



EMPIRICAL STUDY

Domain-General Auditory Processing Partially Explains Second Language Speech Learning in Classroom Settings: A Review and Generalization Study

Kazuya Saito ^{a,b}, Yui Suzukida ^a, Mai Tran,^c
and Adam Tierney^c

^aUniversity College London, ^bTohoku, University, Japan, and ^cBirkbeck, University of London

Abstract: To date, a growing number of studies have shown that domain-general auditory processing, which prior work has linked to L1 acquisition, could explain various dimensions of naturalistic L2 speech proficiency. The current study examined the generalizability of this topic to L2 speech learning in classroom settings. The spontaneous speech samples of 39 Vietnamese English-as-a-foreign-language learners were analyzed for fluent and accurate use of pronunciation and lexicogrammar and linked to a range of variables in their auditory processing profiles. The results identified moderate-to-strong correlations between the participants' accurate use of lexicogrammar and audio-motor sequence integration scores (i.e., the ability to reproduce melodic/rhythmic information). However, the relationship between phonological proficiency and auditory acuity (i.e., the ability to encode acoustic details of sounds) was nonsignificant. Although the findings support the audition-acquisition link to classroom L2 speech learning to some degree, they only suggest that this link is robust for the acquisition of lexicogrammar information.

We are grateful to the anonymous *Language Learning* reviewers and Associate Editor Kara Morgan-Short for providing their constructive comments on drafts of this paper. The project was funded by a Leverhulme Trust research grant (RPG-2019-039), a Spencer Foundation grant (10023068), and the Arnold Bentley New Initiatives Fund.

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

The handling editor for this article was Kara Morgan-Short.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Keywords second language speech; pronunciation; auditory processing; foreign language learning

Introduction

It is widely understood that adult second language (L2) speech learning is subject to a great deal of individual variation. Although some learners are able to achieve high-level L2 proficiency with apparent ease, others experience a tremendous amount of difficulty reaching the same state (e.g., Abrahamsson & Hyltenstam, 2009). One explanation for this variation relates to the amount of exposure learners are experiencing and have previously had to the target language. This corresponds to a usage-based perspective of L2 learning that views proficiency in terms of how much (quantity), in what way (quality), and how recently (timing) learners have practiced the target language (Ellis, 2006). An interesting testing ground for experience effects is the foreign language setting where exposure to L2 input is limited to several hours of language-focused instruction per week (Muñoz, 2014). There is ample evidence that the outcomes of classroom L2 speech learning could be influenced by variables such as the length of foreign language education (Jaekel, Schurig, Florian, & Ritter, 2017), extracurricular activities (e.g., Muñoz, 2014; Peters, Noreillie, Heylen, Bulté, & Desmet, 2019; Pfenninger & Singleton, 2019), type of instruction (e.g., see Norris & Ortega, 2000, for focus on form vs. focus on forms), and timing of learning (e.g., see Larson-Hall, 2008, for early vs. late starters).

At the same time, it has been reported that experience-related variables alone cannot fully explain the variance in foreign language learning success (e.g., Saito & Hanzawa, 2016, found that approximately 15–20% of variance accounted for by the length, type, and timing of L2 experience). Even if two individuals of the same age and with similar levels of motivation engage in the same type of practice for the same period of time, they will most likely end up with different levels of proficiency (Doughty, 2019). Part of this variation can be attributed to learner-internal abilities related to perception and general cognition (e.g., Linck et al., 2013, for working memory, and see Appel, Trofimovich, Saito, Isaacs, & Webb, 2019, for a special issue on the topic covering a range of learner-internal individual differences). These abilities are thought to be instrumental in the acquisition of relatively difficult, complex, and nonsalient linguistic features because they can help learners better encode, analyze, memorize, and internalize the input that they receive (Li, 2016). In this study, we investigated one such ability that has been widely cited in the first language (L1) acquisition literature and that acts as a foundation of human language and music learning: domain-general auditory processing. The

goal of this study was to examine the generalizability of this framework to L2 speech learning in the instructed foreign language setting.

Background Literature

Domain-General Auditory Processing and First Language Acquisition

Research on individual differences in L1 acquisition has mainly investigated two sets of processes: (a) those aspects of neurocognitive functioning specifically devoted to language acquisition (Campbell & Tyler, 2018) and (b) those aspects that can be characterized as domain-general learning mechanisms (e.g., Hamrick, Lum, & Ullman, 2018, for declarative vs. procedural memory; Granena & Yilmaz, 2019, for sequence learning ability). One well-researched domain-general construct is auditory processing, defined as the ability to precisely represent and remember characteristics of sounds (Mueller, Friederici, & Männel, 2012; Tierney & Kraus, 2014; Tierney, White-Schwoch, MacLean, & Kraus, 2017).

Auditory processing essentially comprises two dimensions. The first relates to the type of audio signal that is processed, that is, temporal versus spectral (Flinker, Doyle, Mehta, Devinsky, & Poeppel, 2019; Zatorre & Belin, 2001). Temporal processing is defined as a person's capacity to track changes in amplitude over time. This ability is integral to a range of phenomena related to fluency (e.g., phonation time, pause frequency) and segmental contrasts (e.g., short vs. long vowels), prosody (e.g., durational differences in weak vs. strong vowels, voice onset time), and rhythm (e.g., duration between two syllables, stressed syllables vs. morae). Spectral processing is defined as a person's capacity to track changes in the frequency content of the signal, such as pitch (the frequency of vocal cord vibration) and formants (the energy concentration at different frequency bands). This ability is fundamental to the correct assignment of stress and intonation (e.g., lexical tones in Mandarin), the fine-tuning to isochrony (e.g., syllable, stressed vs. mora-timed), and the refinement of segmental accuracy (e.g., third formant variability for English [r] and [l]).

The second dimension relates to the type of information processing involved, for example, audio-motor integration (i.e., proceduralizing temporal and spectral patterns) versus auditory acuity (encoding temporal and spectral details of sounds). In the early stages of L1 acquisition, infants need not only to detect and remember novel sound patterns and contrasts, but they also must consolidate and make them available for motor action (i.e., reproduction; Flaughnacco et al., 2014; Tierney & Kraus, 2014). At the same time, auditory acuity ability is needed to detect subtle differences in the temporal and spectral aspects of acoustic signals at a fine-grained level. Fine auditory acuity thus

Table 1 Constructs of auditory processing

Type of audio information	Type of information processing	Measures
Temporal	Audio-motor integration	Rhythm reproduction
	Auditory acuity	Duration discrimination
Spectral	Audio-motor integration	Melody reproduction
	Auditory acuity	Pitch discrimination
		Formant discrimination

makes possible phonetic restructuring that will in turn help learners increase the sophistication of their auditory representations and, by extension, to attain more advanced linguistic proficiency (McArthur & Bishop, 2005). The constructs of the auditory processing model and corresponding measures used in the current study are summarized in the first two columns of Table 1.

Auditory acuity develops and reaches its peak around 7 to 10 years of age, and thereafter the degree of precision gradually declines with age (Skoe, Krizman, Anderson, & Kraus, 2015). In comparison, auditory-motor integration continues to improve until the late 20s, and is followed by a great amount of individual variation over the remainder of the lifespan (Thompson, White-Schwoch, Tierney, & Kraus, 2015). Auditory processing is fundamental to every stage of L1 acquisition. Within the first 6 to 8 months of life, for example, infants use temporal and spectral information in speech to distinguish between the probabilities of individual phonemes existing in L1 phonetic inventories (Kuhl, 2000; for a review, see Chládková & Paillereau, 2020). At the same time, they use both temporal and spectral processing to identify word boundaries (Cutler & Butterfield, 1992), to track syntactic structure (Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992), and to detect morphosyntactic cues (the identification of suffixes; Joanisse & Seidenberg, 1998). This eventually enables L1 learners to attend to the temporal and spectral details of sounds and words and hence perceive and produce a number of phonologically similar words with correct morphological markers (Gervain & Werker, 2008). Deficits in auditory processing are known to result in a range of global language problems. For example, the audition profiles of L1 learners vary widely between normal and delayed L1 acquirers (Surprenant & Watson, 2001). Furthermore, individual differences in auditory processing have been linked to L1 learning difficulty (Goswami et al., 2011). In spite of this large body of evidence, the causal nature of the link between auditory processing and L1 acquisition has continued to be debated (cf., Halliday & Bishop, 2006; Rosen & Manganari 2001; Snowling, Gooch, McArthur, & Hulme, 2018).

Domain-General Auditory Processing and Second Language Acquisition

To test the domain generality of auditory processing, a growing number of scholars have begun to examine the extent to which the construct explains success in postpubertal speech L2 learning (for a comprehensive summary, see Table 2). Within the paradigm of novel word learning, there is some evidence that learners with more precise auditory acuity and integration abilities can better perceive foreign sound patterns and contrasts that they have never learned (e.g., Kempe, Bublitz, & Brooks, 2015, for L1 English speakers' perception of Norwegian pitch and vowel contrasts). Furthermore, auditory processing has been shown to predict gains in perception and production following focused training on novel and/or foreign sounds and words (e.g., Wong & Perrachione, 2007, for pseudo words; Li & DeKeyser, 2017, for real Chinese words). In particular, the auditory processing training studies have provided ample evidence of implication in processing and producing L2 sounds. Given that the studies directly tracked how participants with different auditory profiles reacted to short-term in-laboratory training, the findings presented longitudinal evidence regarding precisely how auditory processing helps humans notice, process, and integrate a novel language when they encounter it for the first time. Training studies of this kind have notably allowed researchers to discuss the longitudinal effects of auditory processing in the initial stage of language learning. (For recent examples of research in the broader area of L2 phonological development, see Casillas, 2020; Chládková & Šimáčková, in press; Nagle, 2020; Pelzl, Lau, Jackson, Guo, & Gor, 2020, though these studies did not examine individual differences nor auditory processing). However, the role of auditory processing in long-term L2 learning in various contexts (immersion vs. classroom) has remained unclear, especially when it comes to acquisition of other forms of linguistic knowledge such as lexicogrammar.

Though limited in number, some studies have explored the relationship between auditory processing and L2 phonological and morphosyntax learning when participants have had naturalistic, extensive, and immersive learning experience with a target language (for a summary, see Table 3). For example, for 28 Greek L2 learners of English (with approximately 10 years of L2 learning experience), Lengeris and Hazan (2010) found that their auditory processing profiles correlated with learning gains in L2 vowel accuracy when the learners received 4 hours of training simulating the intensive and highly variable nature of naturalistic L2 speech learning (i.e., high variability phonetic training). More recently, a team at University of London has conducted a series of studies focusing on more than 300 adult L2 learners with diverse L1s, immersion experience, auditory processing, and linguistic profiles (e.g., Kachlicka, Saito,

Table 2 Summary of studies examining auditory processing and novel word learning

Studies	Subjects	L2 tasks	Auditory measures	Findings
Wong and Perrachione (2007)	17 L1 English speakers	Learning 18 English pseudowords with Mandarin-like pitch patterns	Pitch identification (acuity)	Novel word learning success was tied to auditory processing (pitch acuity) and experience (prior music training). Those with strong auditory processing benefitted from high-variability training. Those with weaker auditory processing benefitted from low-variability training.
Perrachione, Del Tufo, & Gabrieli (2011)	64 L1 English speakers	Learning 18 English pseudowords with Mandarin-like pitch patterns	Pitch identification (acuity)	Those with strong auditory processing benefitted from high-variability training. Those with weaker auditory processing benefitted from low-variability training.
Cooper & Wang (2012)	28 L1 Thai & 26 L1 English speakers	Learning novel five Cantonese syllables with five different tones	Composite scores of pitch and rhythm discrimination (acuity)	Novel tone learning success was linked to composite pitch and rhythm discrimination especially among non-tonal (L1 English) speakers
Kempe et al. (2015)	118 L1 English speakers	Discriminating foreign phonological contrasts (Norwegian)	Pitch and formant discrimination (acuity)	Auditory processing mediated the relationship between music experience and novel sound perception.
Li & DeKeyser (2017)	41 L1 English speakers	Learning 16 real Chinese words	Melody reproduction (integration)	Those with more precise auditory processing demonstrated greater gains.

Table 3 Summary of studies examining auditory processing and naturalistic L2 learning

Studies	Subjects	L2 tasks	Auditory measures	Findings
Lengeris & Hazan (2011)	28 L1 Greek speakers	Learning L2 English vowel contrasts (perception)	Formant discrimination (acuity)	Those with more precise auditory processing demonstrated more advanced L2 speech proficiency (accuracy) when they received intensive/immersive speech training.
Kachlicka et al. (2019)	40 L1 Polish speakers	L2 English vowels (perception) L2 English grammaticality judgments	Spectral and temporal discrimination (acuity) and reproduction (integration)	Participants' L2 proficiency (accuracy) was equally determined by auditory processing and experience variable
Saito et al. (2020)	100 L1 Polish speakers	L2 English phonological accuracy and fluency (production)	Spectral and temporal discrimination (acuity) and reproduction (integration)	Auditory acuity predicted phonological accuracy attainment in particular. The link between auditory integration and fluency was minor at best.
Saito et al. (2020a)	30 L1 Chinese speakers	L2 English vowel and prosody (production)	Spectral and temporal discrimination (acuity)	Those with more precise auditory processing demonstrated more learning during 8 months of immersion (accuracy but not fluency).
Saito et al. (in press)	138 L1 Chinese, Polish, and Spanish speakers	L2 English vowel/prosody (perception) L2 English grammaticality judgments	Spectral and temporal discrimination (acuity)	Effects of auditory processing were clearly observed in relatively experienced L2 learners' accuracy performance (> 1 year of immersion).
Sun et al. (in press)	50 L1 Chinese speakers	L2 English vowel and prosody (perception)	Spectral and temporal discrimination (acuity) and reproduction (integration)	Auditory processing predicted the amount of L2 suprasegmental learning (accuracy) over the course of 4 months of immersion.

& Tierney, 2019; Saito, Kachlicka, Sun, & Tierney, 2020). Broadly, these studies have shown that individual differences in integration and acuity predicted phonological and grammatical proficiency even after biographical variables (age, experience) were controlled for. Such auditory processing effects were stronger when the learners had practiced the L2 in naturalistic settings (for more than 1 year).

To further extend this line of L2 research, the current study examined the generalizability of the relationship between auditory processing and various dimensions of L2 speech learning in foreign language classroom settings. According to Larson-Hall (2008, p. 36), foreign language classroom learning is a “minimal input” setting. Input exposure in these settings is limited to several hours of language-focused instruction per week, and opportunities to use the language outside of the classroom are rare. The rate of success in foreign language classrooms can be attributed not only to how much learners have practiced but to how recently, meaningfully, and interactively they have done so (Muñoz, 2014). However, the final outcomes of classroom L2 learning are subject to a great deal of individual variation, even for students in the same classrooms with similar experience profiles (Saito & Hanzawa, 2016).

A well-researched source of this variation is foreign language aptitude—a set of perceptual and cognitive abilities that underlie the development of foreign language proficiency. Previous research has shown that certain abilities are instrumental in the acquisition of relatively difficult linguistic features within a short period of time because these abilities arguably help L2 learners better encode, analyze, memorize, and internalize input at every opportunity (e.g., phonemic coding, grammar inferencing; see Skehan, 2016). Although foreign language aptitude has been found to predict various dimensions of classroom L2 learning (e.g., $r = .49$ in Li’s, 2016, meta-analysis), such an aptitude construct has been operationalized as composite competence specific to foreign language learning that comprises a combination of multiple skills (e.g., phonological awareness, analysis, and memory for phonemic coding; though see a challenge for some this research by Bokander & Bylund, 2020). To further examine precisely what kinds of perceptual-cognitive abilities explain aptitude effects in L2 acquisition, a growing number of scholars have begun to test the predictive power of more fine-grained, domain-general cognitive abilities (e.g., see Linck et al., 2013, for their attempts to include working memory as a part of foreign language aptitude). In line with this goal, the current investigation introduced an ability that represents a perceptual-cognitive foundation of human language learning: domain-general auditory processing (Tierney &

Kraus, 2017). In doing so, we propose and provide evidence for a new framework of aptitude for L2 speech learning in reference to this ability.

The Current Study

Predictions Concerning Auditory Processing, Experience, and Classroom L2 Speech Learning

In the current investigation, we set out to examine the relationships between auditory processing, experience, and L2 speech acquisition. To this end, we formulated the following research question:

To what degree do auditory processing and experience variables predict the accuracy and fluency dimensions of L2 speech learning in the foreign language classroom setting?

L2 speech learning has been characterized as a multifaceted phenomenon that involves the development of accurate and fluent language in extemporaneous speaking (Révész, Ekiert, & Torgersen, 2016; Trofimovich & Isaacs, 2012). Accuracy encompasses the abilities to pronounce consonants and vowels (segmentals), to correctly assign word and sentence stress (prosody), to choose appropriate combinations of words in different contexts (vocabulary), to correctly mark, for example, tense, aspect, agreement for gender, person, and number (morphology), and to correctly use word order and assign relations between words (syntax). Fluency entails the ability to deliver speech at an optimal rate (speed) without too many pauses (breakdown), repetitions, or self-corrections (repair). Experience refers to the extent to which learners have extensively and intensively studied a target language inside and outside classroom settings. Auditory processing has been operationalized as the integration and acuity abilities that deal with temporal and spectral information. Our predictions regarding the relationship between audition, experience, and L2 speech learning were two-fold in accordance with the different measures of learning (accuracy would entail more learning difficulty than fluency).

Prediction 1: Experience Variables Could Be Key Determinants of the Relatively Easy Aspects of L2 Speech Learning (Temporal Fluency)

We posited that most L2 learners will develop speaking fluency (greater speed and less breakdown) regardless of individual differences in auditory processing (i.e., weak audition effects) as long as they have sufficient practice with the target language (i.e., clear experience effects). This agrees with emerging empirical findings showing that much improvement occurs in the

fluency rather than accuracy aspects of language following short periods of L2 immersion (Mora & Valls-Ferrer, 2012) and classroom instruction (Saito & Hanzawa, 2018). In the context of L2 English speakers in Canada, Derwing and her colleagues conducted a series of longitudinal investigations showing that the temporal characteristics of L2 speech (fluency) continue to improve as a function of increased input and conversational experience (e.g., Derwing, Munro, Thomson, & Rossiter, 2009) but that nativelike accuracy in speech remains unchanged regardless of extensive immersion (see Derwing & Munro, 2013, for perceived accentedness; Munro, Derwing, & Saito, 2013, for vowel accuracy). As Table 3 shows, the link between auditory processing and L2 fluency in naturalistic settings has been found to be nonsignificant (Saito, Sun, & Tierney, 2020a) or minor (Saito, Kachlicka et al., 2020).

Prediction 2: Auditory Processing Variables Determine the Rate of Success for the Relatively Difficult Aspects of L2 Speech Learning (Lexicogrammar and Phonological Accuracy)

Auditory processing may play a key role in determining learning success for the accuracy (rather than fluency) aspects of L2 speech learning (i.e., phonological and lexicogrammar accuracy). The development of accuracy has been shown to be relatively resistant to rapid change at both the phonological (Flege, 2016) and lexicogrammatical levels (Saito, 2019). An interesting finding is that, although spoken lexicogrammar slowly, gradually, and continuously becomes more accurate after an extensive amount of experience (e.g., 3–5 years of immersion; Saito, 2015), high-level phonological refinement requires not only ample practice (e.g., 5+ years of immersion; Flege, Takagi, & Mann, 1995) but also special L2 learning aptitude profiles (e.g., high phonemic coding ability; Hu et al., 2013). As Table 3 shows, much evidence has suggested that auditory acuity predicts L2 phonological and morphosyntax accuracy (e.g., Saito et al., in press). Thus, our hypothesis was that more precise auditory acuity (encoding) and integration of spectral and temporal information underlies the successful acquisition of L2 phonological and lexicogrammar accuracy. To perceive and produce L2 segmental and suprasegmental contrasts, learners need to encode, analyze, and integrate novel spectral and temporal patterns (Gervain & Werker, 2008). For lexicogrammar, precise spectral and temporal processing directly relates to the accurate perception of pitch height and contour. This precise spectral and temporal processing directly helps learners establish lexical and syntactic boundaries while attending to perceptually nonsalient morphosyntactic markers (Joanisse & Seidenberg, 1998). Precise auditory processing may also facilitate the comprehension and production of

collocations, that is, lexical constructions that are fundamental to L2 speech accuracy (Saito, 2020) and that are marked by shorter word duration (Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999).

Table 4 summarizes the constructs, predictors, and outcome measures relevant to the auditory processing account of L2 speech acquisition.

Method

Participants

The participants were 39 undergraduate students (7 males, 32 females) who were majoring in a wide range of social science and humanities programs at a large university in Vietnam ($M_{\text{age}} = 20.1$ years, range: 18–21 years). Their proficiency based on TOEIC scores (measuring composite L2 English listening and reading proficiency) corresponded to A1/2 (Basic User) to B1/2 (Independent User; $M = 512.9$ out of 990, range: 370–690) as per the Common European Framework of Reference for Languages. The results of a questionnaire showed that the participants had started learning L2 English at different ages ($M_{\text{age of learning}} = 9.0$ years, range: 8–11 years). They had studied L2 English in classroom settings without any experience abroad ($M_{\text{length of learning}} = 1,182.8$ hours, range: 637.5–2,457.5 hours). At the time of the project, all the participants were registered for 4 hours of English classes per week. They reported that they spent a variable amount of time outside classrooms practicing the target language ($M_{\text{extracurricular L2 practice}} = 9$ hours, range: 6–11 hours). Some participants attempted to have L2 conversation activities with other L1 and L2 English users ($M = 0.7$ hours, range: 0–5 hours). For practical reasons, although we elicited participants' L2 speech performance and EFL backgrounds in individual face-to-face meetings, we collected their auditory data remotely.

L2 Speaking Task

As a part of the EFL curriculum at their university, students' L2 English proficiency was evaluated from multiple angles. Each student participated in a face-to-face tutoring session with an instructor. Not only were students asked to demonstrate their L2 English proficiency through a variety of tasks, but they also received feedback and training from the instructor. The entire session took about 1 hour per participant. Among a set of activities in which the participants engaged, we have reported the results of their oral performance in a monologue task that they completed at the beginning of their tutoring session. We chose this task format because our pilot study had shown that the task was suitable for eliciting sufficiently long L2 speech samples that can index participants'

Table 4 Constructs of L2 speech proficiency and its predictors

Dimensions	Subcomponents	Levels of learning difficulty	Primary predictors	Measures
Accuracy	Phonology	Strong	Strong effects of auditory processing	Segmental and prosodic accuracy
	Lexicogrammar	Mid	Moderate effects of auditory processing	Weighted accuracy; collocation association
Fluency	Speed	Low	Experience effects	Articulation rate
	Breakdown	–	Experience effects	Pauses between and within clauses

extemporaneous use of a wide variety of lexicogrammatical features. The participants were asked to talk about the following topic for 4 minutes: “What was the most recent favorite movie of yours that you watched?” To ensure that the participants would continue to speak for a sufficiently long time (totaling 4 minutes), we also prepared six discussion points that were presented below the topic question: (a) What was it called? (b) What kind of movie was it? (c) When and where did you watch it? (d) Who were the main characters? (e) What happened in the movie? and (f) Why did you like it? To avoid false starts, the talkers had to start their speech by using the following fixed first line: “The favorite movie I recently watched was _____.” All the speech tokens were recorded in a quiet room with a Roland-05 audio recorder, set at 44.1 kHz sampling rate and 16-bit quantization, and with a unidirectional condenser microphone. The audio data were transcribed for the lexicogrammar fluency and accuracy analyses. Although the number of words per token was substantially different for each participant ($M = 174.0$ words, range: 130–240 words), the speech surpassed the suggested threshold for robust L2 vocabulary analyses (Koizumi & In’nami, 2012 for +100 words).

Fluency Measures

We assessed L2 fluency from the speech sample using two measures (see Table 4). In light of Tavakoli and Skehan’s (2005) framework of utterance fluency, we analyzed the speech as follows. First, we assessed speed fluency for articulation rate, calculated by dividing the total number of syllables produced by phonation time. We analyzed phonation time by subtracting all the fillers (*ah, oh, eh*) and extensive silence (greater than 250 milliseconds) from the total length of each sample. Second, we assessed pausing behavior for the frequency of filled and unfilled pauses, calculated by dividing the number of pauses by the total number of words. Following the suggestions in many L2 fluency studies (e.g., Lambert, Kormos, & Minn, 2017), we calculated breakdown fluency separately for pauses in the middle and end of clauses. The frequency of midclause pauses has been assumed to represent the efficiency of L2 linguistic encoding processes, but the ratio of clause-final pauses is supposed to reflect conceptualization processes (Kormos, 2006). Two researchers separately transcribed 10 similar L2 speech samples (used in a different project) and coded them for speed, breakdown, and repair fluency. Next, they had a meeting where they checked the results of their transcripts and fluency analyses. Because there was no evidence of disagreement between the coders, just one of them completed the transcription and fluency analyses of the 39 Vietnamese EFL speakers’ monologues used in the current study.

Accuracy Measures

We assessed L2 accuracy from the speech sample using two measures (see Table 4). Traditionally, accuracy has been dichotomously analyzed by tallying the number of linguistic errors in obligatory contexts (correct vs. incorrect). More recently, however, many scholars have emphasized the notion of error gravity in L2 accuracy judgments (Derwing & Munro, 2015, for comprehensibility; Foster & Wigglesworth, 2016, for weighted accuracy; Saito, Trofimovich, & Isaacs, 2017, for segmental, prosodic and lexical appropriateness). According to this paradigm, there is consensus that certain errors have a greater negative impact on global L2 communicative adequacy than others (Révész et al., 2016), that the relative (rather than dichotomous) quality of accuracy should be evaluated from multiple angles using a combination of objective and subjective analyses (Trofimovich & Isaacs, 2012), and that phonological and lexicogrammar accuracy influence each other and thus should be analyzed separately (Crowther, Trofimovich, Isaacs, & Saito, 2015). We analyzed each dimension of accuracy as follows.

Phonological Accuracy

We adopted the training and rating procedure for subjective analyses of phonological accuracy originally conceptualized, developed, and validated by Saito et al. (2017). First, we recruited two linguistically trained raters: L1 Vietnamese speakers with high-level L2 English proficiency. Both were PhD candidates with an academic background in linguistics, and both had 8 years of EFL teaching experience in Vietnam. As Saito, Suzukida, and Sun (2019) argued, recruiting highly proficient and experienced L2 users rather than native speakers as listeners adds a degree of ecological validity to L2 speech research methodology because such expert L2 raters are believed to be able to adequately evaluate the degree to which speakers of the same L1 are making efforts to acquire and use L2 English rather than continuously relying on their L1 systems.

The two raters first underwent a brief training session with the first author on the three different constructs of L2 phonological accuracy: (a) segmentals (substitution, omission, or insertion of individual consonant and vowel sounds), (b) word stress (misplaced or missing lexical stress in multisyllabic words), and (c) intonation (appropriate, varied use of pitch movements; for training scripts, see Appendix S1 in the online Supporting Information). Then, each rater separately listened to the 39 speech files in a randomized order. For each speech sample, the raters assessed the extent to which a speaker made an effort to approximate the targetlike use of segmentals, word stress, and intonation in L2 English rather than in their L1 Vietnamese on one scale from

1 (*nontargetlike*) to 9 (*targetlike*). The interrater reliability was significantly high: for segmentals, $r = .85$, for word stress, $r = .84$, and for intonation, $r = .81$. Given that the scores of the two raters did not show any clear disagreement (defined as more than a 2-point difference at each component), we decided to use their averaged scores for the subsequent analyses.

Lexicogrammar Accuracy

We adopted two different analyses of lexicogrammatical accuracy in this study: (a) subjective judgments (global weighted accuracy) and (b) corpus-based text analysis (collocation association). Foster and Wigglesworth (2016) originally proposed a scale from 1 (*very serious errors hindering meaning*) to 4 (*entirely accurate*) to assess the weighted accuracy of lexicogrammar in conjunction with its relative impact on global understanding. Following this line of thought, Appel et al. (2019) stressed the importance of evaluating the accurate use of lexicogrammar from the perspective of comprehensibility (i.e., overall ease of understanding). Therefore, the same two expert L1 Vietnamese raters also conducted the lexicogrammar judgments. To factor out the influence of fluency-related phenomena, we eliminated all filled pauses (e.g., *ah, eh, oh*) from the transcripts. After the raters had received training from the researcher on the definition of global lexicogrammar accuracy (for training scripts, see Appendix S2 in the online Supporting Information), they proceeded to read the 39 transcripts in a randomized order and assess each written file for global accuracy on a scale from 1 (*difficult to understand*) to 9 (*easy to understand*). We again identified significantly high agreement, $r = .87$. Because we observed no major disagreement (more than a 2-point difference), we used the raters' averaged scores as a measure of participants' global lexicogrammar accuracy for the subsequent analyses.

For the corpus-based text analysis, a growing number of studies have shown that collocation use forms a crucial component of L2 speech proficiency (Kyle & Crossley, 2015; Tavakoli & Uchihara, 2020) and can serve as a good index of speakers' ability to use lexicogrammar appropriately in context (Saito, 2020). In general, collocation is defined as "the phenomenon surrounding the fact that certain words are more likely to occur in combination with other words in certain contexts" (Baker, Hardie, & McEnery, 2006, p. 36). One useful analytic unit of collocation is n -gram association, that is, what is the frequency and the statistical likelihood of n words occurring together (but not with any other word). In the current study, we operationalized this using mutual information scores (for a comprehensive overview, see Gablasova, Brezina, & McEnery, 2017). To calculate mutual information scores, we submitted all the cleaned

transcripts to the bigram and trigram measures available in the Tool for the Automatic Analysis of Lexical Sophistication (Version 2.0; Kyle & Crossley, 2015). We calculated mutual information scores by dividing the frequency of collocations by the frequency of random co-occurrence of the words. We chose the Corpus of Contemporary American English as the reference corpus (Davies, 2009). Mutual information scores reflect the exclusivity of word combinations; higher scores are assigned to low-frequency associations that do not have many other partner words. To create composite collocation scores for each transcript, we standardized and averaged both bigram and trigram mutual information scores.

Measures of Auditory Processing

As the third column of Table 1 indicates, we measured four different aspects of participants' auditory processing abilities (i.e., audio-motor integration and auditory acuity for temporal and spectral information). We assessed audio-motor integration via reproduction tasks and auditory acuity via discrimination tasks.

Although we collected the L2 English speech samples as a part of the EFL curriculum at the university and as a form of individual tutoring with an instructor, participants were not required to continue with the auditory processing test that took approximately 30 extra minutes of their time. Therefore, we asked for volunteers who would be willing to participate in the auditory test battery. To administer all the extra data collection in an efficient manner and to reduce the burden for the participants, we allowed them to complete the auditory processing tests using their own computer at their convenience. For the participants to do so, we first uploaded the test materials onto our inhouse website and piloted them multiple times. Next, when interested participants contacted the researcher, she (a L1 Vietnamese speaker) held a brief online meeting in which the participants received the instructions for each auditory processing test in L1 Vietnamese. All the participants were explicitly instructed to engage in the test in a quiet room using their computer and headset. When the participants had any questions, they contacted the researcher to ensure that they fully understood the procedure.

All the participants followed the same task order. They first engaged in the audio-motor integration task (rhythm and melody reproduction) and then the auditory acuity task (duration, pitch, and formant discrimination). Initially, 42 participants joined our project and completed the auditory processing tests without any problems. Although we carefully monitored the participants' auditory processing performance, we found that the temporal integration

performances of three participants were not properly recorded due to some technical issues. Thus, we eliminated these three participants from all analyses.

Audio-Motor Integration

Following the procedures used in Tierney et al. (2017), the participants completed two different audio-motor integration tasks: rhythm and melody reproduction. The rhythm reproduction task was designed to tap into the participants' temporal integration ability; and the melody reproduction task into their spectral integration ability.

The rhythm reproduction task evaluated the extent to which the participants could easily remember perceptible rhythmic sequences (i.e., broader levels of temporal information) and reproduce them. In the test, the participants listened to a four-measure sequence three times and were asked to drum out the sequence as if there were a fourth repetition. The participants were presented with a total of 30 trials, with the first 15 being strongly metrical and the remaining 15 being weakly metrical (described by Povel & Essens, 1985). The weakly metrical sequences contained fewer drum hits on the first and third beats than did the strongly metrical sequences. The participants' drumming was quantized by changing each interdrum-interval to the nearest interval in the set (200, 400, 600, and 800 milliseconds). Each of the participants' responses was treated as a sequence of hits and rests such that the program checked whether there was a drum hit or a rest every 200 milliseconds. The participants' sequence of hits and rests was then compared to the sequence of hits and rests in the stimulus. The resulting ratio of correct hits and rests constituted the rhythm integration scores.

For melody reproduction, we designed a new task to evaluate the extent to which the participants could recollect and reproduce a sequence of complex tones that varied in pitch. Each melody consisted of a sequence of seven notes. Each of these notes was drawn from a set of five six-harmonic complex tones with equal amplitude across harmonics and fundamental frequencies equal to the first five notes of the major scale, corresponding to frequencies of 220, 246.9, 277.2, 311.1, and 329.6 Hz. Each note was 300 milliseconds in duration, with a 50 milliseconds cosine ramp at the beginning and end of the note to avoid perception of transients. No silence was interposed between notes within a melody. Melodies were pseudorandomly constructed in the following manner. Each melody began on the third note of the scale, that is, 277.2 Hz. The next note was then randomly chosen to be either one note higher on the scale (311.1 Hz) or one note lower on the scale (246.9 Hz). This process then repeated until all seven notes were chosen. The melody could not descend below

220 Hz or ascend above 329.6 Hz; once the melody reached these limits, the next note was chosen to either be closer to the center of the range or identical to the previous note.

During the test, participants were told that they were completing a memory test in which they would hear melodies repeated three times and that they were to try to remember the melodies and then to repeat them. They were then played an example of a melody, repeated three times. Next, they were shown a set of five buttons vertically arranged on the screen, labeled “1” through “5.” Each of these buttons, when clicked, turned from black to green in outline and played one of the five notes of the scale (with the lowest note linked to the button labeled “1,” which was also arranged at the lowest point on the screen). Participants were encouraged to try clicking on these buttons to familiarize themselves with the tone linked to each button. Finally, participants were explicitly told that each melody would begin with the note linked to the button labeled “3.” The test itself consisted of 10 melodies, each of which was presented three times, with a 1-second pause in between repetitions. After these repetitions finished, the five boxes once again appeared on the screen, and participants were instructed to reproduce the melody by pressing on the boxes. When they clicked each box, the tone that was linked to that number was played. Once they had completed their reproduction, they were asked to click on a “next trial” button to advance to the next melody. To assess performance, we compared the first seven button presses produced by the participant to the target melody, scoring identical notes as 1 and scoring notes that differed to any degree as 0. We then averaged the participants’ performance across all 10 melodies.

Auditory Acuity

We administered three psychoacoustic tests to assess the participants’ ability to capture temporal and spectral details of sounds: duration, pitch, and formant discrimination thresholds (Surprenant & Watson, 2001). We designed duration discrimination thresholds to assess the participants’ temporal acuity, and we designed pitch and formant thresholds to assess the participants’ spectral acuity. For each test, we created 100 synthesized stimuli using custom MATLAB scripts. These stimuli varied along a single acoustic continuum; they either had 100 different durations, 100 different fundamental frequencies (i.e., pitch), or 100 different formant values. In each trial, three different tones were presented with an interstimulus interval of 0.5 s. Upon hearing each sequence, the participants were asked to choose which of the three tones differed from the other two by pressing the number “1” or “3.” On the basis of Levitt’s (1971) adaptive

threshold procedure, we designed the size of the difference to vary from trial to trial in accordance with task performance.

General procedure. Because there were 100 target samples, each file was labeled from Levels 1 to 100. The standard/anchor stimulus was labeled as Level 0. If the participants could perceive the difference between the standard stimulus (Level 0) and Level 1, this represented the highest level of auditory sensitivity. If they could hear the difference only when they compared between the standard stimulus (Level 0) and Level 100, this indicated the lowest level of auditory sensitivity. That is, the lower scores were a proxy for higher auditory sensitivity. Initially, the tests started from the midpoint, Level 50. In other words, in the first trial, the two identical stimuli had a value of 0 on the target acoustic continuum (duration, pitch, or formant frequency), but the different target stimulus had a value of 50 on that continuum. When an incorrect response was made, the difficulty of the task decreased by a degree of 10 steps (with the difference being wider). For example, if the participants answered the first trial incorrectly, for the second trial the target stimulus would have a value of 60. When they provided three consecutive correct responses, the task difficulty increased by a degree of 10 steps (with the difference being smaller). In other words, the level of the target stimulus might change from 50 to 40. The step size decreased when the direction of difficulty between trials reversed, that is, when an increase in difficulty was followed by a decrease in difficulty, or vice versa. After the first reversal, the step size changed from 10 to 5, and then, after the second reversal, from 5 to 1. The logic behind this feature of the test was that large changes were made to test difficulty initially to find the stimulus range where the test was difficult but not impossible, and then fine adjustments were made to test difficulty from that point on so that the participant's threshold could be very precisely measured. The tests stopped either after 70 trials or eight reversals. Participants' auditory processing score was determined by averaging the stimulus levels at which the reversals occurred, starting at the third reversal. This was a measurement of the stimulus level at which a participant could just barely discriminate the stimuli. For example, one participant's third through eighth reversals were at Levels 50, 35, 40, 35, 45, and 41. This participant's score was calculated as the average of these six numbers, or .41. This participant, therefore, could just barely tell the difference between a stimulus at Level 41 and a stimulus at Level 0. What each stimulus level indicated was different across the three subtasks (duration, pitch, and formant discrimination).

Stimuli for duration discrimination. We prepared a total of 100 four-harmonic complex tones with the fundamental frequency set to 330 Hz and

equal amplitude across harmonics. The duration of the standard stimulus (Level 0) was 250 milliseconds. To avoid the perception of transients, we included two amplitude ramps at the onset and endpoint of the stimulus (15 milliseconds each). To differentiate the 100 tones in terms of duration (Levels 1–100), we manipulated the target acoustic dimension (duration) in steps of 2.5 milliseconds (252.5–500 milliseconds). For example, if a participant's reverse happened at Level 10, this meant that the minimum difference in duration that the participant could hear was 25 milliseconds, that is 250 milliseconds (standard stimulus) versus 275 milliseconds (target stimulus).

Stimuli for pitch discrimination. We used the same 100 four-harmonic complex tones from the duration discrimination task. This time, however, the duration dimension remained the same throughout (i.e., 250 milliseconds); but we set an F0 of 330 Hz as the standard stimulus (Level 0), and manipulated F0 as the target acoustic dimension for the remaining stimuli (Levels 1–100). All the 100 stimuli differed between 330.3 and 360 Hz in F0 with a step of 0.3 Hz. For example, if a participant's reversal happened at Level 10, this meant that the minimum difference in pitch that the participants could hear was 3 Hz, that is, 330 Hz (standard stimulus) versus 333 Hz (target stimulus).

Stimuli for formant discrimination. We created a total of 100 complex tones. The duration of each token was 500 milliseconds with a fundamental frequency of 100 Hz and harmonics up to 3,000 Hz. We inserted two 15 milliseconds amplitude ramps at the beginning and endpoint of the stimulus. Using the technique of a parallel formant filter bank (Smith, 2007), we generated three formants at 500, 1,500, and 2,500 Hz. We set an F2 of 1,500 Hz as the standard stimulus (Level 0) and manipulated F2 as the target acoustic dimension for the remaining stimuli (Levels 1–100). All the 100 stimuli differed between 1,502 and 1,700 Hz in F2 with a step of 2 Hz. For example, if a participant's reversal happened at Level 10, this meant that the minimum difference in formant that the participant could hear was 20 Hz, that is 1,500 Hz (standard stimulus) versus 1,520 Hz (target stimulus).

Calculating temporal versus spectral acuity. We used the participants' duration discrimination scores to index their temporal acuity. Following the method of calculating spectral acuity in the precursor research (Kachlicka et al., 2019), we standardized and averaged the participants' pitch and formant discrimination scores. Thus, we considered the composite spectral acuity to represent the participants' sensitivity to lower frequencies (pitch discrimination) and higher frequencies (formant discrimination).

Reliability of Reproduction and Discrimination Tasks

To determine the test–retest reliability of the audio-motor integration and acuity tasks, we conducted a follow-up study. Using the test procedure described above, a total of 30 English users not included in the current study with diverse experience and proficiency levels took the reproduction and discrimination tests online twice one day apart. According to Pearson correlation analyses (summarized in Table 5), the test instruments yielded medium-to-large test–retest effects ($r = .562\text{--}.907$) except for duration discrimination ($r = .284$). The test–retest reliability for the combined spectral discrimination scores (i.e., averaging formant and pitch discrimination scores) was $r = .598$ ($p < .001$).

It is important to note that the results suggested that some parts of our online testing of auditory abilities ($r = .907\text{--}.775$ for spectral and temporal reproduction) reached an acceptable level of test–retest reliability (as identified by Lance, Butts, & Michels, 2006) as well as reaching the level of test–retest reliability previously reported for in-laboratory testing (for example, $r = .75$ in Raz, Willerman, & Yama, 1987). However, the low reliability of the other measures ($r = .598$ and $.284$ for spectral and temporal discrimination) could be ascribed to several scenarios (e.g., the lack of validity, small sample size, and inconsistent sound system across participants; for more details and open data, see Saito, Sun, & Tierney, 2020b).

Results

The research question asked how auditory processing and experience variables are related to the accuracy and fluency dimensions of L2 speech learning among learners in the foreign language classroom setting. For individual differences research of this kind where there are multiple predictor and dependent variables, it is essential to examine how these variables relate to each other to avoid multicollinearity problems. Thus, we first present the auditory processing scores and their associations with experience backgrounds (the predictor variables). After we have summarized the participants' L2 speech accuracy and fluency proficiency (the dependent variables), we finally present the results of multiple and mixed effects regression analyses to shed light on the complex relationship between auditory processing, experience, and L2 speech proficiency. We selected an alpha level of .05 as the level of significance for the statistical tests and applied the Bonferroni correction to this alpha when we used multiple tests within an analysis.

Table 5 Results of test–retest reliability for the reproduction and discrimination tests used to examine integration and acuity, respectively

Melody	Reproduction (Integration)		Discrimination (Acuity)					
	<i>r</i>	<i>p</i>	Rhythm	Formant	Pitch	Duration		
.907	< .001	.775	< .001	.619	< .001	.562	.284	.128

Table 6 Interrelationships of auditory processing measures

Measures	Spectral integration		Temporal acuity		Spectral acuity	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Temporal integration	.307	.069	-.213	.211	-.018	.918
Spectral integration			-.002	.993	.006	.972
Temporal acuity					.438	.005

Note. Adjusted alpha (after Bonferroni correction for multiple correlations, $0.05/3 = .016$).

Characteristics of Auditory Processing

According to the results of Kolmogorov-Smirnov tests, the participants' pitch discrimination scores demonstrated significant deviation from normally distributed patterns, $D(39) = .214, p = .047$, but their formant and duration scores were comparable to a normal distribution, $D(39) = .094, p = .846$, and $D(39) = .117, p = .615$, respectively. Thus, we transformed the raw pitch scores using a log10 function. We standardized and averaged these transformed pitch and raw formant discrimination scores to index the participants' spectral acuity. The composite spectral acuity did not significantly differ from a normal distribution, $D(39) = .139, p = .397$. We used the raw duration discrimination scores for temporal acuity. As for audio-motor integration, both raw rhythm and melody reproduction scores did not significantly differ from a normal distribution, $D(36) = .098, p = .810$, and $D(39) = .096, p = .861$, respectively. We used the raw rhythm reproduction scores to index temporal integration and the raw melody reproduction scores to index spectral integration.

Table S3.1 in Appendix S3 in the online Supporting Information provides a descriptive summary of the participants' raw auditory processing scores. For the remainder of the analyses, we used the four different dimensions of auditory processing summarized in Table 1. In the current study, we operationalized spectral acuity as combined pitch and formant discrimination scores, temporal acuity as duration discrimination scores, spectral integration as melody reproduction scores, and temporal integration as rhythm reproduction scores. For the interrelationships between integration and acuity scores, the results of Pearson correlation analyses (summarized in Table 6) showed a moderate correlation between temporal and spectral acuity scores ($r = .438$). There were no other significant associations between the integration and acuity dimensions (adjusted alpha with Bonferroni correction, $.05/3 = .017$). This supported the

Table 7 External relations between auditory processing measures and other influencing variables

Measures	Past experience		Current experience	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Temporal integration	.495	.002	.331	.049
Spectral integration	.281	.084	.299	.065
Temporal acuity	-.022	.894	-.015	.929
Spectral acuity	-.096	.562	-.036	.829

Note. Adjusted alpha (after Bonferroni correction for multiple correlations, $0.05/2 = .025$).

conceptualization shown in Table 1 and suggested that integration and acuity consist of two theoretically dissociable aspects of auditory processing, at least within the current dataset.

Auditory Processing and Experience

We performed another set of Pearson correlation analyses to examine the relationship between auditory processing and language experience. For the rest of the analyses, the length of foreign language education was labeled as “past experience” (hours in total) and the current L2 use outside classrooms as “current experience” (hours per week). The results presented in Table 7 suggested that temporal integration in particular may be tied to the extent to which participants have accumulatively practiced the target language inside L2 classrooms ($r = .495$); and that the acuity aspect of auditory processing may be independent of experience variables.

Characteristics of L2 Speech Proficiency

Table S3.2 in Appendix S3 in the online Supporting Information summarizes the participants’ L2 fluency and accuracy performance. The results of Kolmogorov-Smirnov tests suggested that the fluency and accuracy scores were normally distributed, $Ds(39) = .073-.193$, $ps = .093-.972$. To examine the relationships among a total of three fluency measures and five accuracy measures, we performed a set of Pearson correlation analyses (adjusted alpha with Bonferroni correction, $0.05/7 = .007$). As Table 8 shows, the three temporal fluency measures (articulation rate, midclause pause ratio, clause-final pause ratio) were significantly or marginally correlated with each other (all $p < .012$). The three phonological accuracy measures (segmentals, word stress,

Table 8 Interrelationships of L2 accuracy and fluency measures

Measures	Midclause		Clause-final		Segmentals		Word stress		Intonation		Global		Collocation	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Temporal fluency														
Articulation rate	-.624	< .001	-.683	< .001	-.043	.793	-.128	.437	-.058	.728	.358	.025	.217	.185
Midclause pause ratio	.400	.012	.017	.918	.177	.282	.103	.533	.103	.533	-.311	.054	.318	.048
Clause-final pause ratio			-.038	.817	.057	.730	.038	.818	.038	.818	-.400	.012	-.052	.755
Phonological accuracy														
Segmentals					.875	< .001	.865	< .001	.875	< .001	.225	.168	.028	.863
Word stress											.115	.486	.081	.625
Intonation														
Lexicogrammar accuracy										.212	.196	.024	.886	
Global accuracy													.101	.540

Note. Correlations in boldface are significant for adjusted alpha (after Bonferroni correction for multiple correlations, 0.05/7) = .007. 95% confidence intervals around *r* are provided in Appendix Table S4.1 in the Supporting Information online.

intonation) demonstrated relatively strong associations ($r_s = .865-.875$). However, a relationship between two lexicogrammar accuracy measures (global accuracy judgments vs. collocation) remained unclear, $r = .101$, $p = .540$. Although global accuracy demonstrated marginally significant negative associations with clause-final pause ratio, $r = -.400$, $p = .012$, the lexicogrammar accuracy measures were not clearly clustered into the other fluent and accuracy measures. Taken together, the results suggested that the eight outcome measures appeared to tap into four broadly different aspects of participants' L2 oral abilities: (a) temporal fluency, (b) phonological accuracy, (c) lexicogrammar accuracy, and (d) collocational use.

Roles of Auditory Processing and Experience in L2 Speech Proficiency

In order to examine the relative weights of auditory processing and experience variables in L2 speech proficiency, we conducted several regression analyses. Just as the existing literature has discussed (Trofimovich & Isaacs, 2012), we found that a total of eight outcome measures (see Table 8) tapped into four different dimensions of L2 speech proficiency

1. Phonological accuracy (segmentals, word stress, and intonation)
2. Fluency (articulation rate, midclause pause ratio, and clause-final pause ratio)
3. Lexicogrammar accuracy (global accuracy judgments)
4. Collocation accuracy (mutual information scores)

Therefore, we constructed a total of four different models relative to six predictors, that is, temporal and spectral integration and acuity, past experience (total hours of foreign language education), and current experience (extracurricular L2 practice). According to the results of our power analyses using G*Power, the observed statistical power for these models explaining a relatively large amount of variance ($f^2 = 0.35$) in the 39 participants' L2 speech proficiency by way of the six predictors was .71, which could be considered acceptable in the field of applied linguistics (Larson-Hall, 2010).

For Models 1 and 2, the dependent variables were hierarchical, where the same participants' performance needed to be tested three times in each model (i.e., segmentals, word stress, and intonation for phonological accuracy; articulation, midclause, and clause-final pauses for fluency). To take random effects of subjects (being tested three times) into account, we constructed two separate mixed effects regression models using the `lm` and `glmer` functions from the `lme` package (Version 1.1-21; Bates, Maechler, Bolker, & Walker, 2015) in R (Version 3.6.3; R Core Team, 2018). Because the directions of articulation

and pause measures were opposite (with faster speech rate and fewer pauses indicating better fluency), we reversed scores on the pause ratio measures.

For Models 3 and 4, the dependent variables involved one single dimension (i.e., global accuracy judgment scores for lexicogrammar accuracy, mutual information scores for collocation accuracy). Thus, we constructed two separate multiple regression models using the `lme` function in R. In essence, both analyses (mixed effects and multiple regression) provided insights and values that can be interpreted in the same way, that is, whether, how, and to what degree each predictor variable was associated with the dependent variables.

Table 9 shows the results of mixed effects modeling for Models 1 and 2, and Table 10 shows the results of multiple regression for Models 3 and 4. In both analyses, auditory processing and experience variables uniquely explained 8.1% to 37.9% of variance for different dimensions of L2 speech. There was no strong evidence of multicollinearity problems (variance inflation factors = 1.01–1.72). We observed the following patterns for the relationship between auditory processing, experience, and L2 speech proficiency. First, the fluency measures demonstrated significant associations with the amount of L2 English use outside of the classroom (extracurricular L2 practice). Second, the degree of L2 lexicogrammar accuracy (overall comprehensibility, collocational use) were primarily determined by auditory processing variables (spectral integration). Third, the driving predictor of phonological accuracy was unclear. Figure 1 graphically summarizes the relationship between auditory processing and L2 speech proficiency when we controlled for the experience variables.

Discussion

Domain-general auditory processing refers to the extent to which learners can capture and internalize broad levels of temporal and spectral information (audio-motor integration) and the extent to which they can perceive temporal and spectral details of sounds to refine the quality of auditory categorization (auditory acuity). In the L1 acquisition literature, integration and acuity measures have been shown to predict the outcomes of normal and abnormal language development (e.g., Tierney & Kraus, 2014, for audio-motor integration; Surprenant & Watson, 2001, for spectral acuity). The main objective of the current investigation was to examine the generalizability of this construct to postpubertal L2 speech learning in the foreign language setting. We gathered data on auditory processing ability (audio-motor integration, auditory acuity), learning experience (quantity, quality, timing) and L2 speech profiles (fluency, phonology, lexicogrammar) from 39 Vietnamese EFL learners with an

Table 9 Summary of mixed-effects modeling analyses reporting standardized coefficients (*b**)

Dependent variables	Fixed effects	<i>b</i> *	95% CI	SE	<i>t</i>	<i>p</i>	Conditional <i>R</i> ²	Marginal <i>R</i> ²
Fluency (articulation rate, pauses)	Temporal integration	-0.143	[-0.401, 0.130]	0.126	-1.142	.263	.593	.379
	Spectral integration	0.041	[-0.175, 0.257]	0.105	0.387	.701		
	Temporal acuity	0.050	[-0.321, 0.422]	0.181	0.276	.784		
	Spectral acuity	-0.128	[-0.501, 0.243]	0.182	-0.706	.486		
	Past experience	0.056	[-0.239, 0.353]	0.144	0.390	.699		
	Current experience	0.619	[0.349, 0.889]	0.132	4.687	< .001		
	Random effects	Variance	95% CI	SD				
	Participants	0.227	[0.121, 0.333]	0.477				
Dependent variables	Fixed effects	<i>b</i> *	95% CI	SE	<i>t</i>	<i>p</i>	Conditional <i>R</i> ²	Marginal <i>R</i> ²
Phonological accuracy (segmentals, prosody)	Temporal integration	-0.147	[-0.543, 0.248]	0.193	-0.761	.453	.881	.081
	Spectral integration	0.057	[-0.274, 0.390]	0.162	0.356	.725		
	Temporal acuity	-0.475	[-1.047, 0.095]	0.279	-1.702	.099		
	Spectral acuity	0.196	[-0.375, 0.768]	0.279	0.704	.487		
	Past experience	0.039	[-0.416, 0.494]	0.222	0.176	.862		
	Current experience	-0.054	[-0.469, 0.359]	0.202	-0.271	.788		
	Random effects	Variance	95% CI	SD				
	Participants	0.231	[0.154, 0.308]	0.311				

Note. Value in boldface is significant for alpha = .05.

Table 10 Summary of multiple regression analyses reporting standardized coefficients (*b**)

Dependent variables	Fixed effects (predictors)	<i>b</i> *	95% CI	SE	<i>t</i>	<i>p</i>	<i>R</i> ²	Adjusted <i>R</i> ²
Lexicogrammar accuracy (global)	Temporal integration	0.091	[-2.894, 4.940]	0.170	0.535	.597	.469	.359
	Spectral integration	0.449	[0.997, 4.716]	0.143	3.142	.003		
	Temporal acuity	0.068	[-0.023, 0.031]	0.245	0.278	.782		
	Spectral acuity	-0.020	[-0.612, 0.563]	0.245	-0.085	.933		
	Past experience	0.018	[-0.001, 0.001]	0.195	0.096	.923		
	Current experience	0.309	[-0.017, 0.206]	0.178	1.736	.093		
Lexicogrammar accuracy (collocation)	Temporal integration	0.060	[-3.098, 4.253]	0.187	0.322	.751	.298	.152
	Spectral integration	0.454	[0.721, 4.211]	0.157	2.889	.007		
	Temporal acuity	0.048	[-0.023, 0.027]	0.270	0.178	.859		
	Spectral acuity	-0.230	[-0.782, 0.322]	0.270	-0.853	.400		
	Past experience	0.056	[-0.001, 0.001]	0.212	-0.263	.794		
	Current experience	0.075	[-0.085, 0.124]	0.196	0.385	.703		

Note. Values in boldface are significant for alpha = .05.

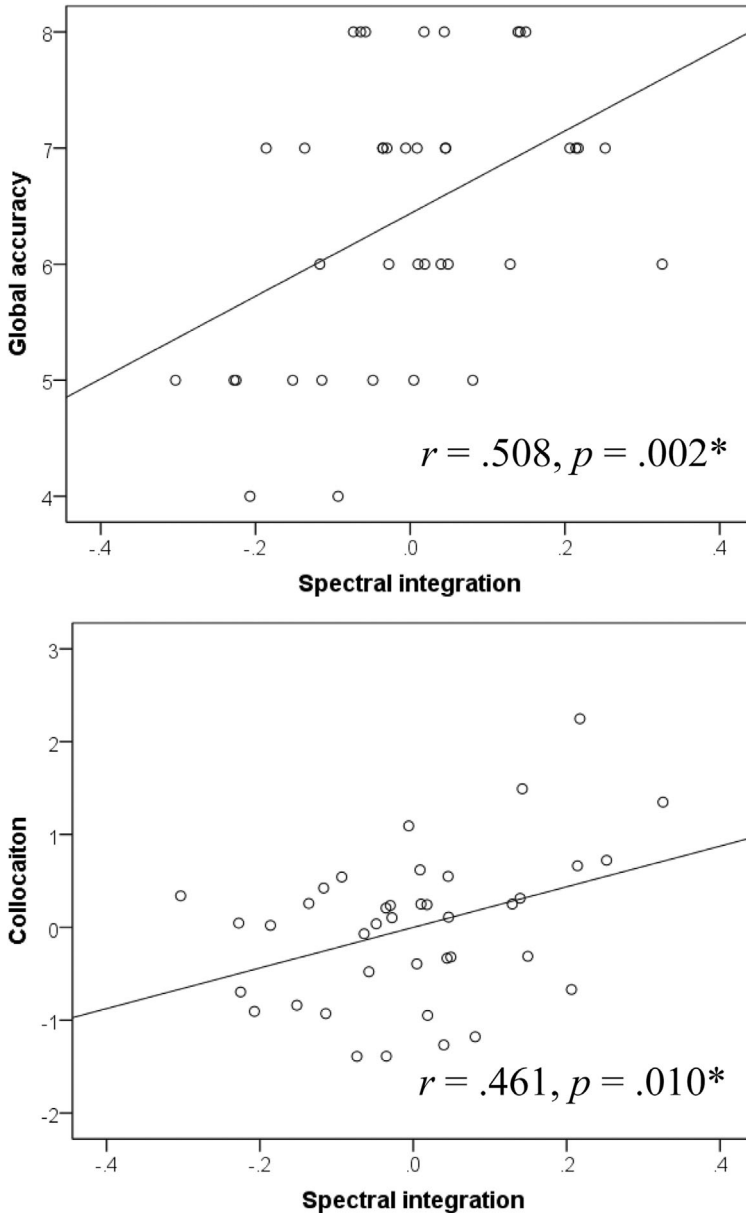


Figure 1 Correlations of auditory processing and L2 speech with experience variables (total hours of foreign language education, extracurricular L2 practice) partialled out. $*p < .05$.

extensive amount of foreign language learning experience in classrooms (>1,000 hours) and submitted them to correlation and regression analysis.

For the relationships between auditory processing, experience, and L2 speech proficiency, both audio-motor integration and experience variables uniquely explained some aspects of L2 speech proficiency attainment in classroom settings. Specifically, the findings confirmed our prediction that the temporal fluency aspects of L2 speech (articulation rate, pause ratio) are tied to learning experience and that lexicogrammar accuracy and collocation use (global judgments, collocation association) are primarily determined by auditory-motor integration (relative to the other variables we included in our analysis). In a broad sense, these findings lead to the conclusion that (a) L2 learners can improve the fluency aspects of L2 speech, even in a foreign language setting, as long as they practice on a regular and frequent basis (Saito & Hanzawa, 2016, 2018) and (b) individual differences in the ability to perceive and reproduce auditory patterns may play a key role in the acquisition of difficult linguistic features (Li, 2016). In conjunction with Plonsky and Ghanbar's (2018) field-specific benchmarks, the strength of the experience effects and the audition effects could be considered as moderate to large ($r_s = .40-.60$). The conclusion here generally concurs with the existing short-term training literature showing auditory processing facilitates the process and product of novel language learning (e.g., Wong & Perrachione, 2007). This conclusion is in line with emerging research showing that auditory processing matters for the acquisition of the relatively difficult aspect of L2 speech acquisition in naturalistic settings (i.e., accuracy rather than fluency; Saito et al., 2020a; Saito et al., in press).

At the same time, however, we would like to emphasize that the conclusions described here need to be interpreted with some caution pending further empirical investigation and replication. Although auditory processing played a significant role according to the findings of the current study, it is important to point out that the predictive capacity of the construct may be doubtful, especially considering the asymmetric relations between the different types of auditory processing and different dimensions of L2 speech. In the current dataset, a large portion of the audition-proficiency link was actually restricted to audio-motor integration (rather than acuity) in the lexicogrammar (rather than phonological) dimensions. This asymmetric pattern found among classroom L2 learners was different from what we previously reported among naturalistic L2 learners (e.g., Kachlicka et al., 2019, for the significant role of acuity and integration in both phonology and morphology). As we reviewed earlier, scholars in cognitive psychology assume the predictive power of

auditory processing in acquisition because more precise auditory processing abilities help encode and integrate aural input in a more efficient and effective fashion (Mueller et al., 2012; Tierney & Kraus, 2014). The precursor research has shed light on the generalizability of the model to various dimensions of L2 acquisition, and we now discuss the findings in the current study in order to update and fill in the theoretical details of an audition-based account of L2 acquisition. In essence, we argue that the quality and quantity of experience that learners go through (naturalistic vs. classroom) may further determine types of auditory processing abilities that learners primarily use (integration vs. acuity) and dimensions of language that auditory processing facilitates (phonology vs. lexicogrammar).

With respect to naturalistic settings, similar to L1 acquisition, L2 learners access ample aural input necessary for the simultaneous development of L2 phonology and lexicogrammar as long as they seek opportunities to use a target language (Derwing & Munro, 2013). In such contexts, for the efficient and effective processing of every input opportunity, learners rely on both auditory integration (converting input into motor action) and acuity (conducting fine-grained analyses of input). Thus, those with more precise auditory processing can demonstrate various dimensions of advanced L2 proficiency (phonology and lexicogrammar), and the trend becomes stronger as a function of increased input (e.g., Saito et al., 2020a, for the longitudinal relationship between auditory processing and L2 speech acquisition within the first 8 months of immersion; Saito, & Tierney, in press, for the first 4 months of immersion).

As for classroom L2 learning (the main focus of the current study), experience is problematic in many ways. Muñoz (2014) has pointed out that the input received by foreign-language learners is limited in source (mainly the teacher), quantity (not all teachers use the target language as the language of communication in the classroom), and quality (there is great variability in teachers' oral fluency and general proficiency, as shown by Graham, Courtney, Marinis, & Tonkyn, 2017). Following our revised model of audition-based account of L2 learning, we argue that these unique characteristics of the input available through an instructed foreign-language experience explain the unpredicted findings of the current study, that is, the significant associations between lexicogrammar (rather than phonological) accuracy aspects of L2 speech learning and the integration (rather than acuity) dimension of auditory processing.

In Vietnamese EFL classrooms, adult L2 learners typically learn the target language through decontextualized teaching methods such as grammar translation and audiolingualism (mechanical repetition and memorization of target sentences). Although learners do receive some form of aural input from

teachers (e.g., choral repetition of target sentences), the input that these approaches provide is known to be not only insufficient, but also skewed. For example, these kinds of EFL approaches are characterized by their exclusive emphasis on production (rather than comprehension) practice of lexicogrammar (rather than phonology). As the current study has shown, this could explain why auditory integration (rather than acuity) abilities could clearly predict L2 lexicogrammar (rather than phonology) proficiency among our participants who had been through years of EFL experience.

First, both grammar translation and audiolingualism recommend that the target language should be mainly learned through repetitive output of oral and written sentences. Although researchers have emphasized the importance of comprehension-based practice, where students receive an abundant amount of contextually rich aural input in order to enhance understanding of language, this approach tends to be neglected in many EFL classrooms (Shintani, Li, & Ellis, 2013). For those interested in the Vietnamese EFL setting in particular, Nguyen (2017) have provided a good reference for context-specific issues related to over-reliance on grammar-translation and the significant lack of authentic L2 input. Given that students lack enough auditory, communicatively authentic input, and input-based practice opportunities in order to develop, refine, and sophisticate their auditory representations, it is reasonable to assume that processing even limited input for production (audio-motor integration rather than acuity) may be a relatively key skill in successful L2 learning in EFL classrooms. This is essentially different from the context of naturalistic L2 speech learning, where auditory integration and acuity have been found to be equally instrumental to success (Kachlicka et al., 2019).

Furthermore, there is a great amount of educational reporting that has revealed that the focus of instruction is exclusively on lexicogrammar, and that pronunciation training has not received enough attention in many L2 classrooms (for a review, Derwing & Munro, 2015). This may be because teachers lack adequate training experience in order to provide research-based pronunciation instruction with confidence (e.g., Burri & Baker, 2019) and/or because EFL learners prioritize the accurate use of lexicogrammar over pronunciation for the purposes of successful L2 communication (e.g., Saito, 2015). It is important to remember that previous training studies have shown a logical sequence: Auditory processing can facilitate L2 phonological acquisition, when learners engage in aural input only and are guided to attend to phonological characteristics of language (e.g., Wong & Perrachione, 2007). Echoing what we found in the current study, therefore, it is unsurprising that auditory processing could determine the degree of success in L2 lexicogrammar rather than

phonology in classroom L2 speech learning because the lexicogrammar is what students primarily practice and strive to improve in classroom settings. Another possible explanation for why performance on the audio-motor integration test battery might relate to lexicogrammar rather than phonology is that the test battery required remembering and integrating information across a relatively long period of time (several seconds). Lexicogrammatical information is conveyed in speech across a longer time span than phonological information, and so auditory memory may be more crucial to the acquisition of lexicogrammar than fine-grained phonological distinctions.

Limitations and Future Directions

The current study took a first step toward examining the role of domain-general auditory processing in classroom L2 speech learning. Given the exploratory nature of the study, there are a number of methodological limitations that should be brought to light. In this section, we acknowledge these issues and call for future investigations to remedy them with a view of obtaining a full-fledged picture of the complex relationship between auditory processing and L2 speech acquisition.

An obvious limitation of the study was the relatively small number of learners involved ($N = 39$). In the current investigation, we found significant effects for auditory processing and experience on L2 lexicogrammar accuracy and fluency but not for phonological accuracy. However, the small sample put the results at greater risk for Type I or Type II error. The true presence and absence of the relationship between auditory processing, experience, and L2 speech proficiency needs to be tested with a sufficiently large sample size. The generalizability of the results should also be treated with caution. We stress that the findings presented in this study should be interpreted solely according to the particular group of L2 learners involved (undergraduate-level Vietnamese learners of English). We recommend that future replication studies use more participants with a wider range of proficiency levels (e.g., low, mid, high, and near-nativelike L2 proficiency; proposed by Abrahamsson & Hyltenstam, 2009), classroom experience (e.g., language vs. content-based classes; Saito & Hanzawa, 2018), and L1–L2 pairings (e.g., linguistically close vs. distant; McAllister, Flege, & Piske, 2002). To further broaden understanding of this relationship, future research should also feature different speaking tasks (formal vs. informal; Crowther et al., 2015), speech analysis techniques (e.g., acoustic vs. rater judgments; Saito & Plonsky, 2019), and auditory processing instruments (e.g., explicit vs. implicit; Saito et al., 2019).

Furthermore, due to the small sample size, it is important to stress that any conclusions regarding the constructs of the auditory processing that we proposed and adopted in the current study are tentative. The results showed that the strength of the relationship between acuity and integration scores was not statistically significant, suggesting the two represent independent constructs (as argued by Tierney & Kraus, 2014). Given that the distinction between spectral and temporal processing reached statistical significance for acuity but not integration, we are hesitant to make any conclusive remarks on the conceptual overlap between spectral versus temporal processing, especially in light of mixed findings in the previous literature (e.g., $r = .05$ in Kempe et al., 2015, vs. $r = .43$ in the current study). It is interesting that our participants' individual differences in integration (but not acuity) appeared to be related to L2 classroom learning experience. This finding is in line with previous research showing that audio-motor integration improves as a function of language and music learning experience (Tierney et al., 2008). In contrast, research has shown that auditory acuity declines as a function of chronological age (i.e., perceptual aging; Skoe et al., 2015), and that practice effects could be considered minor at best (Saito et al., 2020a). Saito et al. (2020b) presented more methodological recommendations.

The third limitation is the possibility—one that we cannot at present rule out—that the auditory processing tasks (reproduction, discrimination) used in the study may have conflated a range of modality-general executive function skills (e.g., attentional control, processing speed, short- and long-term memory). Although we have demonstrated links between auditory perception and L2 speech learning, the extent to which individual differences in auditory processing are distinguishable from variability in higher order cognitive abilities upon which auditory perception may draw is still unclear. This issue aligns with concerns also present in the L1 acquisition literature about the construct validity of auditory processing tests (e.g., Snowling et al., 2018). For example, the audio-motor integration task required the participants to selectively attend to and store melodic and rhythmic sequences for a short period of time in the brain and then to reproduce them with good motor control. There is some research evidence that L2 speech acquisition may be mediated by various components of cognitive abilities (Darcy, Mora, & Daidone, 2016, for inhibitory control; O'Brien, Segalowitz, Collentine, & Freed, 2006, for phonological short-term memory; Reiterer et al., 2011, for working memory). It would be interesting to further examine whether the relationship between audio-motor integration and classroom L2 speech learning remains significant even after participants' phonological short-term memory and processing speed are factored out.

Future studies should adopt both auditory processing and cognitive measures within the same research design so as to check the degree of independence between auditory processing and cognitive abilities and to investigate the separate effects of auditory processing and cognitive abilities on the process and product of learning. (cf. Zheng, Saito, & Tierney, in press for the relative weights of auditory processing and music aptitude in L2 speech learning).

Finally, we need to acknowledge that we collected all the auditory processing data online rather than via in face-to-face meetings. Although we made efforts to ensure that the participants followed the procedure and completed the test in a quiet room, three out of 42 participants who originally joined the current study had to be eliminated due to their confusion and to technical difficulties (i.e., less than 10% of attrition). In the current climate of the COVID-19 pandemic, researchers have been urgently encouraged to avoid face-to-face meetings and to collect data online. We strongly believe that more studies are needed not only to examine the reliability and validity of the online auditory processing tests but also to improve such online data collection platforms. As we reported earlier, the test–retest reliability of the online auditory processing tests (for reproduction and discrimination) was somewhat varied ($r_s = .284-.907$). These results are different from those reported by previous cognitive psychology literature for the reliability of the auditory processing measures in laboratory settings ($r = .75$ in Raz et al., 1987). This indicates that, although the task format of the auditory processing tests has been well accepted (see also Moore, 2012, for an overview of auditory processing test formats in L1 and hearing research), the possibility of delivering the test online remains open to further discussion, validation, and refinement. We stress again that the results of the reliability analyses derived from our small-scale pilot research (Saito et al., 2020b, for 30 L1 and L2 English speakers). In order to better establish the presence or absence of satisfactory test–retest reliability, we plan to redo the analyses with a larger sample size with greater statistical power. As a reviewer pointed out, another reason for the inherent difficulties of online testing is technological in nature. That is, control cannot be maintained over stimulus loudness across participants, given that they are using different hardware and sound settings, both of which could contribute to the range of test reliability observed. More work is needed on how to help deliver identical test settings to participants regardless of their contexts (see Nagle, 2019, for his interesting reliability and validation study on the implementation of online L2 speech ratings and analyses via Amazon Mechanical Turk).

Conclusion

Research to date has shown that auditory acuity and integration are primary determinants of L1 acquisition (e.g., Tierney & Kraus, 2014) and of L2 phonological and lexicogrammar acquisition in naturalistic settings (e.g., Kachlicka et al., 2019; Saito et al., 2020a). The current study extended this line of work to a classroom foreign language setting, showing that auditory processing effects were limited to specific dimensions of auditory processing (integration) and speech (lexicogrammar). These findings may reflect how the participants in the current study (Vietnamese EFL students) usually practice the target language (e.g., through production-based grammar translation practice) and the lack of authentic input that is typical of this setting (which probably impedes the development/refinement of auditory acuity). All in all, the study offers broad support for an audition-based account of language learning whereby domain-general auditory processing could be an important source of individual differences in language learning throughout life (Goswami et al., 2011; Mueller et al., 2012; Tierney & Kraus, 2017). However, we add that the type of learning experience (i.e., naturalistic vs. classroom) could influence which auditory processing abilities learners draw on (integration and/or acuity) and which dimensions of language rely on auditory processing (phonology vs. lexicogrammar).

Final revised version accepted 7 October 2020

Open Research Badges



This article has earned Open Data and Open Materials badges for making publicly available the digitally-shareable data and the components of the research methods needed to reproduce the reported procedure and results. All data and materials that the authors have used and have the right to share are available at <https://www.iris-database.org/>. All proprietary materials have been precisely identified in the manuscript.

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. <https://doi.org/10.1111/j.1467-9922.2009.00507.x>
- Appel, R., Trofimovich, P., Saito, K., Isaacs, T., & Webb, S. (2019). Lexical aspects of comprehensibility and nativeness from the perspective of native-speaking English raters. *ITL-International Journal of Applied Linguistics, 170*, 24–52. <https://doi.org/10.1075/itl.17026.app>

- Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh, UK: Edinburgh University Press.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. Retrieved from <https://www.jstatsoft.org/article/view/v067i01>
- Bokander, L., & Bylund, E. (2020). Probing the internal validity of the LLAMA language aptitude tests. *Language Learning*, *70*, 11–47. <https://doi.org/10.1111/lang.12368>
- Burri, M., & Baker, A. (2019). “I never imagined” pronunciation as “such an interesting thing”: Student teacher perception of innovative practices. *International Journal of Applied Linguistics*, *29*, 95–108. <https://doi.org/10.1111/ijal.12247>
- Casillas, J. V. (2020). The longitudinal development of fine-phonetic detail: Stop production in a domestic immersion program. *Language Learning*, *70*, 768–806. <https://doi.org/10.1111/lang.12392>
- Campbell, K. L., & Tyler, L. K. (2018). Language-related domain-specific and domain-general systems in the human brain. *Current Opinion in Behavioral Sciences*, *21*, 132–137. <https://doi.org/10.1016/j.cobeha.2018.04.008>
- Chládková, K., & Paillereau, N. (2020). The what and when of universal perception: A review of early speech sound acquisition. *Language Learning*, *70*, 1136–1182. <https://doi.org/10.1111/lang.12422>
- Chládková, K., & Šimáčková, Š. (in press). Distributional learning of speech sounds: An exploratory study into the effects of prior language experience. *Language Learning*, *70*, <https://doi.org/10.1111/lang.12432>
- Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *The Journal of the Acoustical Society of America*, *131*, 4756–4769. <https://doi.org/10.1121/1.4714355>
- Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015). Does a speaking task affect second language comprehensibility? *Modern Language Journal*, *99*, 80–95. <https://doi.org/10.1111/modl.12185>
- Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*, 218–236. [https://doi.org/10.1016/0749-596X\(92\)90012-M](https://doi.org/10.1016/0749-596X(92)90012-M)
- Darcy, I., Mora, J. C., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*, *66*, 741–773. <https://doi.org/10.1111/lang.12161>
- Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, *14*, 159–190. <https://doi.org/10.1075/ijcl.14.2.02dav>
- 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*, 163–185. <https://doi.org/10.1111/lang.12000>

- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam, Netherlands: John Benjamins.
- Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*, 533–557. <https://doi.org/10.1017/S0272263109990015>
- Doughty, C. J. (2019). Cognitive language aptitude. *Language Learning*, *69*, 101–126. <https://doi.org/10.1111/lang.12322>
- Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*, 1–24. <https://doi.org/10.1093/applin/aml015>
- Flaugnacco, E., Lopez, L., Terribili, C., Zoia, S., Buda, S., & Tilli, S. (2014). Rhythm perception and production predict reading abilities in developmental dyslexia. *Frontiers in Human Neuroscience*, *8*, <https://www.frontiersin.org/articles/10.3389/fnhum.2014.00392/full>
- Flège, J. (2016, June). *The role of phonetic category formation in second language speech acquisition*. Plenary address delivered at New Sounds, Aarhus, Denmark.
- Flège, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English/r/and/l/accurately. *Language and Speech*, *38*, 25–55. <https://doi.org/10.1177/002383099503800102>
- Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O., & Poeppel, D. (2019). Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nature Human Behaviour*, *3*, 393–405. Retrieved from <https://www.nature.com/articles/s41562-019-0548-z>
- 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. <https://doi.org/10.1017/S0267190515000082>
- 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. <https://doi.org/10.1111/lang.12225>
- Gervain, J., & Werker, J. F. (2008). How infant speech perception contributes to language acquisition. *Language and Linguistics Compass*, *2*, 1149–1170.
- 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. <https://doi.org/10.1111/j.1749-818X.2008.00089.x>
- Graham, S., Courtney, L., Marinis, T., & Tonkyn, A. (2017). Early language learning: The impact of teaching and teacher factors. *Language Learning*, *67*, 922–958. <https://doi.org/10.1111/lang.12251>
- Granena, G., & Yilmaz, Y. (2019). Corrective feedback and the role of implicit sequence-learning ability in L2 online performance. *Language Learning*, *69*, 127–156. <https://doi.org/10.1111/lang.12319>

- 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. *Proceedings of the Chicago Linguistic Society*, 35, 151–166.
- Halliday, L. F., & Bishop, D. V. (2006). Auditory frequency discrimination in children with dyslexia. *Journal of Research in Reading*, 29, 213–228.
<https://doi.org/10.1121/1.2890749>
- Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences*, 115, 1487–1492.
<https://doi.org/10.1073/pnas.1713975115>
- Hu, X., Ackermann, H., Martin, J. A., Erb, M., Winkler, S., & Reiterer, S. M. (2013). Language aptitude for pronunciation in advanced second language (L2) learners: Behavioural predictors and neural substrates. *Brain and Language*, 127, 366–376.
<https://doi.org/10.1016/j.bandl.2012.11.006>
- Jaekel, N., Schurig, M., Florian, M., & Ritter, M. (2017). From early starters to late finishers? A longitudinal study of early foreign language learning in school. *Language Learning*, 67, 631–664. <https://doi.org/10.1111/lang.12242>
- Joanisse, M. F., & Seidenberg, M. S. (1998). Specific language impairment: A deficit in grammar or processing? *Trends in Cognitive Sciences*, 2, 240–247.
[https://doi.org/10.1016/S1364-6613\(98\)01186-3](https://doi.org/10.1016/S1364-6613(98)01186-3)
- Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and Language*, 192, 15–24.
<https://doi.org/10.1016/j.bandl.2019.02.004>
- Kempe, V., Bublitz, D., & Brooks, P. J. (2015). Musical ability and non-native speech-sound processing are linked through acuity to pitch and spectral information. *British Journal of Psychology*, 106, 349–366.
<https://doi.org/10.1111/bjop.12092>
- Koizumi, R., & In'nami, Y. (2012). Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System*, 40, 554–564.
<https://doi.org/10.1016/j.system.2012.10.012>
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Science*, 97, 11850–11857. <https://doi.org/10.1073/pnas.97.22.11850>
- Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49, 757–786.
<https://doi.org/10.1002/tesq.194>
- Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, 39, 167–196.
<https://doi.org/10.1017/S0272263116000085>

- Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*, 202–220. <https://doi.org/10.1177/1094428105284919>
- Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research, 24*, 35–63. <https://doi.org/10.1177/0267658307082981>
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York, NY: Routledge.
- Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America, 128*, 3757–3768. <https://doi.org/10.1121/1.3506351>
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*, 467–477. <https://doi.org/10.1121/1.1912375>
- Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition, 38*, 801–842. <https://doi.org/10.1017/S027226311500042X>
- Li, M., & DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition, 39*, 593–620. <https://doi.org/10.1017/S0272263116000358>
- Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., ... Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning, 63*, 530–566. <https://doi.org/10.1111/lang.12011>
- Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *The Quarterly Journal of Experimental Psychology Section A, 45*, 73–87. <https://doi.org/10.1080/14640749208401316>
- McArthur, G. M., & Bishop, D. V. (2005). Speech and non-speech processing in people with specific language impairment: A behavioural and electrophysiological study. *Brain and Language, 94*, 260–273. <https://doi.org/10.1016/j.bandl.2005.01.002>
- McAllister, R., Flege, J., & Piske, T. (2002). The influence of the L1 on the acquisition of Swedish vowel quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics, 30*, 229–258. <https://doi.org/10.1006/jpho.2002.0174>
- Moore, B. C. (2012). *An introduction to the psychology of hearing*. San Diego, CA: Academic Press.
- Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly, 46*, 610–641. <https://doi.org/10.1002/tesq.34>
- 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*, 15953–15958. <https://doi.org/10.1073/pnas.1204319109>

- Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, 35, 463–482. <https://doi.org/10.1093/applin/amu024>
- Munro, M. J., Derwing, T. M., & Saito, K. (2013). English L2 vowel acquisition over seven years. In: J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 112–119). Ames, IA: Iowa State University.
- Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50, 417–528. <https://doi.org/10.1111/0023-8333.00136>
- Nagle, C. L. (2019). Developing and validating a methodology for crowdsourcing L2 speech ratings in Amazon Mechanical Turk. *Journal of Second Language Pronunciation*, 5, 294–323. <https://doi.org/10.1075/jslp.18016.nag>
- Nagle, C. L. (2020). Revisiting perception–production relationships: Exploring a new approach to investigate perception as a time-varying predictor. *Language Learning*, 70, <https://doi.org/10.1111/lang.12431>
- Nguyen, H. T. M. (2017). *Models of mentoring in language teacher education*. Cham, Switzerland: Springer.
- O’Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics*, 27, 377–402. <https://doi.org/10.1017/S0142716406060322>
- Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. E. (2011). Human voice recognition depends on language ability. *Science*, 333, 595. <https://doi.org/10.1126/science.1207327>
- Pelzl, E., Lau, E. F., Jackson, S. R., Guo, T., & Gor, K. (2020). Behavioral and neural responses to tone errors in foreign-accented mandarin. *Language Learning*, 70, <https://doi.org/10.1111/lang.12438>
- Peters, E., Noreillie, A.-S., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to foreign language input on learners’ vocabulary knowledge in two languages. *Language Learning*, 69, 747–782. <https://doi.org/10.1111/lang.12351>
- Pfenninger, S. E., & Singleton, D. (2019). Starting age overshadowed: The primacy of differential environmental and family support effects on second language attainment in an instructional context. *Language Learning*, 69, 207–234. <https://doi.org/10.1111/lang.12318>
- Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R^2 values. *The Modern Language Journal*, 102, 713–731. <https://doi.org/10.1111/modl.12509>
- Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception: An Interdisciplinary Journal*, 2, 411–440. <https://doi.org/10.2307/40285311>

- Raz, N., Willerman, L., & Yama, M. (1987). On sense and senses: Intelligence and auditory information processing. *Personality and Individual Differences*, 8, 201–210. [https://doi.org/10.1016/0191-8869\(87\)90175-9](https://doi.org/10.1016/0191-8869(87)90175-9)
- Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., ... Ackermann, H. (2011). Individual differences in audio-vocal speech imitation aptitude in late bilinguals: Functional neuro-imaging and brain morphology. *Frontiers in Psychology*, 2, 1–12. <https://doi.org/10.3389/fpsyg.2011.00271>
- Révész, A., Ekiert, M., & Torgersen, E. N. (2016). The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics*, 37, 828–848.
- Rosen, S., & Manganari, E. (2001). Is there a relationship between speech and nonspeech auditory processing in children with dyslexia? *Journal of Speech, Language, and Hearing Research*, 44, 720–736. [https://doi.org/10.1044/1092-4388\(2001/057\)](https://doi.org/10.1044/1092-4388(2001/057))
- R Core Team. (2018). *R: A language and environment for statistical computing [Computer software]*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Saito, K. (2015). How do non-native speakers perceive the role of phonology and lexicogrammar in improved comprehensibility? *Bunka Ronshu*, 46, 167–179.
- Saito, K. (2019). To what extent does long-term foreign language education improve spoken second language lexical proficiency? *TESOL Quarterly*, 53, 82–107. <https://doi.org/10.1002/tesq.468>
- 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. <https://doi.org/10.1111/lang.12387>
- Saito, K., & Hanzawa, K. (2016). Developing second language oral ability in foreign language classrooms: The role of the length and focus of instruction and individual differences. *Applied Psycholinguistics*, 37, 813–840. <https://doi.org/10.1017/S0142716415000259>
- Saito, K., & Hanzawa, K. (2018). The role of input in second language oral ability development in foreign language classrooms: A longitudinal study. *Language Teaching Research*, 22, 398–417. <https://doi.org/10.1177/1362168816679030>
- Saito, K., Kachlicka, M., Sun, H., & Tierney, A. (2020). 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. <https://doi.org/10.1016/j.jml.2020.104168>
- Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, 69, 652–708. <https://doi.org/10.1111/lang.12345>
- Saito, K., Sun, H., & Tierney, A. (2020a). Domain-general auditory processing determines success in second language pronunciation learning in adulthood: A

- longitudinal study. *Applied Psycholinguistics*, *41*, 1083–1112.
<https://doi.org/10.1017/S0142716420000491>
- Saito, K., Sun, H., Kachlicka, M., Alayo, J., Nakata, N., & Tierney, A. (in press). Domain-general auditory processing explains multiple dimensions of L2 acquisition in adulthood. *Studies in Second Language Acquisition*,
<https://doi.org/10.1017/S0272263120000467>
- Saito, K., Sun, H., & Tierney, A. (2020b). Brief report: Test-retest reliability of explicit auditory processing measures. *bioRxiv*. <https://doi.org/10.1101/2020.06.12.149484>
- Saito, K., Suzukida, Y., & Sun, H. (2019). Aptitude, experience and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition*, *41*, 201–225.
<https://doi.org/10.1017/S0272263117000432>
- Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*, 439–462.
<https://doi.org/10.1093/applin/amv047>
- Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language Learning*, *63*, 296–329. <https://doi.org/10.1111/lang.12001>
- Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in L2 processing and acquisition* (pp. 15–38). Amsterdam, Netherlands: John Benjamins.
- Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2015). Stability and plasticity of auditory brainstem function across the lifespan. *Cerebral Cortex*, *25*, 1415–1426.
<https://doi.org/10.1093/cercor/bht311>
- Smith, J. O. (2007). *Introduction to digital filters with audio applications*. Gardena, CA: W3K Publishing.
- 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*, 1270–1282. <https://doi.org/10.1177/0956797618763090>
- Sun, H., Saito, K., & Tierney, A. (in press). Domain-general auditory processing and L2 segmental and prosody acquisition: A longitudinal study. *Studies in Second Language Acquisition*,
- Surprenant, A. M., & Watson, C. S. (2001). Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *The Journal of the Acoustical Society of America*, *110*, 2085–2095. <https://doi.org/10.1121/1.1404973>
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–277). Amsterdam, Netherlands: Benjamins.

- Tavakoli, P., & Uchihara, T. (2020). To what extent are multiword sequences associated with oral fluency? *Language Learning*, *70*, 506–547. <https://doi.org/10.1111/lang.12384>
- Tierney, A.T., Bergeson-Dana, T.R., & Pisoni, D.B. (2008). Effects of early musical experience on auditory sequence memory. *Empirical musicology review: EMR*, *3*, 178. <https://doi.org/10.18061/1811/35989>
- Tierney, A., & Kraus, N. (2014). Auditory-motor entrainment and phonological skills: Precise auditory timing hypothesis (PATH). *Frontiers in Human Neuroscience*, *8*, 1–9. <https://doi.org/10.3389/fnhum.2014.00949>
- Tierney, A., & Kraus, N. (2017). 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. <https://doi.org/10.1111/ejn.13171>
- Thompson, E. C., White-Schwoch, T., Tierney, A., & Kraus, N. (2015). Beat synchronization across the lifespan: Intersection of development and musical experience. *PLoS ONE*, *10*, 1–13. <https://doi.org/10.1371/journal.pone.0128839>
- 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. https://doi.org/10.1162/jocn_a_01092
- Thompson, E. C., White-Schwoch, T., Tierney, A., & Kraus, N. (2015). Beat synchronization across the lifespan: Intersection of development and musical experience. *PLoS ONE*, *10*, 1–13. <https://doi.org/10.1371/journal.pone.0128839>
- Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*, 905–916. <https://doi.org/10.1017/S1366728912000168>
- Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*, 565–585. <https://doi.org/10.1017/S0142716407070312>
- Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *11*, 946–953. <https://doi.org/10.1093/cercor/11.10.946>
- Zheng, C., Saito, K., & Tierney, A. (in press). Successful second language pronunciation learning is linked to domain-general auditory processing rather than music aptitude. *Second Language Research*. <https://doi.org/10.1177/0267658320978493>

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Training Materials for Phonological Accuracy.

Appendix S2. Training Materials for Lexical Accuracy.

Appendix S3. Descriptive Statistics of Raw Auditory Processing and Speech Scores.

Appendix: Accessible Summary (also publicly available at <https://oasis-database.org>)

In Foreign Classroom Contexts, Having a “Good Ear” Does Not Reliably Predict All Aspects of L2 Speech

What This Research Was About and Why It Is Important

Successful learning of a second language (L2) after puberty is characterized by great variation between individuals. One source of this variation could be, just as in first language (L1) acquisition, the extent to which learners have “a good ear,” that is, the ability to represent single acoustic dimensions (such as melody, rhythm, and frequencies). The current study set out to review existing research into the role of auditory processing in L1 and L2 acquisition, and to examine the generalizability of the research to date to *classroom* L2 learners. With respect to adult L2 learners in *naturalistic* (non-classroom) settings, our review showed that learners with more precise auditory processing abilities can benefit more from conversation opportunities, leading to better L2 speech (more accurate, fluent, and complex use of language). However, the findings of the current study only *partially* supported the idea that auditory processing has an influential role in *classroom* settings.

What the Researchers Did

- We recruited 39 Vietnamese adult learners who had studied English for a few hours a week, for an average of 11 years, in the classroom (i.e., without any naturalistic experience of the language overseas).
- We elicited their spontaneous L2 English speech via an oral interview, and analyzed the accuracy and fluency of their phonology (pronunciation) and lexicogrammar (vocabulary and grammar).
- We tested two types of auditory processing abilities: (a) acuity (the ability to *discriminate* sounds so as to be able to establish meaningful categories of the sounds of the language) and (b) integration (the ability to convert what you hear into speech motor actions so as to be able *reproduce* those sounds).
- We examined the importance of (a) the amount of prior L2 English instruction and (b) their two auditory processing abilities in determining how good their L2 speech was (the accuracy and fluency of their phonology and lexicogrammar).

What the Researchers Found

- Those who attained higher levels of L2 speech tended to have not only more extensive foreign language learning experience, but also more precise auditory processing abilities, relative to those with lower levels of L2 speech.
- However, a reliable link between auditory processing and L2 speech was found only for one type of auditory processing (“integration” [reproduction], not “acuity” [discrimination]) and for one measure of L2 speech (i.e., lexicogrammar, not phonology).
- This finding was different to previous findings with learners from naturalistic settings, where both types of auditory processing (acuity and integration) can be associated with both aspects of L2 speech (phonology and lexicogrammar).

Things to Consider

- These findings may reflect how the participants in the current study (Vietnamese EFL students) usually practice the language (e.g., through production-based grammar translation), and their lack of authentic input, which could impede the refinement of auditory acuity.
- The type of learning experience (i.e., naturalistic vs. classroom) could influence which type of auditory processing abilities (integration and/or acuity) learners draw on, and which aspects of language (phonology and/or lexicogrammar) come to rely on auditory processing capacity.
- Given that L1 acquisition researchers have shown that training can improve auditory processing, future studies could investigate the extent to which training auditory processing (versus teaching the language itself) can boost the rate and ultimate attainment of L2 learning in various settings.

Materials, data, open access article: Materials and data are publicly available at <https://www.iris-database.org>.

How to cite this summary: Saito, K. (2021). In foreign classroom contexts, having a “good ear” does not reliably predict all aspects of L2 speech. *OASIS Summary* of Saito, Suzukida, Tran, Tierney. (2021) in *Language Learning*. <https://oasis-database.org>

This summary has a CC BY-NC-SA license.