language learners with greater rhythmic sensitivity spoke more fluently (with faster speech rate and fewer pauses and repetitions) (Saito et al., 2018). This relationship between rhythm perception and second language learning may help explain prior findings of links between musical training and second language learning success (Slevc & Miyake, 2006; Martinez-Montes et al., 2013; Swaminathan & Gopinath, 2013; Cooper, Wang, & Ashley, 2017; Dittinger, D'Imperio, & Besson, 2018). Musicians demonstrate enhanced rhythm skills, including more precise synchronization and more accurate rhythm memory (Bailey & Penhune, 2010; Krause, Pollok, & Schnitzler, 2010), which may help them benefit from temporal structure in speech.

The auditory processing impairments which we find to be tied to difficulties with second language acquisition are strikingly similar to the auditory difficulties which have been linked to developmental language disorders. Children with dyslexia, for example, are more likely to display problems with the perception of temporal and spectral features (Ahissar, Protopapas, & Merzenich, 2000; Talcott et al., 2000; Amitay, Ahissar, & Nelken, 2002; McArthur & Bishop, 2005; Gibson, Hogben, & Fletcher, 2006; Goswami et al., 2010; Casini, Pech-Georgel, & Ziegler, 2017), to struggle to synchronize to a metronome (Thomson & Goswami, 2008), to have difficulty remembering rhythmic patterns (Flaugnacco et al., 2014), and to have neural responses to sound which are more variable across trials (Hornickel & Kraus, 2013; Lizarazu et al., 2015). Some research in second language acquisition assumes that first language acquisition in childhood and second language acquisition in adulthood are characterized by different underlying mechanisms (Abrahamsson & Hyltenstam, 2009; DeKeyser, Alfi-Shabta, & Ravid, 2010). According to this theory (i.e., Critical Period Hypothesis), upon leaving a putative critical period learners shift from implicit learning mechanisms (statistical learning) to explicit learning mechanisms (conscious inference of linguistic rules). Our results, on the other hand, suggest that difficulties with auditory encoding can be a bottleneck for language acquisition both in childhood and later in life. Thus, some of the mechanisms which facilitate language learning may continue to play similar roles throughout the lifespan (Flege et al., 1999; Birdsong & Molis, 2001; Hamrick, Lum, & Ullman, 2018).

Based on prior work reporting correlations between the robustness of the frequency-following response and the precision of auditory processing (Purcell et al., 2004; Krishnan et al., 2010, 2012; Marmel et al., 2013; Tierney & Kraus, 2013, 2016; Bharadwaj et al., 2015; Tierney et al., 2017), we predicted that neural encoding would relate to both spectral and temporal processing. Instead, although we found that greater phase-locking at the fundamental frequency was linked to more precise synchronization (replicating Tierney & Kraus, 2013), we found no significant correlations between psychoacoustic thresholds and neural encoding. To some extent this null result may be explained by methodological differences between the current study and previous studies. In particular, Purcell et al. (2004) and Bharadwaj et al. (2015) measured temporal processing using amplitude modulation *detection*, while we measured duration and amplitude rise time *discrimination*. Given that our design required participants to hold stimuli in memory and compare them, our measurements may have had a greater cognitive load, reflecting individual differences in cognitive skills such as attention and auditory short-term memory. As the frequency-following response is relatively unaffected by cognitive state (Varghese et al., 2015), measures of auditory processing under a greater cognitive load may be less closely tied to the early auditory neural encoding measured by the FFR (Coffey et al., 2016). This could explain why the only temporal processing measure that related to neural encoding was synchronization variability, a measure which in part reflects rapid, very precise subconscious auditory-motor integration (Repp, 2000). This explanation, however, cannot account for our non-replication of the finding that frequency-following response encoding is linked to frequency discrimination (Krishnan et al., 2010, 2012; Marmel et al., 2013). Even so, given that there was a weak tendency in our dataset for more robust neural encoding to be linked to more precise frequency discrimination, it is possible that we lacked the power to detect this relationship.

Cross-sectional studies cannot conclusively distinguish between predictors and consequences of language learning success. We did not find any significant correlations between either years of immersion in the UK or years of in-class training in Poland and auditory processing, arguably because much rapid learning happens within first few months of immersion, followed by relatively slow and plateaued development patterns (Munto & Derwing, 2008). Focusing on moderately experienced second language learners (length of residence > 1 year), our cross-sectional investigation of their biodemographic, audition and linguistic profiles suggests that pre-existing individual differences in auditory processing help determine success beyond the early phase of L2 learning. Nevertheless, it is plausible that experience learning a second language could have subtle effects on auditory processing which our results may partially reflect. Indeed, prior research has shown that bilinguals have enhanced neural phase-locking to the fundamental frequency of speech (Krizman, Marian, Shook, Skoe, & Kraus, 2012; Krizman, Skoe, Marian, & Kraus, 2014; Krizman, Skoe, & Kraus, 2016; Omote et al., 2017), greater grey matter volume within auditory cortex (Ressel et al., 2012), and enhanced musical rhythm perception (Roncaglia-Denissen, Roor, Chen, & Sadakata, 2016). Future longitudinal work examining auditory processing and L2 knowledge before and after immersion in a second language environment could help disentangle the roles of pre-existing differences versus neural plasticity in mediating the relationship between auditory processing and language learning.

Acknowledgements

The authors would like to thank all of the participants for taking part in this study. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

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

Ahissar, M., Protopapas, A., Reid, M., & Merzenich, M. (2000). Auditory processing parallels reading abilities in adults. *PNAS*, *97*, 6832-6837.

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

Amitay, S., Ahissar, M., & Nelken, I. (2002). Auditory processing deficits in reading disabled adults. *JARO*, *3*, 302-320.

Bailey, J., & Penhune, V. (2010). Rhythm synchronization performance and auditory working memory in early- and late-trained musicians. *Exp Brain Res*, *204*, 91-101.

Basu, M., Krishnan, A., & Weber-Fox, C. (2010). Brainstem correlates of temporal auditory processing in children with specific language impairment. *Developmental Science*, *13*, 77-91.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, *56*, 289-300.

Bharadwaj, H., Masud, S., Mehraei, G., Verhulst, S., & Shinn-Cunningham, B. (2015). Individual differences reveal correlates of hidden hearing deficits. *Journal of Neuroscience*, *35*, 2161-2172.

Bishop, D. V., & McArthur, G. M. (2005). Individual differences in auditory processing in specific language impairment: a follow-up study using event-related potentials and behavioural thresholds. *Cortex*, *41*(3), 327-341.

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

Carcagno, S., & Plack, C. (2011). Subcortical plasticity following perceptual learning in a pitch discrimination task. *JARO*, *12*, 89-100.

Carroll, J., & Sapon, S. (1959). *Modern Langauge Aptitude Test: Manual*. Psychological Corporation.

Casini, L., Pech-Georgel, C., & Ziegler, J. (2017). It's about time: revisiting temporal processing deficits in dyslexia. *Developmental Science*, *21*, e12530.

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

Coffey, E., Herholz, S., Chepesiuk, A., Baillet, S., & Zatorre, R. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature Communications*, *7*, 11070.

Cooper, A., Wang, Y., & Ashley, R. (2017). Thai rate-varied vowel length perception and the impact of musical experience. *Language and Speech*, *60*, 65-84.

Corriveau, K., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: tapping to the beat. *Cortex*, *45*, 119-130.

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

David, D., Wade-Wolley, L., Kirby, J., & Smithrim, K. (2007). Rhythm and reading development in school-age children: a longitudinal study. *Journal of Research in Reading*, *30*, 169-183.

de Pijper, J., & Sanderman, A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *JASA*, *96*, 2037-2047.

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

DeKeyser, R., Alfi-Shabta, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*, 413–438.

Dellatolas, G., Watier, L., Le Normand, M., Lubart, T., & Chevrie-Muller, C. (2009). Rhythm reproduction in kindergarten, reading performance at second grade, and developmental dyslexia theories. *Archives of Clinical Neuropsychology*, *24*, 555-563.

Derwing, T., & Munro, M. (2013). The development of L2 oral language skills in two L1 groups: A seven-year study. *Language Learning*, *63*, 163–185.

Dittinger, E., D'Imperio, M., & Besson, M. (2018). Enhanced neural and behavioral processing of a non-native phonemic contrast in professional musicians. *European Journal of Neuroscience*, *47*, 1504-1516.

Doughty, C. (2018). Cognitive language aptitude. *Language Learning*. DOI: https://doi.org/10.1111/lang.12322

Douglas, S., & Willatts, P. (1994). The relationship between musical ability and literacy skills. *Journal of Research in Reading*, *17*, 99-107.

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

Fear, B., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *JASA*, *97*, 1893-1904.

Flaugnacco, E., Lopez, L., Terribili, C., Zoia, S., Buda, S., Tilli, S., Monasta, L., Montico, M., Sila, A., Ronfani, L., & Schön, D. (2014). Rhythm perception and production predict reading abilities in developmental dyslexia. *Frontiers in Human Neuroscience*, *8*, 392.

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

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

Flege, J., Yeni-Komshian, G., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*, 78-104.

Flege, J., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *SSLA*, *23*, 527-552.

Foster, P., Bolibaugh, C., & Kotula, A. (2014). Knowledge of nativelike selections in a L2: the influence of exposure, memory, age of onset, and motivation in foreign language and immersion settings. *Studies in Second Language Acquisition*, *36*, 101-132.

Freed, B. F., Dewey, D. P., Segalowitz, N., & Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition*, *26*(2), 349-356.

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

Godfroid, A., Loewen, S., Jung, S., & Park, J. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: evidence from eye-movement patterns. *Studies in Second Language Acquisition*, *37*, 269-297.

Golestani, N., & Pallier, C. (2007). Anatomical correlates of foreign speech sound production. *Cerebral Cortex*, *17*, 929-934.

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex*, *17*, 575-582.

González-Trujillo, M., Defior, S., & Gutiérrez-Palma, N. (2014). The role of nonspeech rhythm in Spanish word reading. *Journal of Research in Reading*, *37*, 316-330.

Gordon, R., Shivers, C., Wieland, E., Kotz, S., Yoder, P., & McAuley, J. (2015). Musical rhythm discrimination explains individual differences in grammar skills in children. *Developmental Science*, *18*, 635-644.

Goswami, U., Wang, H., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2010). Language-universal sensory deficits in developmental dyslexia: english, spanish, and chinese. *Journal of Cognitive Neuroscience*, *23*, 325-337.

Goswami, U., Mead, N., Fosker, T., Huss, M., Barnes, L., & Leong, V. (2013). Impaired perception of syllable stress in children with dyslexia: a longitudinal study. *Journal of Memory and Language*, *69*, 1-17.

Hamrick, P., Lum, J., & Ullman, M. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *PNAS*, *115*, 1487-1492.

Hu, X., Ackermann, H., Martin, J., Erb, M., Winkler, S., & Reiterer, S. (2013). Language aptitude for pronunciation in advanced second language (L2) learners: Behavioural predictors and neural substrates. *Brain & Language*, *127*, 366-376.

Huss, M., Verney, J., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. *Cortex*, *47*, 674-689.

Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, *24*, 131-161.

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

Kempe, V., Bublitz, D., & Brooks, P. (2015). Musical ability and non-native speech-sound processing are linked through sensitivity to pitch and spectral information. *British Journal of Psychology*, *106*, 349-366.

Krause, V., Pollok, B., & Schnitzler, A. (2010). Perception in action: the impact of sensory information on sensorimotor synchronization in musicians and non-musicians. *Acta Psychologica*, *133*, 28-37.

Krishnan, A., Bidelman, G., & Gandour, J. (2010). Neural representation of pitch salience in the human brainstem revealed by psychophysical and electrophysiological indeces. *Hearing Research*, *268*, 60-66.

Krishnan, A., Bidelman, G., Smalt, C., Ananthakrishnan, S., & Gandour, J. (2012). Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia*, *50*, 2849-2859.

Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences*, *109*, 7877-7881.

Krizman, J., Skoe, E., Marian, V., & Kraus, N. (2014). Bilingualism increases neural response consistency and attentional control: evidence for sensory and cognitive coupling. *Brain and Language*, *128*, 34–40.

Krizman, J., Skoe, E., & Kraus, N. (2016). Bilingual enhancements have no socioeconomic boundaries. *Developmental Science*, *19*, 881–891

Langus, A., Marchetto, E., Bion, R., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language*, *66*, 285-306.

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. *JASA*, *128*, 3757-3568.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *JASA*, *49.2B*, 467-477.

Li, X. (2010). RTbox: a device for highly accurate response time measurements. *Behavior Research Methods*, *42*, 212-225.

Li, S. (2014). The associations between language aptitude and second language grammar acquisition: a meta-analytic review of five decades of research. *Applied Linguistics*, *36*, 385-408.

Li, S. (2016). The construct validity of language aptitude. *Studies in Second Language Acquisition*, *38*, 801-842.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*, 249-336.

Linck, J., Hughes, M., Campbell, S., Silbert, N., Tare, M., Jackson, S., Smith, B., Bunting, M., & Doughty, C. (2013) Hi-LAB: a new measure of aptitude for high-level language proficiency. *Language Learning*, *63*, 530-566.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, *33*, 42-49.

Lizarazu, M., Lallier, M., Molinaro, N., Bourguignon, M., Paz-Alonso, P., Lerma-Usabiaga, G., & Carreiras, M. (2015). Developmental evaluation of atypical auditory sampling in dyslexia: functional and structural evidence. *Human Brain Mapping*, *36*, 4986-5002.

Marmel, F., Linley, D., Carlyon, R., Gockel, H., Hopkins, K., & Plack, C. (2013). Subcortical neural synchrony and absolute thresholds predict frequency discrimination independently. *JARO*, *14*, 757-766.

Marslen-Wilson, W., Tyler, L., Warren, P., Grenier, P., & Lee, C. (1992). Prosodic effects in minimal attachment. *The Quarterly Journal of Experimental Psychology*, *45A*, 73-87.

Martinez-Montes, E., Hernández-Pérez, H., Chobert, J., Morgado-Rodriguez, L., Suárez-Murias, C., Valdés-Sosa, P., & Besson, M. (2013). Musical expertise and foreign speech perception. *Frontiers in Systems Neuroscience*, *7*, 84.

McArthur, G., & Bishop, D. (2005). Speech and non-speech processing in people with specific language impairment: a behavioural and electrophysiological study. *Brain and Language*, *94*, 260-273.

McGivern, R., Berka, C., Languis, M., & Chapman, S. (1991). Detection of deficits in temporal pattern discrimination using the seashore rhythm test in young children with reading impairments. *Journal of Learning Disabilities*, *24*, 58-62.

Meara, P. (2005). *LLAMA language aptitude tests: the manual.* Swansea: Lognostics.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*, 36-47.

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

Omote, A., Jasmin, K., & Tierney, A. (2017). Successful non-native speech perception is linked to frequency following response phase consistency. *Cortex*, *93*, 146-154.

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, *38*, 422-430.

Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels. *JASA*, *24*, 175-184.

Povel, D., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, *2*, 411-440.

Purcell, D., John, S., Schneider, B., & Picton, T. (2004). Human temporal auditory acuity as assessed by envelope following responses. *JASA*, *116*, 3581-3593.

Repp, B. (2000). Compensation for subliminal timing perturbations in perceptual-motor synchronization. *Psychological Research*, *63*, 106-128.

Ressel, V., Pallier, C., Ventura-Campos, N., Díaz, B., Roessler, A., Avila, C., & Sebastián-Gallés, N. (2012). An effect of bilingualism on the auditory cortex. *Journal of Neuroscience*, *32*, 16597-16601.

Richards, S., & Goswami, U. (2015). Auditory processing in specific language impairment (SLI): relations with the perception of lexical and phrasal stress. *JSLHR*, *58*, 1292-1305.

Rojczyk, A. (2010). Forming new vowel categories in second language speech: The case of Polish learners' production of English /ɪ/ and /e/. *Research in Language*, *8*, 85-97.

Rojczyk, A. (2012). Non-native speech perception in noise: The voicing contrast in English. *Linguistica Silesiana*, *33*, 7-17.

Roncaglia-Denissen, M., Roor, D., Chen, A., & Sadakata, M. (2016). The enhanced musical rhythm perception in second language learners. *Frontiers in Human Neuroscience*, *10*, 288.

Saito, K., & Hanzawa, K. (2016). Developing second language oral ability in foreign language classrooms: The role of the length and focus of instruction and individual differences. *Applied Psycholinguistics*, *37*(4), 813-840.

Saito, K. (2017). Effects of Sound, Vocabulary, and Grammar Learning Aptitude on Adult Second Language Speech Attainment in Foreign Language Classrooms. *Language Learning*, *67*(3), 665-693.

Saito, K., Sun, H., & Tierney, A. (2018). 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*. doi: 10.1017/S1366728918000895

Schwartz, G., Aperlinski, G., Jekiel, M., & Malarski, K. (2016). Spectral dynamics in L1 and L2 vowel perception. *Research in Language*, *1*, 61-77.

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

Smith, J. (2007). *Introduction to Digital Filters with Audio Applications.* W3K Publishing.

Smith, M., Cutler, A., Butterfield, S., & Nimmo-Smith, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech and Hearing Research*, *32*, 912-920.

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

Swaminathan, S., & Gopinath, J. (2013). Music training and second-language English comprehension and vocabulary skills in Indian children. *Psychological Studies*, *58*, 164-170.

Talcott, J., Witton, C., McLean, M., Hansen, P., Rees, A., Green, G., & Stein, J. (2000). Dynamic sensory sensitivity and children's word decoding skills. *PNAS*, *97*, 2952-2957.

Thomson, J., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *Journal of Physiology*, *102*, 120-129.

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

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

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

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

Turk, A., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, *27*, 171-206.

Varghese, L., Bharadwaj, H., & Shinn-Cunningham, B. (2015). Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. *Brain Research*, *1626*, 146-164.

Wang, L., Li, X., & Yang, Y. (2014). A review on the cognitive function of information structure during language comprehension. *Cognitive Neurodynamics*, *8*, 353-361.

Whiteford, K., & Oxenham, A. (2018). Learning for pitch and melody discrimination in congenital amusia. *Cortex*, *103*, 164-178.

Won, J., Tremblay, K., Clinard, C., Wright, R., Sagi, E., & Svirsky, M. (2016). The neural encoding of formant frequencies contributing to vowel identification in normal-hearing listeners. *JASA*, *139*, 1-11.

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

Woodruff Carr, K., White-Schwoch, T., Tierney, A., Strait, D., & Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proceedings of the National Academy of Sciences*, *111*, 14559-14564.

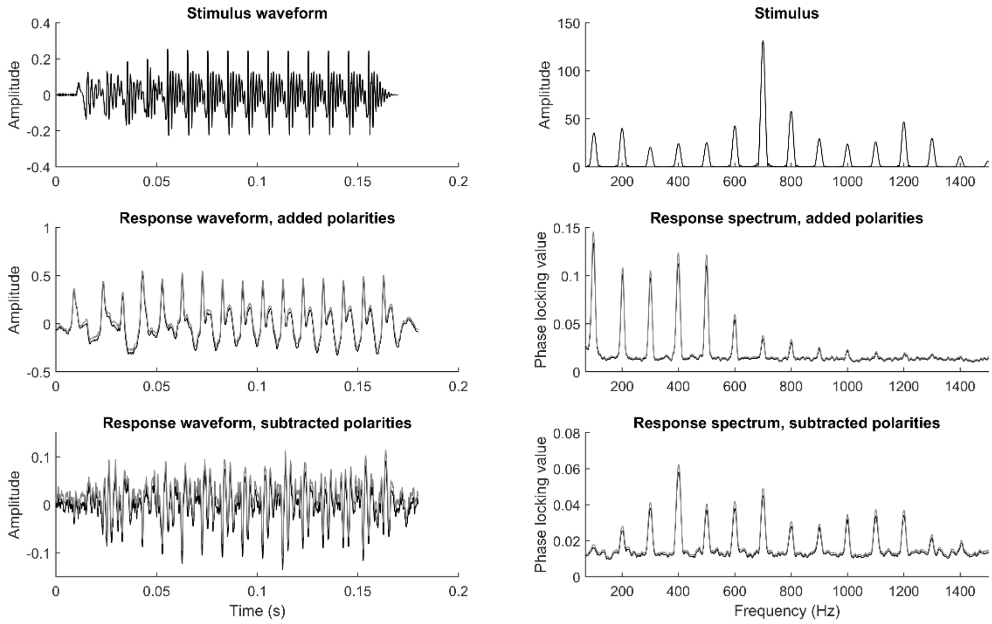

Figure 1. Schematic displaying similarities between stimulus and response waveforms and spectra. (Top left) Stimulus waveform. (Top right) Stimulus spectrum calculated from 50 to 170 ms using a Hanning-windowed fast Fourier transform. (Middle left) Average response waveform across all participants, calculated by adding both polarities. (Bottom left) Average response waveform, subtracted polarities. (Middle right) Inter-trial phase-locking calculated without manipulating the phase of either polarity (equivalent to adding polarities). (Bottom right) Inter-trial phase-locking calculated with a 180-degree shift of one polarity relative to the other (equivalent to subtracting polarities). Across the bottom four panels, the grey line indicates +1 standard error of the mean.

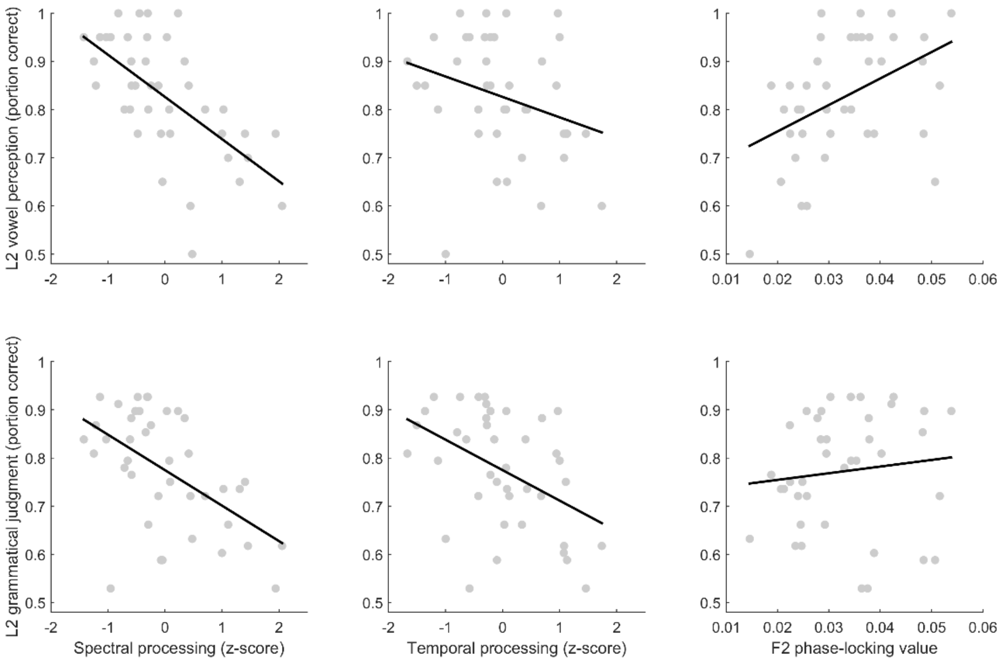

Figure 2. (Top left) Relationship between spectral processing (frequency and formant discrimination thresholds) and English vowel perception. (Top middle) Relationship between temporal processing (duration and rise time discrimination thresholds) and English vowel perception. (Top right) Relationship between neural encoding of F2 (phase-locking value) and English vowel perception. (Bottom left) Relationship between spectral processing and English grammatical judgment. (Bottom middle) Relationship between temporal processing and English grammatical judgment. (Bottom right) Relationship between neural encoding of F2 and English grammatical judgment.

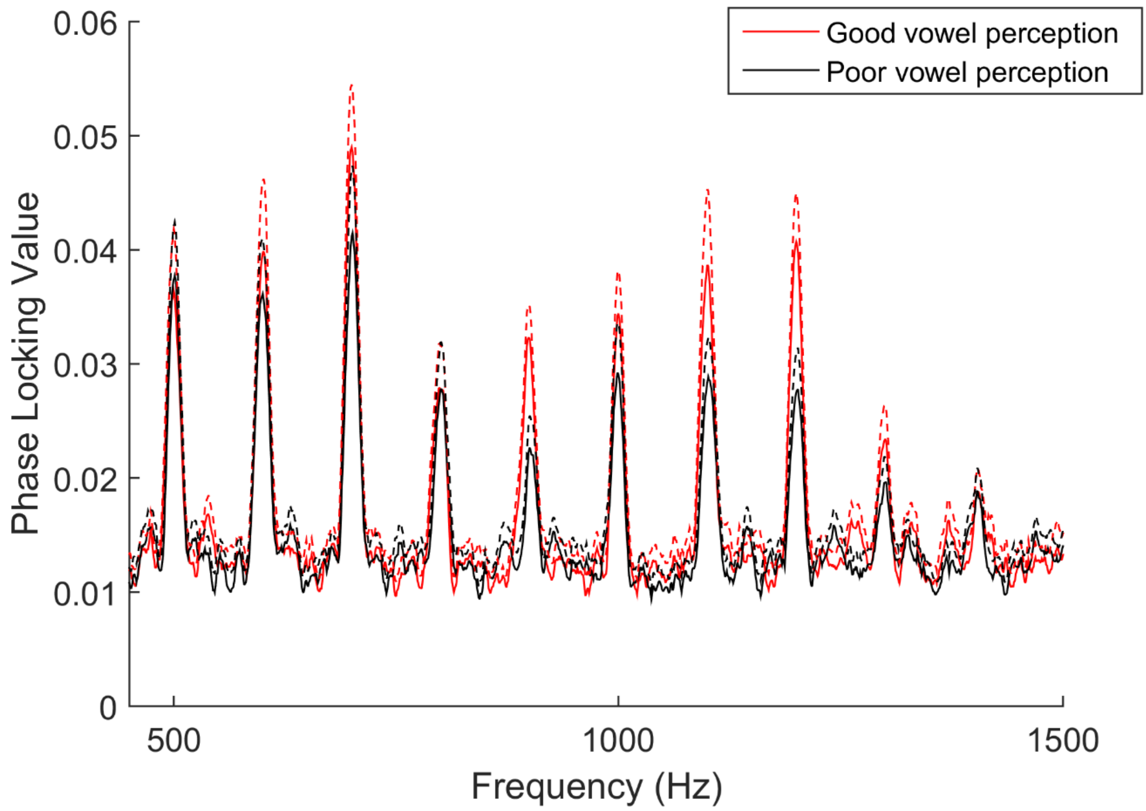

Figure 3. Subtracted polarities inter-trial phase locking in participants with good vowel perception (red) and poor vowel perception (black). Participants were separated into groups based on a median split in performance. The dotted lines indicate plus one standard error.

Predicted variable	Predictor	Standardized Beta	t
Grammatical judgment	Temporal processing	-0.43	-3.53***
	Synchronization variability	-0.37	-2.86**
	Neural encoding of F1	-0.28	-2.29*
	Length of residence	0.37	2.60*
	Age	-0.38	-2.79**
Speech perception	Spectral processing	-0.44	-3.84***
	Synchronization variability	-0.35	-3.06**
	Neural encoding of F1	-0.27	-2.58*
	Neural encoding of F2	0.39	3.57***
	Age of acquisition	-0.34	-3.28**
	Length of in-class training	-0.31	-2.72*

* p < 0.05, ** p < 0.01, *** p < 0.001

Table 1. Final models predicting grammatical judgment and speech perception performance after multiple linear regression with backward elimination, with only significant predictors retained (p < 0.05).

	Speech perception	Grammatical judgment
Spectral processing	-.63***	-.43*
Temporal processing	-.27	-.44*
Synchronization variability	-.44*	-.39
Rhythmic memory	.22	.22
Neural encoding of F0	.32	.35
Neural encoding of F1	-.05	-.15
Neural encoding of F2	.49*	.15
Linguistic aptitude	.31	.39

* p < 0.05, ** p < 0.01, *** p < 0.001

Table 2. Partial correlations between behavioural and neural measures of auditory processing and L2 speech and syntax processing, covarying for language experience and age. P-values have been FDR corrected for multiple comparisons.

	F0 encoding	F1 encoding	F2 encoding
Spectral processing	-.18	-.12	-.32
Temporal processing	-.10	-.29	-.12
Synchronization variability	-.54*	-.15	-.10
Rhythmic memory	-.01	-.06	-.13

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3. Pearson's correlations between behavioural and neural measures of auditory processing. P-values have been FDR corrected for multiple comparisons.