

Spoken Second Language Vocabulary Attainment Beyond the Critical Period: Roles of Domain-
General Auditory Processing

AUDITORY PROCESSING & VOCABULARY

Abstract

Recently, scholars have begun to explore the hypothesis that individual differences in domain-general auditory perception, which has been identified as an anchor of L1 acquisition, could explain some variance in post-pubertal L2 learners' segmental and suprasegmental learning in immersive settings. The current study set out to examine the generalizability of the topic to the acquisition of higher-level linguistic production skills—i.e., the appropriate use of diverse, rich, and abstract vocabulary. The speech of 100 Polish-English bilinguals was elicited using an interview task, submitted to corpus-/rater-based linguistic analyses, and linked to their ability to discriminate sounds based on individual acoustic dimensions (pitch, duration, amplitude).

According to the results, those who attained more advanced L2 lexical proficiency demonstrated not only more relevant experience (extensive immersion, earlier age of arrival), but also more precise auditory perception ability.

Key words: Second language speech, aptitude, auditory processing, vocabulary, grammar

AUDITORY PROCESSING & VOCABULARY

Over the past 50 years, scholars have shown that outcomes of post-pubertal second language (L2) speech learning are characterized by a great deal of individual differences. Many scholars have ascribed such variation to a range of biographical factors, such as earlier age of acquisition (Hopp & Schmid, 2013), longer immersion experience (Trofimovich & Baker, 2006), more frequent L2 use (Jia & Aaronson, 2003), and greater levels of willingness to communicate (Derwing & Munro, 2013). However, such experience-related factors do not fully explain the observed variances, especially when it comes to the incidence of highly advanced L2 learners (Abrahamsson & Hyltenstam, 2009) and the acquisition of relatively complex, non-salient, and difficult linguistic features (e.g., Li, 2016). According to the aptitude-acquisition view (Doughty, 2019), even learners with comparable biographical backgrounds who spend the same amount of time practicing a target language in the same fashion may vary greatly in their resulting levels of proficiency. This is arguably due to perceptual-cognitive individual differences (i.e., aptitude) that in turn determine the extent to which L2 learners can take advantage of every input opportunity, maximizing the long-term learning gains. Following this line of thought, a growing number of scholars have begun to demonstrate that domain-general auditory processing, which cognitive psychology literature has identified as a foundation of first language (L1) acquisition, can explain some variances in phonological dimensions of L2 speech learning (Kachlicka, Saito, & Tierney, 2019; Omote, Jasmin, & Tierney, 2017; for a comprehensive review, Saito & Tierney, forthcoming). In this paper, we report the results of an empirical study examining the extent to which the link between auditory precision and acquisition can be generalized to adult L2 learners' processing and acquisition of higher-order linguistic information, i.e., the appropriate use of diverse, rich, and abstract vocabulary during spontaneous speech in the context of 100 Polish-English bilinguals with varied age and experience backgrounds in the UK.

Background

Modeling, Assessing, and Developing Spoken L2 Vocabulary Proficiency

Whereas vocabulary is considered to be an integral unit of L2 learning, much of the existing work has been concerned with the assessment and development of *receptive* vocabulary knowledge. It has been shown (a) that such receptive knowledge can be operationalized as vocabulary size (2-3k frequent word families for beginner L2 speakers; 24k frequent word families for L1 speakers; Webb & Nation, 2017); (b) that L2 speakers' vocabulary size continues to improve as a function of increased input (for a discussion on the number of encounters vs. acquisition, see Pellicer-Sánchez, 2016); (c) that many learners can achieve nativelike vocabulary size as long as they engage in a great deal of L2 immersion experience (Hellman, 2011); and (d) that vocabulary size may be strongly correlated with a wide range of global L2 skills (e.g., listening, reading, speaking, and writing; see Schmitt, 2010).

In contrast, *productive* L2 vocabulary has remained understudied. For example, word frequency does not serve as a reliable index of advanced L2 speakers' spoken vocabulary use as they do not necessarily use more infrequent words while speaking (Crossley, Skalicky, Kyle, & Monteiro, 2019). For a long time, scholars have debated on how productive vocabulary knowledge can be assessed, how L2 speakers develop it, and what kinds of factors matter for its acquisition (for a review, see Koizumi, 2012). With respect to *spoken* L2 vocabulary, prior studies have indicated that even highly experienced L2 speakers' productive vocabulary use is subject to a great deal of individual variation, hinting at the possibility that some form of *aptitude* may play a very critical role in determining the incidence of high-level productive L2 vocabulary attainment (e.g., Hyltenstam, 1988).

AUDITORY PROCESSING & VOCABULARY

Recently, Crossley and his colleagues have proposed, developed, and refined a computational model of L2 learners' spoken vocabulary use (Crossley, Salsbury, & McNamara, 2015; Kyle & Crossley, 2015). Within this framework, the lexical dimensions of L2 speech are analyzed from two different perspectives. The first dimension (appropriateness) is defined as the ability to use a combination of words in a contextually appropriate and nativelike manner with the correct assignment of morphological markers. For similar accounts of appropriateness, see the semantic, collocational, and grammatical functions of word in Nation's (2001) model of L2 vocabulary knowledge. The second dimension (richness) is defined as the ability to use more infrequent, context-specific, and abstract words. This corresponds to the width and breadth and depth of word knowledge in Ellis's (2002) model of lexical acquisition. Though few in number, empirical studies have examined how the appropriateness and richness aspects of L2 lexical knowledge develop among various types of L2 learners.

In terms of the initial phase of immersion (length of residence [LOR] < 1 year), much of the learning appears to benefit L2 learners' use of rich and varied vocabulary. Crossley and his colleagues longitudinally analyzed the lexical richness of six L2 learners' L2 speech development over one year. Participants' spoken vocabulary quickly became more abstract, using more hypernyms and less concrete words, especially within the first four months (Salsbury, Crossley, & McNamara, 2011). As for the ultimate attainment of more experienced and advanced L2 learners' vocabulary use, the literature has been severely limited. Bartning, Lundell, and Hancock (2012) investigated the spoken morphological accuracy of the speech of 20 experienced late native Swedish learners of L2 (LOR > 5 years). The results demonstrated that the participants' accuracy performance was significantly distinguishable from inexperienced learners (LOR < 2 years) and native controls.

In the context of 100+ Japanese learners of English with varied experience profiles in naturalistic and classroom settings, Saito (2015, 2019, in press) examined the degree of vocabulary appropriateness (lexical, collocational, and morphological accuracy) and richness (frequency, range/context-specificity, abstractness). Experienced learners' (LOR > 6 years) spoken vocabulary use was significantly more accurate, varied, and richer than that of inexperienced learners in spontaneous speech. Interestingly, whereas few ultimately attained nativelike lexical accuracy, many experienced participants' richness performance was indistinguishable from native controls. The results indicate that the rate and ultimate attainment of spoken L2 vocabulary learning may differ in appropriateness and richness. On the one hand, many L2 learners can expand L2 vocabulary richness and reach a nativelike level within a short period of immersion (< 1 year), as long as they practice and use the target language. On the other hand, whereas L2 learners' vocabulary use tends to be more accurate as a result of increased immersion, the incidence of *nativelike* accuracy appears to be limited to very few individuals (cf. Hyltenstam, 1988).

In the current study, we test the hypothesis that the outcomes of spoken L2 vocabulary development can be explained not only by experience-related factors (length, quality, and timing of L2 use), but also by learners' aptitude profiles (i.e., auditory processing). More specifically, we assume that the aptitude and acquisition link can be most clearly observed especially in the relatively difficult aspects of L2 vocabulary learning—i.e., appropriateness rather than richness (cf. see the Results section for the benchmark analyses of L1 and L2 speakers' vocabulary proficiency). In this way, those with greater aptitude are expected to attain high-level L2 lexical proficiency after years of immersion as they can make the most of every practice opportunity in naturalistic settings (Doughty, 2019).

AUDITORY PROCESSING & VOCABULARY

Domain-General Auditory Processing in L1 Acquisition

In the field of cognitive psychology, one major theoretical debate concerns whether, to what degree, and how certain regions of the brain are specifically involved in human language acquisition (for an overview, see Campbell & Tyler, 2018). One influential view states that the same perceptual-cognitive faculties govern a range of general purpose learning behaviors including language learning, and an example of such a domain-general capacity that has received much attention is auditory processing. This ability is collectively referred to as a set of basic, low-level perception skills to encode, represent, and remember frequency and time dimensions of sounds (e.g., pitch, formants, duration, and amplitude). Many scholars have argued that individual differences in such auditory perception skills play a key role in the speed, development, and delay in first language (L1) acquisition (i.e., the auditory-deficit theory; Goswami, 2015; Tallal, 2004).

Auditory processing serves as “the gateway to spoken language” (Mueller, Friederici, & Männel, 2012, p. 15953), as it anchors every stage of phonological, lexical, and morphosyntactic processing. In order to detect phonetic and phonological categories, it is necessary to encode the relative weights of multiple acoustic cues, such as formant height, shape, and length for vowels (Kuhl, 2000) and approximants (Espy-Wilson et al., 2000), pitch and voice onset time for stop consonants (Shultz, Francis, & Llanos, 2012), and pitch height and contour for lexical tones (Chandrasekaran, Sampath, & Wong, 2010). More robust, prompt, and automatic phonetic and phonological analyses directly relate to the activation of contextually appropriate target words (Norris & McQueen, 2008), the detection of word and sentence boundaries (Cutler & Butterfield, 1992), and the refinement of morphological details (Joanisse & Seidenberg, 1998).

Among typical language development, auditory sensitivity continues to grow up to the age of 8 to 9, followed by a gradually declining curve through older adulthood (Skoe, Krizman, Anderson, & Kraus, 2015). There is ample evidence that when toddlers experience difficulties at the level of basic lower-level auditory perception, their acquisition of phonetic, phonological, lexical, and morphosyntactic knowledge is slowed down, resulting in a range of global language problems (for a meta-analysis, see Hämäläinen, Salminen, & Leppänen, 2012). For example, global language skills, such as reading and phonological awareness, are linked to the perception of non-verbal spectral (pitch, formants) and temporal (duration, amplitude) encoding (Foxton, Talcott, Witton, Brace, McIntyre, & Griffiths, 2003; Grube, Kumar, Cooper, Turton, & Griffiths, 2012). Thus, there is much correlational evidence showing that dyslexic children are more likely to have auditory deficits (Casini, Pech-Georgel, & Ziegler, 2018; Goswami, Wang, Cruz, Fosker, Mead, & Huss, 2011; Won, Tremblay, Clinard, Wright, Sagi, & Svirsky, 2016). Some scholars have suggested auditory processing measures as a diagnostic tool for dyslexia (Hornickel & Kraus, 2013) and other language-related disorders (Russo et al., 2008).

As for normal-hearing *children* (i.e., children who have not been diagnosed with specific language impairment or dyslexia), there is ample research examining the relationship between individual differences in auditory processing and language skills (e.g., Anvari, Trainor, Woodside, & Levy, 2002; Bavin, Grayden, Scott, & Stefanakis, 2010; Boets, Wouters, Van Wieringen, De Smedt, & Ghesquiere, 2008; Douglas & Willatt, 1994; Lamb & Gregory, 1993; Talcott, Witton, McLean, Hansen, Rees, Green, & Stein, 2000; Tierney, Gomez, Fedele, & Kirkham, 2021). In essence, these studies have indicated (a) that children without hearing impairment nonetheless vary in auditory abilities; and (b) that this variability is linked to a range of language outcomes (speech-in-noise perception, vocabulary use, literacy, and phonological awareness).

AUDITORY PROCESSING & VOCABULARY

When it comes to normal-hearing *adults*, similar individual variation has been observed (e.g., Kidd, Watson, Gygi, 2007). However, the correlations between auditory processing and speech perception abilities appear to be unclear (e.g., Surprenant & Watson, 2001; but see Ahissar, Protopapas, Reid, & Merzenich, 2000). One possible reason for this could be related to the various *redundancies* in speech perception. Every phonological contrast involves the complex integration of multiple acoustic signals. Due to the existence of multiple redundant cues, when listeners fail to perceive one, they may still accurately perceive the phoneme based on a different cue (e.g., the perception of English stops using voice onset time and/or pitch in following vowels; Toscano & McMurray, 2010). As L1 speakers regularly engage in language-based interactions during which they receive input for prolonged periods of time, accumulating a great deal of relevant speech perception experience, even those with particular auditory deficits may identify/adopt unique cue weighting strategies to optimize speech recognition (e.g., Jasmin, Dick, Holt, & Tierney, 2019 for the case of amusics using duration rather than pitch cues for the normal perception of speech and music).

Domain-General Auditory Processing in L2 Acquisition

More recently, some scholars (e.g., Saito & Tierney, forthcoming) have begun to argue not only that auditory processing could explain some variance in adult L2 learners' speech learning outcomes, but also that it may play an even more influential role in L2 than L1 acquisition because of the quantitative and qualitative differences between L1 and L2 learning processes. In L1 acquisition, even infants with auditory perception deficits may overcome acquisition problems with extensive exposure to input for a long period of time (Rosen, 2003). Contrastingly, adult L2 learners typically have limited access to exposure to their target language, even under immersion conditions (Jia & Aaronson, 2003). Unlike L1 learners, the lack of sufficient exposure opportunities may prevent L2 speakers from compensating for any perceptual deficit hindering their L2 comprehension development. In L2 learning contexts, any perceptual advantage or disadvantage can more strongly predict the extent to which L2 learners can benefit most from such limited input opportunities (Doughty, 2019).

Compared to L1 acquisition, in which auditory category learning takes place on a *blank* state (free of prior phonetic experience), it is important to note that adult L2 learners filter a new language input through their already-established auditory representations. In particular, they have to attend to new cues when L2 phonetic and phonological categories differ from L1. For example, Japanese speakers must learn to perceive difference in the third formant to acquire English [r] and [l] (Iverson et al., 2003), as well as to adjust and re-tune to existing analysis patterns when the cue weightings only partially overlap between L1 and L2 sounds (e.g., Chinese speakers need to deprioritize pitch and prioritize duration cues to acquire English word and sentence stress patterns; Jasmin, Sun, & Tierney, 2020). Developing or/and adjusting perceptual strategies to rely on new sources of input may draw on the ability to precisely and explicitly encode auditory dimensions, and so individual differences in auditory processing may demonstrate even greater predictive power for adult L2 speech learning success (Doughty, 2019).

With regards to testing the auditory-deficit hypothesis in L2 acquisition, there is a growing body of empirical research featuring a wide range of adult L2 speakers with diverse experience backgrounds. This work has found that individuals with more precise auditory processing abilities can hear and remember unfamiliar sounds and words more quickly when they are exposed to them (e.g., Kempe et al., 2015; Wong & Perrachione, 2007). In addition, individuals who demonstrate greater sensitivity to key acoustic information in a gradient manner (perceiving fine differences within sound categories) are more capable of integrating multiple

AUDITORY PROCESSING & VOCABULARY

cues to a single percept (e.g., Kim, Clayards, & Kong, 2020 for the weighting of vowel quality and quantity in the perception of synthesized vowel contrasts; Kong & Edwards, 2016 for the weighting of voice onset time and pitch in the perception of synthesized stop voicing contrasts; Chandrasekaran, Sampath, & Wong, 2010 for the weighting of pitch direction and height in English speakers' perception of Mandarin lexical tones).

When it comes to naturalistic L2 speech learning, it is individual variation in auditory processing that determines learning success even after the length and quality of experience is controlled for (Omote et al., 2017 for segmental and suprasegmental perception; Saito, Sun, & Tierney in press-a for segmental and suprasegmental production). Interestingly, the relationship between auditory processing and acquisition tends to be stronger when L2 learners have engaged in a sufficient amount of immersion experience (e.g., > 1 year: Saito, Sun, Kachlicka, Alayo, Nakata, & Tierney, 2020), and when the analyses focus on the relatively difficult aspects of L2 speech learning (e.g., phonological accuracy rather than fluency; Saito et al., in press-a). In contrast, the predictive power of auditory processing may be smaller when learners lack opportunities to be exposed to extensive, interactive and varied aural input (Saito, Tran, Suzukida, & Tierney, 2021 for classroom L2 learners).

Not surprisingly, all the aforementioned studies have exclusively concerned the relationship between auditory perception and L2 phonology since the role of auditory input processing is most directly relevant to segmental and suprasegmental acquisition. If we take the theoretical stance that auditory processing is a bottleneck for various dimensions of L2 speech acquisition, the question now becomes the extent to which auditory processing influences the acquisition of higher-level linguistic competence beyond phonological refinements —i.e., appropriate use of rich and varied vocabulary items. The current study is designed to address this issue.

Motivation for Current Study

There are several reasons to predict the presence of a relationship between more precise auditory processing and spoken L2 vocabulary development. *Although the context of the topic is exclusively limited to English, there is some discussion to support the hypothesis that auditory processing drives the lexical, morphosyntactic, and global aspects of L2 learning.*

At the lexical level, the detection of L2 lexical and sentence stress patterns is claimed to be fundamental to segmenting and making input available for word analyses (Field, 2005). Given that these linguistic phenomena are marked by changes in pitch, duration, and amplitude, it is reasonable to assume that individuals with greater sensitivity to the relevant acoustic dimensions can better encode, notice, and internalize novel or L2 prosodic patterns relative to L1 counterparts (Chandrasekaran et al., 2010). Relatedly, speech corpus research has shown that more frequent collocations are characterized by shorter word duration (Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999) and that more frequent, predictable, or/and redundant words have shorter durations as well as reduced pitch and amplitude range (Shattuck-Hufnagel & Turk, 1996; see also Bybee & Scheibman, 1996 for the relationship between collocational strength and vowel reduction). Thus, we hypothesize that robust prosodic processing may help L2 learners infer from every instance of input not only which parts of speech could be chunked together, but also whether they serve as frequent collocational units.

At the morphosyntactic level, it has been shown that the linguistic features with which L2 learners have the most difficulty tend to have fewer phonemes, and low syllabicity and sonority (e.g., Goldschneider & DeKeyser, 2001). According to the prosodic account of L2 grammar (Goat & White, 2019), the accurate encoding of prosodic cues is believed to be a necessary

AUDITORY PROCESSING & VOCABULARY

condition for the acquisition of complex morphology (e.g., inflection), syntax (e.g., word order), and semantics (e.g., articles). Thus, we hypothesize that individual differences in pitch, duration, and amplitude rise time may determine the extent to which learners can extract L2 morphosyntactic information from aural input and that those with more precise prosodic processing can demonstrate more advanced L2 morphosyntactic proficiency (e.g., Kachlicka et al., 2019; Saito et al., 2020).

At the global level, whereas L2 vocabulary knowledge is instrumental to global reading and listening skills, it has been shown that those who have attained highly advanced L2 listening and reading proficiency are likely to have greater working memory, attentional control, and phonological awareness (Vafaei & Suzuki, 2020; Wallace, 2020). Importantly, scholars have also demonstrated that auditory processing and cognitive abilities are interwoven with each other (Ahissar, Lubin, Putter-Katz, & Banai 2006; Snowling, Gooch, McArthur, & Hulme, 2018; Grube, Kumar, Cooper, Turton, & Griffiths, 2012). As such, auditory processing and memory abilities can simultaneously help learners hold aural information for a longer period of time, thereby making it available for more robust acoustic analyses (Zhang, Moore, Guiraud, Molloy, Yan, & Amitay, 2016).

Given that auditory processing is an important determinant of phonological aspects of L2 speech learning (e.g., Omote et al., 2017), and that the mechanisms underlying the individual differences in L2 lexical production development and attainment have remained open to investigation (Saito, 2015, 2020), the current study explored the relationship between a total of 100 late Polish-English bilinguals' profiles of auditory processing (pitch, duration, amplitude rise time), biographical backgrounds (length of immersion, musical training, age of arrival), and spoken vocabulary proficiency (appropriateness, richness). The following research question, followed by predictions, was formulated:

- Whether, to what degree and how does auditory processing relate to post-pubertal L2 learners' spoken vocabulary proficiency when biographical factors are controlled for?

According to the cross-sectional and longitudinal investigations, much vocabulary learning can be observed in richness (e.g., Salsbury et al., 2011), and an extensive period of regular and frequent L2 use may be needed to make a perceptible change in appropriateness (e.g., Saito, 2019 for 10+ years of immersion). As stated in the aptitude-acquisition hypothesis (Doughty, 2019), it is in such relatively difficult aspects (i.e., appropriateness rather than richness) where aptitude (including auditory processing) may play a key role in determining the extent to which certain L2 learners can attain advanced L2 lexicogrammatical proficiency (rich *and* accurate). In particular, more precise auditory processing (pitch, duration and amplitude in particular) may help L2 learners (a) segment aural input into words with the accurate use of lexical stress, (b) detect, internalize and use more frequent, strongly combined collocational chunks in a contextually appropriate manner, and (c) access perceptually non-salient morphological markers (fewer phonemes, low syllabicity and sonority; e.g., Goldschneider & DeKeyser, 2001 for regular past tense).

Method

Participants

The participants were 100 Polish residents in the UK whose pronunciation performance was assessed in the precursor project. The length of immersion varied widely from 0.1 to 19 years. The data collection (speech and auditory processing tests) was conducted with a researcher

AUDITORY PROCESSING & VOCABULARY

at a university in London. While the frequency of daily L2 use varied across different contexts (work, home, family), according to individual interviews, the participants reported that their main language of communication, at work and/or at home, was L2 English. After experiencing 6-15 years of English-as-a-Foreign-Language education in Poland, the participants arrived in the UK after puberty (age of arrival [AOA] > 17 years). None reported any hearing nor reading problems. All the biographical information is detailed in Table 1.

Table 1

Biographical Backgrounds of 100 Participants

	<i>M</i>	<i>SD</i>	<i>Range</i>	
			<i>Min</i>	<i>Max</i>
Chronological age	30.2 years	7.2	19	45
Age of arrival	22.4 years	3.6	17	36
Length of residence	7.3 years	5.1	0.1	19
Age of EFL education	9.9 years	3.4	4	16
Length of EFL education	9.5 years	3.5	1	20
Current L2 use @ work	80.1%	23.5	10	100
Current L2 use @ home	56.3%	37.5	0	100
Current L2 use @ social	62.5%	25.9	10	100

In the precursor research (Saito, Kachlicka, Sun, & Tierney, 2020a), the differential effects of age, experience, and auditory processing on the participants' *phonological* accuracy and fluency were measured. The analyses and findings were based on short speech samples (30 seconds per participant) elicited using a semi-structured speaking task (i.e., picture description), which would fall short of the length threshold for spoken lexicogrammar analyses. With a view of conducting robust analyses of spoken lexicogrammar, scholars have suggested 100+ words as a minimum length requirement (Koizumi & In'nami, 2012). To elicit sufficiently long spontaneous speech samples, and to capture appropriate use of vocabulary and grammar, longer speech samples from the same participants were elicited using a different task, i.e., oral interview.

Speaking Task

Free speech tasks, such as the oral interview task in this study, have been widely used in L2 vocabulary research (e.g., Crossley et al., 2015) and high-stakes speaking-ability tests (e.g., IELTS). First, participants were asked to talk about the following topic (i.e., *What was the hardest and toughest challenge in your life?*). After one minute of planning time, they spoke for roughly two minutes. Finally, the researcher asked a few follow-up questions in response to the content of their speech (for the materials used in the study, see **Supporting Information-A**). Compared to the highly structured task used in the precursor project, wherein participants focused on describing already provided information (picture narratives), the format of the interview task could be considered less structured, encouraging participants to produce longer and more complex speech while talking about more informal, familiar and personal topics with freedom (see Skehan, 1998).

To control for the effects of phonological quality on L2 analyses, all the recordings were transcribed and cleaned by removing filled pauses (e.g., "ah, eh, um") and fixing obvious mispronunciation problems (e.g., *life*, pronounced as *rife*, would still be spelled as *life*). The duration of the transcripts widely varied ($M = 503.1$ words, $Range = 106-1264$ words). Four researchers initially transcribed the same five speech samples (out of the entire dataset, 100

AUDITORY PROCESSING & VOCABULARY

speech samples) to compare their agreement. While their transcripts largely agreed with each other, they discussed any discrepancies and agreed on some transcription conventions (see **Supporting Information-B**). Afterwards, the remaining 95 samples were divided between the four researchers, each of whom individually transcribed 20-25 samples. Whenever they encountered ambiguous situations, they consulted with each other to ensure that they had consistently followed the agreed conventions.

Analyses of Appropriateness

To capture the multifaceted nature of appropriateness, three different approaches were adopted:

1. Holistic Appropriateness. To account for the potentially different degrees of error gravity on global comprehension and communicative adequacy, scholars have emphasized the importance of expert raters' holistic judgements (Foster & Wigglesworth, 2016). Following the training procedure in Saito (2019), a total of five linguistically trained raters were recruited to assess semantic and morphosyntactic dimensions of appropriateness.

Raters. The raters included three native speakers of English (2 from the UK, 1 from the USA) and two near-native speakers of English (1 from Estonia, 1 from Germany). All of them received several years of linguistics training at universities in London and had a significant amount of experience in L2 speech analyses as they regularly participated in empirical research projects of this kind. They reported high levels of familiarity with vocabulary use in British English and foreign accented English in the UK. As reported below, all the raters demonstrated relatively high inter-rater agreement (see below).

Procedure. The rating sessions took place individually under the supervision of a researcher. The raters first received definitions for the two different areas of appropriateness—(a) semantic and (b) morphosyntactic. For training scripts and onscreen labels, see **Supporting Information-C**. During the assessment, the samples were displayed on a computer screen in a randomized order using MATLAB software. For each token, the degree of appropriateness was assessed using a moving slider. If the slider was placed at the leftmost end of the continuum, labeled with a frowning face (indicating “non-targetlike”), the rating was recorded as 0. If the slider was placed at the rightmost end of the continuum, labeled with a smiley face (indicating “targetlike”), the rating was recorded as 1000. The scoring method was explicitly explained to the raters. None of them asked any questions. To avoid any confusion (as reported in some L2 assessment research using a numbered scale; Isaacs & Thomson, 2013), no numerical values were displayed on the screen.

To ensure the raters' understanding of the procedure, they evaluated three practice transcripts (not included in the main dataset) and explained/justified their decisions. For each response, the researcher gave feedback to ensure that the raters handled the three different categories without confusion. Finally, the raters moved onto the main dataset of 100 transcripts.

Reliability. In our pilot run, the length of a session turned out to be a problem as some transcripts were long (> 1000 words). To reduce rater fatigue, all transcripts were equally cut down to 250 words, except for several samples that were already less than 250 words. Each session lasted for approximately three hours (including training and practice), and the raters took a short break (10 minutes) halfway through. A Cronbach alpha analysis revealed that the five raters demonstrated relatively strong agreement for semantic appropriateness ($\alpha = .81$) and morphosyntactic appropriateness ($\alpha = .83$). According to the post-rating questionnaire, the raters reported that they not only understood but also handled the three rubrics through the judgment sessions without confusion ($M = 9$ out of “1 = “very difficult”, 9 = “very easy and

AUDITORY PROCESSING & VOCABULARY

comfortable”). The five raters’ scores were averaged to generate two scores for each transcript, quantifying its semantic and morphosyntactic appropriateness.

2. Local Appropriateness. Given that the task was designed to elicit the participants to use past tense while speaking, local morphosyntactic accuracy was operationalized via tallying the number of past-tense errors by the number of obligatory contexts per sample (for a similar approach, see Kourтали & Révész, 2020). The past tense in English was considered perceptually non-salient (Goldschneider & DeKeyser, 2001), semantically redundant and less intrusive to communicative success (VanPatten, 2002), and was reported to be difficult among many adult L2 learners of English (Ellis, Loewen, & Erlam, 2006). Two linguistically trained coders first analyzed 20 (out of 100) samples. The inter-coder agreement was relatively high ($r = .91$). The first coder completed the rest of the analyses. The participants’ accuracy ratio widely ranged ($M = 24.9\%$; $SD = 18.3$; $Range = 0-92\%$). Since the obligatory context analysis could be influenced by text length, we also calculated residual accuracy scores with the length factor statistically controlled for.

3. Collocational Appropriateness. Collocation is broadly defined as a meaningful combination of multiword expressions (Gablasova, Brezina, & McEnery, 2017), and found to serve as a primary determinant of humans’ intuitive of lexical appropriateness (Saito, 2020). To this end, two corpus-based association measures were used, Mutual Information (MI) bigram and trigram. Conceptually, MI indicates the strength of the partnership between two and three-word expressions, while controlling for the probability of random groupings of words. Collocations with higher MI scores consist of combinations of words which likely have a fewer number of partner words. These words likely exhibit greater coherence, more distinctive meaning and clearer discourse functions. To calculate MI, random co-occurrences of words were first estimated by dividing the number of any possible combinations within a fixed window size ($n = \pm 5$ words in TAALES) by the total number of tokens in the reference corpus (British National Corpus). Then, the frequency of collocations was divided by the frequency of random co-occurrence among the words, and then logarithmized.

Analyses of Richness

The multifaceted nature of richness was approached from three different perspectives via TAALES 2.0 (Kyle & Crossley, 2015):

1. Frequency. Word frequency refers to the extent to which less frequent and common words are used per sample. The index was calculated by dividing the total sum of frequency scores in reference to the British National Corpus by the number of all the words with frequency scores. In order to control for Zipfian effects in word frequency lists (higher-frequency words are more likely to be recycled), the raw scores were logarithmically transformed. Lower frequency scores indicate the use of less frequent words and more infrequent words, which is characteristic of more advanced L2 lexical proficiency (Crossley et al., 2009).

2. Range. Word range refers to the extent to which L2 speakers used more specific words which are narrowly used and observed in certain contexts and genres (rather than across diverse contexts). The index was calculated by dividing the total sum of range scores by the number of words in the texts with range scores. Like frequency, the raw scores were logarithmically transformed. Words with lower range scores indicate the use of more context-specific words (restricted to certain genres), which is considered as an index of more advanced L2 lexical proficiency (Kyle & Crossley, 2015).

3. Abstractness. Abstractness refers to the extent to which words that are less concrete, imageable and familiar are used per sample. In TAALES, native speakers’ perceived judgements

AUDITORY PROCESSING & VOCABULARY

of concreteness and imageability were stored for 4,000 content words based on the MRC psycholinguistics database (Coltheart, 1981). The average concreteness, imageability and familiarity scores were calculated for each transcript (0-1000 points). Thus, those who often use words with lower judgement scores (i.e., less concrete, imageable, and familiar words) could be considered to have more advanced L2 lexical proficiency (Salsbury et al., 2011).

Auditory Processing Measures

Following the methodology widely used in cognitive psychology (e.g., Surprenant & Watson, 2001), participants' domain-general perception ability was assessed using a battery of psychophysical assessments. The materials used for the current study were developed in precursor studies (e.g., Kachlicka et al., 2019). As reviewed earlier, the acoustic dimensions relevant to L2 vocabulary acquisition were assumed to involve participants' thresholds for discrimination of pitch, duration, and amplitude. For each test, three complex tone stimuli were presented, with either the first or the third sounding different from the other two. Participants indicated which sound was different by either pressing the number '1' or '3' on a keyboard. An adaptive three-alternative forced-choice procedure was used, such that the difficulty of the task would decrease after every incorrect response and increase after every third correct response. The program continued until eight reversals had been reached, i.e., incorrect answers after a string of successes or correct answers after a string of failures.

Stimulus. For each test, 100 continuous synthesized stimuli (500 ms in length) were created via custom MATLAB scripts. They differed at 100 steps along the target acoustic dimension (Levels 1-100). A total of 100 four-harmonic complex tones were created with F0 set to 330 Hz and the amplitude of each harmonic set to 40 dB. The target acoustic dimension for each test varied by a step of 0.3 Hz in F0 (330.3-360 Hz), 2.5 ms in duration (252.5-500 ms), and 1.22 ms in amplitude rise time (178-300 ms), respectively.

Procedure. When three different tones were presented with an inter-stimulus interval of 0.5 s, the participants were asked to choose which of the three tones differed from the other two by pressing the number "1" or "3." Based on Levitt's (1971) adaptive threshold procedure, the level of difficulty changed from trial to trial according to participants' performance. The initial difficulty, i.e. the level of the target stimulus, was set to level 50. When three correct responses were made in a row, the difference became smaller by a degree of 10 steps (more difficult). When their response was incorrect, the difference became wider by a degree of 10 steps (easier). The reverse happened 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 decreased (more difficult) from 10 to 5, and then from 5 to 1 after the second reversal. The tests stopped either after 70 trials or eight reversals. For participants' auditory processing scores, the stimulus levels after the third reversal were averaged. Since the scores indicate how small of a difference participants can perceive, lower scores indicate more precise auditory processing scores.

Reliability. To check the reliability of the auditory processing tests, a follow-up project was conducted with 30 English users with diverse experience and proficiency levels (not included in the current study). They took a range of auditory processing tests (including pitch, duration, and amplitude discrimination) twice with an interval of one day. The results of Spearman correlation analyses demonstrated small-to-medium strength for the individual tests ($r = .632$ for pitch, $.333$ for duration, $.737$ for risetime). As for the composite auditory processing scores (averaging pitch, duration, and rise time discrimination), the reliability ($r = .720$) could be considered satisfactory and comparable to similar research (e.g., $r = 0.75$ in Raz, Willerman, &

AUDITORY PROCESSING & VOCABULARY

Yama, 1987). The results suggest that although using individual test scores may result in low reliability (e.g., duration discrimination), composite test scores may serve as a more reliable proxy of one's auditory precision (for methodological details, see **Brief Report** in Saito, Sun, & Tierney, 2020b).

Composite Scores. The descriptive results of pitch, duration and amplitude rise time discrimination test scores were summarized in **Supporting Information-D**. Since the data significantly differed from the normal distribution ($p < .01$), their raw scores were transformed via a log10 function. To calculate participants' overall prosodic encoding abilities, their raw scores were standardized, and averaged. According to the results of the normality test (Kolmogorov-Smirnov) the resulting averaged scores were comparable to the normal distribution ($p > .05$), and thus were used for the subsequent analyses as a composite index of participants' auditory processing of prosodic cues. Lower factor scores indicate more precise encoding of pitch, duration, and amplitude information.

For the sound stimuli used in the auditory processing tests (duration, pitch, risetime), see the team's website (www.sla-speech-tools.com, under construction).

Results

First, we present the results of preliminary analyses to examine what characterizes spoken L2 vocabulary proficiency among 100 Polish-English bilinguals relative to L1 counterparts. Second, we show the results of factor analyses to explore what underlies spoken L2 vocabulary proficiency (which we analyzed via 11 outcome measures) and auditory processing abilities (which we analyzed via 3 outcome measures). Subsequently, we present the results of multiple regression analyses to probe how a range of predictor variables related to experience and auditory processing are uniquely associated with various dimensions of participants' L2 lexical proficiency.

Spoken L2 vs. L1 Vocabulary Proficiency

The descriptive results of the 11 vocabulary measures are summarized in Table 2. To examine what characterizes spoken L2 vocabulary proficiency, a set of 95% confidence interval analyses were performed to check the extent to which Polish-English bilinguals' performance overlapped with (or deviated from) that of L1 speakers. In the prior project (Saito, in press), a total of 10 monolingual speakers of English (born and raised in the English-speaking areas of Canada) completed the same oral interview task. The results indicated two overall patterns: (a) some Polish-English bilinguals reached nativelike proficiency in terms of richness (overlaps in 95% intervals in all measures); and (b) appropriateness could be considered as a relatively difficult dimension of spoken L2 vocabulary proficiency as L2 speakers' proficiency was significantly distinguishable from the native benchmark in five of six measures (i.e., lexical, morphosyntactic, and collocational accuracy).

Table 2

Descriptive Summary of Spoken L2 Vocabulary Proficiency Relative to Native Benchmark

	A. Polish-English bilinguals				B. Native benchmark ^a			
	<i>M</i>	<i>SD</i>	95% CI		<i>M</i>	<i>SD</i>	95% CI	
			<i>Min</i>	<i>Max</i>			<i>Min</i>	<i>Max</i>
<u>A. Richness</u>								
Frequency	2.57	0.10	2.55	2.59	2.74	0.23	2.66	2.82
Range	-0.60	0.07	-0.61	-0.58	-0.57	0.04	-0.58	-0.56
Familiarity	585	4	584	586	583	2	581	585
Concreteness	325	13	322	327	338	9	331	345
Imageability	363	13	361	366	366	11	358	374
<u>B. Appropriateness</u>								
Lexical judgements (1000 points)	633	138	606	660	841	82	782	907
Morphosyntactic judgements (1000 points)	576	140	547	603	799	88	735	862
Past tense error ratio (raw, %)	24.93	18.32	21.30	28.57	0	n.a.	n.a.	n.a.
Past tense error ratio (adjusted)	0	0.99	-0.19	0.19	0	n.a.	n.a.	n.a.
Spoken bigram MI	9.48	0.26	9.43	9.53	9.85	0.21	9.70	10.01
Spoken trigram MI	8.46	0.24	8.41	8.51	8.60	0.22	8.44	8.76

Note. ^aThe native control data derives from Saito (in press).

Constructs of Spoken L2 Vocabulary Proficiency

To check whether and to what degree they were assumed to tap into the constructs that we intended to measure ($n = 6$ for appropriateness, $n = 5$ for richness), they were submitted to an exploratory factor analysis with Varimax rotation. The factorability of the entire dataset was considered adequate according to Bartlett's test of sphericity ($\chi^2 = 1506.938$, $p < .001$) and the Kaiser-Meyer-Olkin measure of sampling adequacy (.659). Using the standard of an eigenvalue beyond 1.0, a five-factor solution was suggested, accounting for 90.713% of the variance in the outcomes of the auditory processing measures.

In terms of factor loadings, 0.6 was used as a cut-off point in line with Hair, Black, Babin, Anderson, and Tatham's (1998) recommendation for factor analyses of relatively small sample size ($n < 100$). In light of the grouping patterns in Table 3, Factor 1 was labeled as "Holistic Accuracy" as it clustered both of the appropriateness judgement scores; Factor 2 was labeled as "Breadth" as it corresponded the use of infrequent, context-specific and unfamiliar words on a surface level; Factor 3 was labeled as "Local Accuracy"; Factor 4 was labeled as "Abstractness" as it clustered the MRC psycholinguistics database of word concreteness and imageability; and Factor 5 was labeled as "Collocational Accuracy" as it included both corpus-based n -gram measures. According to the results of the Kolmogorov-Smirnov tests, the distribution of the resulting factor scores was not significantly different from the normal distribution ($p > .05$), and thus the scores were used for the subsequent analyses without transformation.

AUDITORY PROCESSING & VOCABULARY

Table 3*Summary of a Five-Factor Solution Based on a Factor Analysis of Spoken L2 Lexicogrammar Proficiency*

	Factor 1 ^a (Holistic Accuracy)	Factor 2 (Breadth)	Factor 3 (Local Accuracy)	Factor 4 (Abstractness)	Factor 5 ^a (Collocational Accuracy)
<u>Richness</u>					
Frequency	-.044	.942	.003	-.235	.020
Range	.041	.941	.019	-.226	.023
Familiarity	-.165	.722	.384	.173	-.070
Concreteness	-.161	-.121	.071	.938	-.136
Imageability	-.088	-.173	-.134	.940	-.098
<u>Appropriateness</u>					
Lexical judgements	.959	-.062	-.129	-.121	.137
Morphosyntactic judgements	.945	-.045	-.150	-.127	.196
Past tense error ratio (raw)	-.129	.100	.977	-.027	-.080
Past tense error ratio (adjusted)	-.133	.100	.976	-.047	-.079
Spoken bigram MI	.180	.012	-.091	-.218	.810
Spoken trigram MI	.118	-.017	-.057	-.012	.887

Note. ^a the direction of the factor scores was reversed to proxy what the original scores indicate (more accurate, more collocational)

AUDITORY PROCESSING & VOCABULARY

Roles of Experience and Auditory Processing in Spoken L2 Vocabulary

To examine the relative weights of the biographical and auditory processing factors in the outcomes of spoken L2 vocabulary proficiency, a set of stepwise multiple regression analyses were conducted on each proficiency dimension as per a set of predictors related to auditory processing and experience. To avoid multicollinearity problems, the composite auditory processing scores (pitch, duration, and risetime discrimination) were used as a global index of Auditory Processing. Four experience factors were included as they were extensively discussed in the existing literature as crucial affecting factors L2 speech acquisition (Flege, 2018 for Age of Arrival; Saito, 2015 for Length of Residence; Flege & Liu, 2001 for Current L2 Use; Muñoz, 2014 for Length of EFL).¹ The mechanisms underlying L2 speech learning are said to differ between the early and later phase of immersion (DeKeyser, 2013). To this end, five interaction terms were included to see whether and to what degree the five predictors differentially related to the L2 vocabulary proficiency among two different groups of L2 learners; dummy codes (1, 2) were given to interlanguage learner group ($n = 50$; LOR = 0.1-5 years) and ultimate attainer group ($n = 50$; LOR = 6+ years). Finally, given that the length of participants' speech widely varied (106-1264 words), this variable was also entered as a covariate. For each of the vocabulary proficiency dimensions (Holistic Accuracy, Local Accuracy, Collocational Accuracy, Breadth, Abstractness), the following model was constructed:

- Vocabulary Proficiency = Auditory Processing + Age of Arrival + Length of Residence + Current L2 Use + Length of EFL + Length of Speech + Auditory Processing \times Group + Age of Arrival \times Group + Length of Residence \times Group + Current L2 Use \times Group + EFL \times Group

Model selection was conducted via SPSS based on the results of F tests. Backward elimination was chosen. After all the independent variables were entered, the largest probability of F was removed at each step (using $p = .10$ as a benchmark). The selection was completed when no variables were eligible for elimination. The details of the model building processes for each vocabulary domain were found in **Supporting Information-E**.

The final models were summarized in Table 4. The results generally showed that L2 accuracy was primarily predicted by auditory processing factors (composite prosodic processing scores), and secondarily by biographical factors (Length of Residence, Age of Arrival). More specifically, the link between auditory processing and acquisition was weak in local accuracy (related to the use of past tense) relative to holistic and collocational accuracy (related to vocabulary use in general). The roles of biographical factors uniquely related to different types of accuracy. Holistic accuracy was tied to length of residence, and local accuracy was associated with age of arrival. Interestingly, all the interaction effects were excluded in the final models in all instances, suggesting that the findings were generalizable across different stages of L2 acquisition (LOR = 0.1 to 40 years). None of the models of the richness factors (breadth, abstractness) reached statistical significance ($p > .05$). No clear sign of multi-collinearity was found in any contexts (Variance Inflation Factor < 1.231).²

¹ Six participants did not report the length of foreign language education prior to their arrival in the UK. Their missing values were replaced with average (i.e., 9.5 years). Not surprisingly, neither of the age-related variables (Chronological Age, Age of EFL) was significantly correlated with any aspects of L2 vocabulary proficiency attainment ($p > .05$). This corresponds to the existing research evidence that what matters for L2 speech acquisition is age of arrival rather than chronological age (e.g., Flege, 2018), and the length of EFL rather than age of learning (e.g., Muñoz, 2014).

² The low Variance Inflation Factor (< 1.231) suggest that participants' biographical backgrounds and auditory processing abilities were relatively independent with each other at least within the current dataset. In cognitive psychology, it has been shown that auditory processing is susceptible

AUDITORY PROCESSING & VOCABULARY

Table 4*Summary of Stepwise Multiple Regression Models Featuring Only Significant Predictors of Spoken L2 Vocabulary Proficiency*

Predicted variables	<i>R</i>	<i>Adjusted R²</i>	Change in <i>R²</i>	Predictors	Standardized β	<i>t</i>	<i>p</i>
Holistic Accuracy	.346	.111	.111	Auditory Processing	-.339	-3.717	< .001
	.437	.175	.064	Length of Residence	.267	2.9240	.004
Local Accuracy ^a	.262	.059	.059	Age of Arrival	.238	2.454	.016
	.324	.087	.028	Auditory Processing	.193	1.991	.049
Collocational Accuracy	.265	.070	.070	Auditory Processing	-.265	-2.717	.008
Breadth							<i>n.s.</i>
Abstractness							<i>n.s.</i>

Note. ^aLower scores indicate lower error ratio (more accurate use of past tense)

Discussion

Drawing on the auditory deficit theory in L1 acquisition, there is an emerging hypothesis that individual differences in experience, auditory processing, and L2 acquisition are interwoven (Mueller et al., 2012). According to the precursor research, auditory processing is an important determinant of segmental and suprasegmental accuracy (rather than fluency) aspects of L2 speech, even when all the biographical factors (age, immersion experience, music training) are controlled for (Kachlicka et al., 2019; Saito et al. in press). To further scrutinize the generalizability of the topic to higher-order dimensions of post-pubertal L2 speech acquisition, we aimed to examine the effects of auditory processing in spoken L2 vocabulary development and attainment among a total of 100 late Polish-English bilinguals in the UK.

According to the results of the statistical analyses, L2 learners who attained more advanced L2 vocabulary proficiency had not only more relevant experience (extensive immersion and earlier age of arrival), but also more precise auditory processing ability. As predicted earlier, our findings here generally align with the view that one's ability to track individual dimensions of prosodic information (i.e., pitch, duration, and amplitude) serves as a key driving force for detecting lexical and syntactic boundaries (de Pijper & Sanderman, 1994). Thus, it is possible that with more precise prosodic processing abilities, learners can better represent, encode, and segment ambient input into lexical and syntactic units, resulting in the development of more robust phonological and morphosyntactic knowledge (Jiang, 2000; Best & Tyler, 2007). Additionally, more precise auditory processing abilities are linked to greater phonological awareness and executive function, which in turn facilitates L2 reading and listening complementarily (Linck et al., 2013). Finally, those with more precise sound timing may detect more closely and frequently used multiword units, as they are delivered faster than other less common and less predictable combinations of words (Gregory et al., 1999).

Importantly, auditory processing could be fundamental, especially concerning appropriateness rather than breadth and abstractness dimensions of L2 lexicogrammar development. This is arguably because the former dimensions (appropriateness) are claimed to be more difficult than the latter dimensions (breadth, abstractness). As shown in the current study (the results of the benchmark analyses), there was considerably larger distance between L1 and L2 speakers in appropriateness than breadth and abstractness. The development of accuracy has been found to take place over a great deal of immersion experience at lexical (Saito, 2019) and morphosyntactic levels (Bartning et al., 2012). According to the aptitude-acquisition view, it is the *relatively difficult* L2 learning aspects that are subject to a substantial amount of L2 experience, and susceptible to the effects of individual differences in aptitude (Doughty, 2019). In terms of the breadth and abstractness aspects of spoken L2 vocabulary proficiency, the participants were comparable with each other regardless of experience and auditory profiles. This is arguably because many L2 learners' vocabulary use could be sufficiently abstract even without much immersive experience (see Saito, 2019; Salsbury et al., 2011).

While the facilitative role of auditory processing is germane to higher-level linguistic skills to some degree, such as the *production* of L2 *vocabulary*, it is also important to remember that the outcomes of spoken L2 vocabulary development are moderately related to auditory processing. In fact, the strength of the audition-acquisition link could be considered small (e.g., $r = -.346$ for Holistic Accuracy in Table 4). In prior research, the predictive power of auditory perception appeared to be more clearly observed in lower-order linguistic skills which directly involve auditory information, such as segmental and suprasegmental perception (e.g., Kachlicka et al. for $r = -.6$) and production (e.g., Saito et al., in press-a for $r = -.4$ to $-.5$). Therefore, it

AUDITORY PROCESSING & VOCABULARY

would be intriguing to further examine whether and to what degree other cognitive measures may explain the remaining variance in spoken L2 vocabulary acquisition. Such potential predictors include working memory (Martin & Ellis, 2012), selective attention (Nicolay & Poncelet, 2013), and foreign language aptitude (Li, 2016).

Limitations

Given that the current study took an exploratory approach to delving into the role of auditory processing in spoken L2 vocabulary development, there are several methodological limitations that future studies should further remedy and expand. First, all the findings were based on the cross-sectional analyses of 100 late Polish-English bilinguals. To further examine the causal relationship between auditory processing, experience, and L2 speech learning, it is necessary to conduct a longitudinal investigation. For example, future studies should explore the variance in phonological and lexical aspects of L2 proficiency in participants with various auditory processing profiles over a certain period of training (Chandrasekaran et al., 2010) and immersion (Sun et al., in press).

Secondly, participants' auditory processing was analyzed via the psychoacoustic tests. However, it has been argued that the test format (A×B discrimination) may not only reflect participants' auditory precision, but also involve a range of cognitive abilities, such as attentional control (Snowling et al., 2018). To control for the separate effects of perceptual and cognitive individual differences, future studies should adopt both auditory processing and executive function tests (cf. Saito et al., forthcoming for the relationship between memory, auditory processing, and L2 speech learning).

Thirdly, whereas participants' spoken vocabulary proficiency was elicited from a single task condition (oral interview), it has been shown that L2 learners' speech performance is susceptible to change as per task conditions (see Ellis, 2009 for an overview on task effects on appropriateness, richness, and fluency). The findings of the current investigation need to be replicated using multiple tasks differing in terms of the timing and length of planning time (Ahmadian & Tavakoli, 2011), and the degree of structural complexity (Foster & Tavakoli, 2009), and conceptualization (Saito, forthcoming).

Fourthly, the generalizability of the findings (i.e., prosodic processing vs. spoken L2 vocabulary) needs to be tested for diverse L1-L2 pairings. Although we argued that prosodic acuity matters for L2 vocabulary acquisition due to its relevance to word segmentation, it is important to note that the relative weights of prosodic cues may be highly language-specific. For example, it would be interesting to replicate the findings in L2 French speakers who use stress to parse linguistic units at sentence but not word-level (e.g., Dupoux, Pallier, Sebastian, Mehler, 1997 for the cross-linguistic differences in word and sentence stress assignment and its impact on tone deafness)

Finally, whereas the current study indicated a potential link between auditory processing and the acquisition of L2 English past tense, it needs to be acknowledged that little is known about how auditory processing is related to L2 morphosyntax at a fine-grained level. In the field of second language acquisition, a growing amount of attention has been directed towards detangling how phonology interfaces with various areas of grammar (for a comprehensive summary of the prosodic account of L2 behaviors, see Goad & White, 2019). Given that Goldschneider and DeKeyser (2001) presented a plausible hierarchical framework for the perceptual acuity and morphosyntactic learning, one promising enquiry concerns the extent to which L2 learners with different levels of auditory processing abilities master L2 morphosyntax with different levels of perceptual salience (e.g., sonority). There is a possibility that individual

AUDITORY PROCESSING & VOCABULARY

differences in auditory processing (a core component of phonology) may be integral to the acquisition of grammar which interfaces lexicon, morphology, and syntax (e.g., inflection; Austin, Chang, Kim, & Daly, 2021) and semantics and discourse (e.g., articles; Demuth & McCullough, 2009).

Conclusion

All in all, our findings concur with the mounting empirical evidence that auditory processing is a determinant of how much L2 learners can benefit from immersion experience, resulting in more advanced outcomes (Saito & Tierney, forthcoming), and the theoretical view that the same driving faculty of L1 acquisition (i.e., auditory processing) is tied to every stage of L2 acquisition throughout an individual's lifetime (Flege, 2018). Building on the prior work (e.g., Saito et al., 2020 for segmental and suprasegmental production), we add that such audition effects are more clearly observed not only in the acquisition of relatively difficult features (accuracy rather than fluency, breadth, and abstractness), but also in the dimensions more closely related to the speech signal (phonology rather than lexicogrammar). Interestingly, hearing research has shown that auditory deficits can be remedied via focused training (e.g., Carcagno & Plack, 2011 for 10 hours of pitch discrimination training). In light of the significant relationship between auditory processing and L2 speech learning (though its strength varies across different linguistic dimensions), our study hints at the possibility that auditory training may help L2 learners amplify and optimize their acquisition processes, if it is provided at the same time that they engage in a certain period of immersive experience in a target language-speaking country (e.g., study abroad), or when they receive intensive or/and meaning-oriented speech training (e.g., Barriuso, & Hayes-Harb, 2018 for high-variability phonetic training; Lee & Lyster, 2016 for focus on form; Lim & Holt, 2011 for incidental video-gaming; Mora & Levkina, 2017 for task-based pronunciation teaching; Shao, Saito, & Tierney, forthcoming for repetition-based training)—i.e., a new interdisciplinary direction that linguistics, psychology, education, and hearing researchers can further pursue together.

AUDITORY PROCESSING & VOCABULARY

References

- Ahissar, M., Lubin, Y., Putter-Katz, H., & Banai, K. (2006). Dyslexia and the failure to form a perceptual anchor. *Nature Neuroscience*, *9*, 1558-1564.
- Ahissar, M., Protopapas, A., Reid, M., & Merzenich, M. M. (2000). Auditory processing parallels reading abilities in adults. *Proceedings of the National Academy of Sciences*, *97*, 6832-6837.
- Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of experimental child psychology*, *83*, 111-130.
- Austin, G., Chang, H., Kim, N., & Daly, E. (2021). Prosodic transfer across constructions and domains in L2 inflectional morphology. *Linguistic Approaches to Bilingualism*.
- Barriuso, T. A., & Hayes-Harb, R. (2018). High Variability Phonetic Training as a Bridge from Research to Practice. *CATESOL Journal*, *30*, 177-194.
- Bartning, I., Lundell, F. F., & Hancock, V. (2012). On the role of linguistic contextual factors for morphosyntactic stabilization in high-level L2 French. *Studies in Second Language Acquisition*, *34*, 243-267.
- Bavin, E. L., Grayden, D. B., Scott, K., & Stefanakis, T. (2010). Testing Auditory Processing Skills and their Associations with Language in 4-5-year-olds. *Language and Speech*, *53*, 31-47.
- Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception. *Language experience in second language speech learning: In honour of James Emil Flege*, 13-34.
- 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*, 29-40.
- Bybee, J., & Scheibman, J. (1999). The effect of usage on degree of constituency: The reduction of don't in American English. *Linguistics*, *37*, 575-596.
- Carcagno, S., & Plack, C. J. (2011). Subcortical plasticity following perceptual learning in a pitch discrimination task. *Journal of the Association for Research in Otolaryngology*, *12*, 89-100.
- Casini, L., Pech-Georgel, C., & Ziegler, J. C. (2018). It's about time: revisiting temporal processing deficits in dyslexia. *Developmental Science*, *21*, e12530.
- Chandrasekaran, B., Sampath, P. D., & Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*, 456-465.
- Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497-505.
- Crossley, S. A., Salsbury, T., & McNamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, *36*(5), 570-590.
- Crossley, S. A., Skalicky, S., Kyle, K., & Monteiro, K. (2019). Absolute frequency effects in second language lexical acquisition. *Studies in Second Language Acquisition*, *41*(4), 721-744.
- DeKeyser, R. M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning*, *63*, 52-67.
- Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*(2), 163-185.
- Doughty, C. J. (2019). Cognitive language aptitude. *Language Learning*, *69*, 101-126.

AUDITORY PROCESSING & VOCABULARY

- Douglas, S., & Willatts, P. (1994). The relationship between musical ability and literacy skills. *Journal of Research in Reading*, *17*, 99-107.
- Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*, 143-188.
- Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied linguistics*, *30*, 474-509.
- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*(2), 339-368.
- 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*, 343-356.
- Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*, 399-423.
- Flege, J. E. (2018). It's input that matters most, not age. *Bilingualism: Language and Cognition*, *21*(5), 919-920.
- Flege, J. & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*, 527-552.
- Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language learning*, *59*, 866-896.
- Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics*, *36*, 98-116.
- Foxton, J. M., Talcott, J. B., Witton, C., Brace, H., McIntyre, F., & Griffiths, T. D. (2003). Reading skills are related to global, but not local, acoustic pattern perception. *Nature Neuroscience*, *6*, 343-344.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language learning*, *67*, 155-179.
- Goad, H., & White, L. (2019). Prosodic effects on L2 grammars. *Linguistic Approaches to Bilingualism*, *9*(6), 769-808.
- Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language learning*, *51*(1), 1-50.
- Goswami, U. (2015). Sensory theories of developmental dyslexia: three challenges for research. *Nature Reviews Neuroscience*, *16*, 43-54.
- Goswami, U., Wang, H. L. S., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2011). Language-universal sensory deficits in developmental dyslexia: English, Spanish, and Chinese. *Journal of Cognitive Neuroscience*, *23*, 325-337.
- Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. In *Chicago Linguistic Society* (Vol. 35, pp. 151-166).
- Grube, M., Kumar, S., Cooper, F. E., Turton, S., & Griffiths, T. D. (2012). Auditory sequence analysis and phonological skill. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 4496-4504.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis* (Vol. 5, No. 3, pp. 207-219). Upper Saddle River, NJ: Prentice Hall.

AUDITORY PROCESSING & VOCABULARY

- Hopp, H., & Schmid, M. S. (2013). Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied psycholinguistics*, 34(2), 361-394.
- Hornickel, J., & Kraus, N. (2013). Unstable representation of sound: a biological marker of dyslexia. *Journal of Neuroscience*, 33, 3500-3504.
- Hyltenstam, K. (1988). Lexical characteristics of near-native second-language learners of Swedish. *Journal of Multilingual & Multicultural Development*, 9, 67-84.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1), B47-B57.
- Jasmin, K., Dick, F., Holt, L., & Tierney, A. T. (2019). Tailored perception: Individuals' speech and music perception strategies fit their perceptual abilities. *Journal of Experimental Psychology: General*, 149(5), 914-934.
- Jasmin, K., Sun, H., & Tierney, A. T. (2020). Effects of language experience on domain-general perceptual strategies. *Cognition*, 206(104481), 1-14.
- Jasmin, K., Dick, F., Holt, L. L., & Tierney, A. (2020). Tailored perception: Individuals' speech and music perception strategies fit their perceptual abilities. *Journal of Experimental Psychology: General*, 149, 914.
- Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, 24(1), 131-161.
- Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21, 47-77.
- Joanisse, M. F., & Seidenberg, M. S. (1998). Specific language impairment: A deficit in grammar or processing? *Trends in Cognitive Sciences*, 2, 240-247.
- 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.
- Kim, D., Clayards, M., & Kong, E. J. (2020). Individual differences in perceptual adaptation to unfamiliar phonetic categories. *Journal of Phonetics*, 81, 100984.
- Koizumi, R. (2012). Relationships Between Text Length and Lexical Diversity Measures: Can We Use Short Texts of Less than 100 Tokens? *Vocabulary Learning and Instruction*, 1(1), 60-70.
- Koizumi, R., & In'nami, Y. (2012). Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System*, 40(4), 554-564.
- Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, 59, 40-57.
- Krizman, J., Slater, J., Skoe, E., Marian, V., & Kraus, N. (2015). Neural processing of speech in children is influenced by extent of bilingual experience. *Neuroscience letters*, 585, 48-53.
- Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, 97, 11850-11857.
- Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4), 757-786.
- Lamb, S. J., & Gregory, A. H. (1993). The relationship between music and reading in beginning readers. *Educational psychology*, 13, 19-27.
- Lee, A. H., & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, 38, 35.

AUDITORY PROCESSING & VOCABULARY

- Leppänen, P. H. T., Hämäläinen, J. A., Guttorm, T. K., Eklund, K. M., Salminen, H., Tanskanen, A., ... & Lyytinen, H. (2012). Infant brain responses associated with reading-related skills before school and at school age. *Neurophysiologie Clinique/Clinical Neurophysiology*, 42(1-2), 35-41.
- Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, 38, 801-842.
- Lim, S. J., & Holt, L. L. (2011). Learning foreign sounds in an Alien World: Videogame training improves non-native speech categorization. *Cognitive science*, 35, 1390-1405.
- Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., ... & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language learning*, 63, 530-566.
- Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, 34, 379-413.
- Mora, J. C., & Levkina, M. (2017). Task-based pronunciation teaching and research: Key issues and future directions. *Studies in Second Language Acquisition*, 39, 381-399.
- Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, 35, 463-482.
- 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.
- Nicolay, A. C., & Poncet, M. (2013). Cognitive advantage in children enrolled in a second-language immersion elementary school program for three years. *Bilingualism: Language and Cognition*, 16, 597.
- Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological review*, 115, 357.
- Omote, A., Jasmin, K., & Tierney, A. (2017). Successful non-native speech perception is linked to frequency following response phase consistency. *Cortex*, 93, 146-154.
- Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, 38(1), 97-130.
- Raz, N., Willerman, L., & Yama, M. (1987). On sense and senses: Intelligence and auditory information processing. *Personality and Individual Differences*, 8(2), 201-210.
- Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything? *Journal of Phonetics*, 31, 509-527.
- Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., & Kraus, N. (2008). Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clinical Neurophysiology*, 119, 1720-1731.
- Saito, K. (2015). The role of age of acquisition in late second language oral proficiency attainment. *Studies in Second Language Acquisition*, 37, 713-743.
- Saito, K. (2019). To what extent does long-term foreign language education help improve spoken second language lexical proficiency? *TESOL Quarterly*, 53(1), 82-107.
- Saito, K. (2020). Multi-or single-word units? The role of collocation use in comprehensible and contextually appropriate second language speech. *Language Learning*, 70, 548-588.
- Saito, K. (forthcoming). Age effects in spoken second language vocabulary attainment in adulthood. *Language and Speech*.
- Saito, K., Cui, H., Suzukida, Y., Suzuki, Y., Jeong, Hyeonjeong., Revesz, m A., & Sugiura, M. (forthcoming). Does domain-general auditory processing uniquely explain the outcomes

AUDITORY PROCESSING & VOCABULARY

- of second language speech acquisition, even once cognitive and demographic variables are accounted for?
- Saito, K., Kachlicka, M., Sun, H., & Tierney, A. (2020a). Domain-general auditory processing as an anchor of post-pubertal second language pronunciation learning: Behavioural and neurophysiological investigations of perceptual acuity, age, experience, development, and attainment. *Journal of Memory and Language*, *115*, 104168.
- Saito, K., Sun, H., & Tierney, A. T. (2020b). Brief report: Test-retest reliability of explicit auditory processing measures. *bioRxiv*. DOI: <https://doi.org/10.1101/2020.06.12.149484>
- Saito, K., Sun, H., & Tierney, A. (in press-a). Domain-general auditory processing as a perceptual-cognitive anchor of L2 pronunciation learning in adulthood: A longitudinal study. *Applied Psycholinguistics*.
- Saito, K., Sun, H., Kachlicka, M., Alayo, J. R. C., Nakata, T., & Tierney, A. T. T. (2020). Domain-general auditory processing explains multiple dimensions of L2 acquisition in adulthood. *Studies in Second Language Acquisition*, 1-30.
- Saito, K., Suzukida, Y., Tran, M. & Tierney, A. T. (2021). Domain-general auditory processing partially explains second language speech learning in classroom settings: A review and generalization study. *Language Learning*, 1-47.
- Saito, K., & Tierney, A. (forthcoming). Domain-general auditory processing as a conceptual and measurement framework for second language speech learning aptitude.
- Salsbury, T., Crossley, S. A., & McNamara, D. S. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research*, *27*(3), 343-360.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.
- Shao, Y., Saito, K., & Tierney, A. (forthcoming). How does having a good ear promote instructed second language pronunciation development? Roles of domain-general auditory processing in choral repetition training.
- Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*, 193-247.
- Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, *132*, EL95-EL101.
- Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2015). Stability and plasticity of auditory brainstem function across the lifespan. *Cerebral Cortex*, *25*, 1415-1426.
- 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.
- 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.
- Talcott, J. B., Witton, C., McLean, M. F., Hansen, P. C., Rees, A., Green, G. G., & Stein, J. F. (2000). Dynamic sensory sensitivity and children's word decoding skills. *Proceedings of the National Academy of Sciences*, *97*, 2952-2957.
- Tallal, P. (2004). Improving language and literacy is a matter of time. *Nature Reviews Neuroscience*, *5*(9), 721-728.

AUDITORY PROCESSING & VOCABULARY

- Tierney, A., Gomez, J. C., Fedele, O., & Kirkham, N. Z. (2021). Reading ability in children relates to rhythm perception across modalities. *Journal of Experimental Child Psychology*, *210*, 105196.
- Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*, 434-464.
- 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.
- Vafaei, P., & Suzuki, Y. (2020). The relative significance of syntactic knowledge and vocabulary knowledge in second language listening ability. *Studies in Second Language Acquisition*, *42*, 383-410.
- VanPatten, B. (2002). Processing instruction: An update. *Language learning*, *52*, 755-803.
- Wallace, M. P. (2020). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning*.
- Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.
- Won, J. H., Tremblay, K., Clinard, C. G., Wright, R. A., Sagi, E., & Svirsky, M. (2016). The neural encoding of formant frequencies contributing to vowel identification in normal-hearing listeners. *The Journal of the Acoustical Society of America*, *139*, 1-11.
- Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*, 565.
- Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and Aging*, *27*, 410.
- Zhang, Y. X., Moore, D. R., Guiraud, J., Molloy, K., Yan, T. T., & Amitay, S. (2016). Auditory discrimination learning: role of working memory. *PLoS one*, *11*, e0147320.