

## How Does Having a Good Ear Promote Successful Second Language Speech Acquisition in Adulthood? Introducing Auditory Precision Hypothesis-L2

Kazuya Saito<sup>1</sup>, University College London, [k.saito@ucl.ac.uk](mailto:k.saito@ucl.ac.uk)

### Abstract

In this paper, I first provide a brief review of how scholars have conceptualized, tested, and elaborated aptitude frameworks relevant to L2 speech learning. Specifically, I highlight an emerging paradigm that assigns a fundamental role to domain-general auditory processing (i.e., having a good ear) in L1 speech acquisition and propose that the same faculty acts as a cornerstone of L2 speech learning (i.e., the Auditory Precision Hypothesis-L2). This hypothesis predicts that learners with more precise auditory processing ability will be able to make the most of every input opportunity, which will result in more advanced L2 speech proficiency. To close, I will provide suggestions on how scholars can assess L2 students' auditory processing ability (e.g., our team's offline test deposited at L2 Speech Tools for Researchers & Teachers [<http://sla-speech-tools.com/>]) and discuss how the results can be used to maximize learners' L2 speech learning opportunities via optimal, profile-matched training programs (e.g., explicit vs. incidental training; naturalistic vs. classroom learning; phonetic vs. auditory training).

---

<sup>1</sup> This paper originates from my plenary speech delivered at EuroSLA 30 (Barcelona, Spain). I am grateful to the following team members for their huge contributions to all the foundation projects that we worked on together: Adam Tierney, Hui Sun, Magdalena Kachlicka, Yui Suzukida, Ingrid Mora-Plaza, & Katya Petrova. I also greatly acknowledge the funders who have supported our research activities: Leverhulme Trust (RPG-2019-039), Spencer Foundation (202100074), and Economic and Social Research Council (ES/S013024/1).

## Introduction

Learning speech in a second language (L2) after puberty is a difficult task that is characterized by a great deal of individual variation. Some learners can achieve high-level L2 oral proficiency while others have a tremendous amount of difficulty doing so. These differences could be due to not only the amount of time spent practicing the target language but also to learners' ability to make the most of every opportunity for input via a range of perceptual and cognitive abilities relevant to L2 acquisition (i.e., *aptitude*; Wen & Skehan, 2021). Examining these abilities can provide a deeper theoretical understanding of the mechanisms that underlie and drive how learners process input, convert it into intake, and acquire a new language. There has been much debate about whether such mechanisms are specific to language learning or generalizable across various kinds of learning behaviours (domain-specific vs. domain-general; e.g., Hamrick, Lum, & Ullman, 2018); and whether they differ in first language (L1) and L2 acquisition (the degree of awareness; e.g., Diaz, Mitterer, Broersma, Escera, & Sebastian-Galles, 2016). An examination of this topic also has considerable pedagogical relevance. An understanding of individual aptitude profiles could help teachers identify students who would likely benefit more from certain types of instructional approaches. For example, those with explicit language learning aptitude would likely benefit more from a language-focused approach (e.g., metalinguistic instruction), while those with stronger implicit learning aptitude may benefit from implicit and meaning-oriented instruction (i.e., aptitude-treatment interaction; DeKeyser, 2012).

In this paper, I will first briefly review a range of aptitude frameworks relevant to L2 speech learning and then introduce an emerging paradigm that holds that having a good ear<sup>2</sup> (i.e., auditory processing precision) serves as an anchor of L1 acquisition and L2 speech learning in adulthood (i.e., the Auditory Precision Hypothesis-L2). Auditory processing is a complex of domain-general perception abilities related to encoding the acoustic characteristics of sounds. Since auditory processing is the first ability that learners rely on to extract linguistic information from spoken input, any individual differences in this ability are thought to affect various dimensions (segmentals, suprasegmentals, vocabulary, morphosyntax) and phases (speed of learning, ultimate attainment) of language learning. Finally, I will discuss how we can assess L2 students' auditory processing ability (e.g., our team's offline test available at <http://sla-speech-tools.com/>) and make a range of pedagogical suggestions about how such assessments could be used to provide more effective instruction. Following the aptitude-treatment interaction

---

<sup>2</sup> "Having a good ear" is listed in the Cambridge Dictionary as a commonly used expression meaning "good at hearing, repeating, and understanding...sounds [of music and language]." This expression is often used when discussing what is needed for the attainment of advanced L2 pronunciation proficiency. Given that this paper was written for *Language Teaching*, whose readers include practitioners as well as researchers, having "a good ear" was included in the title (and elsewhere) to ensure that the relatively novel and highly complex subject matter was accessible to the entire readership. Having said that, I must stress that the dichotomy implicated in this expression (i.e., good vs. bad) has been problematized in L1 acquisition literature. While many tasks have thresholds for the diagnosis of auditory processing disorders in various populations (Moore, 2006), it has been suggested that the nature of the link between auditory processing and language learning could be better characterized as a "spectrum" rather than a "dichotomy" (for a critical review on considerable heterogeneity and variability in the operationalization of auditory processing, see Protopapas, 2014). While some children with certain language impairments (e.g., dyslexia) may have less precise auditory processing, the degree of auditory precision and language learning is still subject to a great deal of individual variation even among so-called normal hearing children (Kalashnikova et al., 2019).

paradigm, I will explain how L2 learners with diverse aptitude profiles (explicit vs. implicit; acuity vs. integration; strong vs. poor) can be encouraged to understand, speak, and master their L2 through profile-matched training programs (explicit vs. incidental; naturalistic vs. classroom; phonetic vs. auditory).

### **What Is L2 Speech Learning?**

According to Saito and Plonsky's (2019) framework, L2 speech proficiency comprises (a) the ability to perceive and produce novel (or partially acquired) consonantal and vocalic sounds in an L2 without deleting and substituting them for L1 counterparts (i.e., *segmental* proficiency); (b) the ability to use adequate and varied stress (characterized by longer, louder, and higher pitch) at the word (e.g., correct assignment of word stress) and sentence (appropriate use of intonation for declarative and interrogative intensions) levels (i.e., *melodic* and *prosodic* proficiency); and (c) the ability to deliver speech at an optimal tempo without making too many pauses or repetitions/self-corrections (*rhythmic* and *temporal* proficiency). The last two dimensions have often been collectively described as "*suprasegmental* proficiency" (Trofimovich & Baker, 2006). The development of precise, robust, and refined L2 segmental and suprasegmental representations is fundamental for reaching advanced levels of listening (Field, 2008) and speaking proficiency (Levis, 2006). With solid L2 segmental and suprasegmental representations, L2 learners can more easily process phonologically similar and complex words (Saito, 2013), perceptually non-salient morphosyntactic markers (Goldschneider & DeKeyser, 2001), and a range of discourse functions (Brazil, 1997), all of which underpin successful oral communication (Isaacs, Trofimovich, & Foote, 2018).

Scholars have extensively examined how learners' L1 phonetic systems influence their L2 speech acquisition. Major frameworks addressing this topic include the Speech Learning Model (Flege & Bohn, 2021), the Perceptual Assimilation Model (Best & Tyler, 2007), Structural Conformity Hypothesis (Eckman, 2004), and the Optimality-Theoretic Model (Escudero & Boersma, 2004). These theoretical accounts share the view that the phonetic distance between the L1 and L2 systems is partially responsible for determining the degree of speech learning difficulty. For example, very few Japanese speakers can perceive and produce English [r] and [l] contrast at a nativelike level because the relevant auditory and articulatory cues are not actively used in the L1 system (third formant frequencies and labial, alveolar, and pharyngeal constrictions; Iverson et al., 2003).

Another line of research has explored which individual difference factors predict advanced L2 speech acquisition. For instance, a wide body of research suggests that factors related to both the quantity (how much learners are exposed to and practice a target language) and quality (with whom learners use a target language [L1 vs. L2 users]), and timing of language experience (how early participants have started learning a target language and have arrived at an L2 speaking country; e.g., Derwing & Munro, 2013) are related to L2 speech learning outcomes.

However, research has shown that experience factors alone cannot fully explain the variability in ultimate L2 speech attainment. In one of my projects, for example, I examined the accuracy of English [r] production among approximately 200 L1 Japanese L2 English late bilinguals in Canada (Saito, 2015). All participants had an extensive amount of immersion experience (length of residence > 6 years), had arrived in Canada after puberty, and used their L2 (English) every day as a primary language of communication. Despite their similar backgrounds and overall speaking proficiency, analysis of the participants' performance on word reading and picture description tasks suggested that the degree of their English [r] pronunciation attainment

widely varied—some demonstrated nativelike pronunciation while others had detectable L1 accents.

One hypothesized source of the individual variation observed in Saito (2015) and other studies (e.g., Abrahamsson & Hyltenstam, 2008) is *aptitude*—a talent for processing the L2 input more efficiently and/or effectively, resulting in larger learning gains in the long run (Doughty, 2019). But what characterizes aptitude for successful L2 speech learning? To answer this question, I will first provide a selective review on the role of aptitude in adult L2 speech learning.

### **What is L2 Speech Aptitude?**

Fifty years of research have provided evidence that aptitude plays an important role in L2 vocabulary and morphosyntax learning (for comprehensive overviews, see Li, 2016; Wen & Skehan, 2021). This has led to the development of thorough conceptual and methodological aptitude frameworks, such as the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959), LLAMA (Meara, 2005), and Hi-LAB (Linck et al., 2013). Although the existing aptitude tests do include audio materials (e.g., sound recognition in LLAMA-D) and refer to their relevance to speech learning on a broad level (e.g., phonemic coding in MLAT for “oral proficiency”; Baker Smemoe & Haslam, 2013), very few studies have examined in depth to what degree, how, and why their specific aptitude tests tap into the development of segmental, melodic, and temporal proficiency. In their focused review, Trofimovich, Kennedy, and Foote (2015) pointed out that “there has been little systematic research on the relationship between various components of aptitude and L2 pronunciation learning” (p. 354).

To fill this gap, I surveyed a range of studies on this topic (i.e., aptitude and speech learning), focusing particularly on those published since Trofimovich et al.’s (2015) call for further research (Table 1). Skehan (2016) provided a set of useful frameworks that researchers can use to survey and categorize different types of L2 aptitude. The frameworks include: (a) linguistic focus (which dimensions of language is the aptitude related to?), (b) domain generality (is the aptitude specific to L2 learning or applicable to all learning behaviors?), and (c) explicitness (is the aptitude associated with explicit or implicit learning?). Accordingly, the following criteria were considered in the creation of a framework for L2 speech learning aptitude:

- Is the aptitude relevant to segmental learning (enhancing consonant and vocalic accuracy) or suprasegmental learning (melody and rhythm)?
- Is the aptitude domain-specific or domain-general?
- Is the aptitude associated with explicit or implicit learning?

**Table 1**  
Summary of Perceptual and Cognitive Abilities Relevant to L2 Speech Learning

Proficiency	Domain	Abilities	Type	Measures	Examples
Segmental proficiency	Specific	Phonemic coding	Explicit	LLAMA-E, Modern Language Aptitude Test (MLAT)	Hu et al. (2013); Saito (2017)
		Sound sequence learning	Implicit	LLAMA-D	Granena (2019); Saito et al. (2019)
	General	Working memory	Explicit	Forward span, backward span, nonword repetition	Darcy et al. (2016); Darcy & Mora (2017)
		Procedural memory	Implicit	Serial reaction time, systematic multimodal associations reaction	Granena (2019); Gabay et al. (2015)
Suprasegmental proficiency	Specific	Rote memorization	Explicit	LLAMA-B, MLAT	Saito (2017); Silbert et al. (2015)
	General	Tonal and rhythm imagery	Explicit	Music Aptitude Profile	Slevc & Miyake (2006)
		Procedural memory	Implicit	Serial reaction time, systematic multimodal associations reaction	Granena (2019); Gabay et al. (2015)

*Note.* Phonemic coding refers to sound-symbol correspondence (featured in Carroll and Sapon's MLAT framework); tonal and rhythm imagery refers to sensitivity to differences in melody and rhythm (featured in Gordon's [1995] notion of Music Aptitude Profile)

As summarized in Table 1, though limited in number, a growing number of studies have explored the relationship between aptitude and L2 speech learning outcomes. The results of this body of work have shown that different types of aptitude (explicit vs. implicit; domain-general vs. -specific) uniquely relate to different areas of L2 speech learning (segmental vs. suprasegmental learning).

### **Auditory Processing as an Emerging Aptitude Framework**

More recently, some scholars (including our team) have begun to conceptualize, test, and elaborate on an aptitude framework based on a very simple hypothesis: that having a “good ear” (i.e., domain-general auditory processing ability) is the root of language acquisition (Mueller, Friederici, & Männel, 2012). Since auditory processing is the first ability that infants rely on to parse incoming linguistic input, the detection and interpretation of acoustic information underlies every stage of phonetic, phonological, lexical, and morphosyntactic learning and delay. Thus, it is possible that auditory processing can explain the rate and ultimate attainment of L2 acquisition as well.

Auditory processing refers to a set of lower-order abilities related to precisely perceiving individual dimension of acoustic information, such as pitch (the perception of the lowest, fundamental frequency of a sound wave), formants (acoustic energy concentrations resulting from resonance), duration (length of sounds), and intensity (loudness of sounds). Corresponding to an influential view in cognitive psychology, auditory processing can be considered domain-general, and forms the basis of multiple domain-specific phenomena, such as music, emotion, environmental sounds, and language (Kraus & Banai, 2007). To measure such domain-general abilities, a number of synthesized stimuli are prepared. Since these stimuli comprise very simple acoustic characteristics (e.g., completely flat fundamental frequencies and formant contours), normal hearing listeners will not perceive them as speech. While exposed to the *nonverbal* stimuli, participants are assessed for their abilities to precisely perceive one particular acoustic dimension (e.g., pitch, duration).

In the context of language learning, this domain-general ability is thought to play a key role in the development of phonology, vocabulary, and morphosyntax. For example, infants rely on auditory processing to detect the probabilities of individual phonemes in the L1 system within the first six to eight months of their life (Werker, 2018). During this critical period, every phoneme can be statistically defined in accordance with the different weighting of multiple acoustic cues, such as pitch (F0), first formant (F1), second formant (F2), third formant (F3), duration, and intensity (Kuhl, 2004). Auditory processing is also instrumental to the identification of word and phrase boundaries (Cutler & Butterfield, 1992), syntactic structures (Penner, Wymann, & Weissenborn, 2001), and morphosyntactic markers (Joanisse & Seidenberg, 1998; Koester et al., 2004).

In terms of development trajectory, children reach adult-like auditory processing within the first eight to ten years of life (e.g., Thompson, Cranford, & Hoyer, 1999 for pitch discrimination; Elfenbein, Small, & Davis, 1993 for duration discrimination). From the early 20s onwards, however, auditory processing gradually declines over the rest of the lifespan (Skoe et al., 2015; but see the relatively slow peak and decline curve on the development of audio-motor integration abilities, see Thomson, White-Schwoch, Tierney, & Kraus, 2015).

Based on these observations, many L1 acquisition researchers have put forth the hypothesis that auditory impairments are a source of many language problems (Goswami, 2015); that is, if someone experiences deficits in auditory processing, it immediately affects their speech perception, which could, in turn, prevent them from detecting, developing, and consolidating the

speech categories, and could lead to a range of global language problems. For example, auditory processing measures have been suggested to be a diagnostic tool for dyslexia (Hornickel & Kraus, 2013) and other language-related disorders (Russo et al., 2008).

There is ample cross-sectional and longitudinal evidence showing that auditory individual differences among normal hearing children are significantly tied to a range of L1 outcomes (e.g., speech-in-noise perception, vocabulary use, literacy, and phonological awareness) (Anvari, Trainor, Woodside, & Levy, 2002; Bavin, Grayden, Scott, & Stefanakis, 2010; Boets, Wouters, Van Wieringen, De Smedt, & Ghesquiere, 2008; Tierney, Gomez, Fedele, & Kirkham, 2021; for evidence as to how auditory processing influences L1 vocabulary development over the first 3 years of life, see Kalashnikova, Goswami, & Burnham, 2019). In addition, correlation studies have shown a medium-to-large relationship between reading difficulty and auditory deficits for various dimensions of nonverbal sounds (McArthur & Bishop, 2005 for frequency; Casini, Pech-Georgel, & Ziegler, 2018 for duration; Goswami, Fosker, Huss, Mead, & Szűcs, 2011 for amplitude rise time).

Because the auditory precision hypothesis concerns causality, it is naturally subject to a great deal of controversy. Specifically, some scholars have argued that not all dyslexic children and adults have auditory deficits (Rosen, 2003 for an overview). From a methodological point of view, it is important to remember that behavioral tasks for measuring auditory perception (e.g., A×B discrimination; for details, see below) inevitably tap into a set of higher-order executive skills (e.g., attentional control, memory), in addition to lower-order skills. For instance, the highly repetitive and abstract nature of laboratory tasks may make it difficult for participants to maintain auditory information in working memory and thus may limit how much information is available for acoustic analysis (Zhang et al., 2016). Accordingly, individuals with language impairments may perform poorly on auditory processing tasks because of problems with both auditory processing and executive functioning, which suggests that any link between auditory processing and linguistic deficits could be confounded with higher-order cognitive abilities (Gooch, Hulme, Nash, & Snowling, 2014; Henry, Messer, & Nash, 2012; Snowling et al., 2018).

### **Auditory Precision Hypothesis-L2**

More recently, researchers have begun to explore how well the auditory precision hypothesis generalizes to adult L2 speech learning (i.e., Auditory Precision Hypothesis-L2; Mueller et al., 2012). This concurs with the assumptions underlying major L2 speech theories that the mechanisms in successful L1 speech acquisition remain active throughout the lifespan, and are germane to any new speech learning experience (e.g., Flege & Bohn, 2021). In this paper and elsewhere (e.g., Saito, Sun, & Tierney, 2021), I would like to further argue that auditory processing could be particularly consequential in post-pubertal L2 learning (relative to L1 acquisition). This is arguably due to the *quantitative* and *qualitative* differences between the L1 and L2 learning experiences.

Because L1 learners are normally exposed to an extensive amount of spoken language, they may be able to overcome auditory-based difficulties via remedial strategies. For example, those with pitch deficits (amusics) can still process phrase boundaries normally using durational rather than pitch information (Jasmin, Dick, Holt, & Tierney, 2020). In contrast, the amount of communicatively authentic and interactive input that L2 learners receive is generally highly limited in classroom settings (Muñoz, 2014), and subject to a great deal of individual variation in naturalistic settings (Derwing & Munro 2013). Thus, L2 learners may have more difficulty developing a similar range of remedial strategies.

Furthermore, different from L1 acquisition, which is free of influence from prior language learning experience, L2 speakers need to encode spectro-temporal patterns through already-developed and automatized L1 perception strategies (see McAllister, Flege, & Piske, 2002 for the feature account of adult L2 speech learning). That is, to acquire new speech categories, L2 speakers need to not only adjust their already-attuned cue weighting patterns (e.g., Chinese speakers need to use both pitch and duration to perceive L2 English prosody; Jasmin, Sun, & Tierney, 2021), but also need to learn and develop new perception strategies that they do not actively use in their L1 (e.g., Japanese speakers need to discriminate variation in F3 to perceive English [r] and [l]; Iverson et al., 2003).

### **Components of Auditory Processing**

Extending several popular aptitude frameworks in SLA (Skehan, 2016) and L2 speech (see Table 1), I propose *four* different auditory process abilities particularly relevant to adult L2 speech learning. Under this  $2 \times 2$  model (see Table 2), the key distinctions concern (a) whether the abilities relate to L2 speech learning with or without awareness (i.e., explicit vs. implicit) and (b) whether the abilities concern the processing of formants or prosodic information, such as pitch, duration, and intensity. Scholars have operationalized auditory processing of duration and intensity via amplitude rise time (i.e., the time/duration from the onset of a sound to its maximum amplitude; Goswami, 2015).



**Table 2**  
 Summary of Auditory Processing Relevant to L2 Speech Learning

Proficiency	Abilities	Type	Measures	Examples
Segmental proficiency	Formant acuity	Explicit	Discrimination (F1, F2, F3)	Lengeris & Hazan (2010)
		Implicit	EEG response (F1, F2, F3)	Kachlicka et al. (2019); Saito et al. (2019, 2021); Sun et al. (2021)
Suprasegmental proficiency	Prosodic acuity	Explicit	Discrimination (F0, duration, amplitude rise time)	Qin, Zhang, & Wang (2021)
		Implicit	EEG response (F0)	Chandrasekaran, Kraus, & Wong (2012)

### **Explicit Acuity**

Explicit acuity concerns how subtle of a difference in a particular acoustic dimension (e.g., formant, pitch, duration, and intensity) learners can encode. This ability is behaviourally measured via A × B discrimination tasks, where participants hear three nonverbal sounds, one of which is different from the other two, and must indicate which sound differs. The sounds featured in this task are typically synthesized stimuli whose acoustic dimensions are identical except for one dimension. As shown in Table 2, learners' sensitivity to first, second, and third formants (F1, F2, and F3) is thought to relate to segmental learning; and their sensitivity to prosodic information (fundamental frequency [F0], duration, and amplitude rise time) is thought to relate to suprasegmental learning.

Lengeris and Hazan (2010) used this type of task to measure L1 Greek English learners' *formant* acuity. A total of 51 stimuli were developed which differed in terms of a single formant (analogous to vowel F2 = 1250-1500 Hz), and presented to participants. Those who were capable of perceiving the smaller differences in formants demonstrated more learning gains when intensively exposed to multi-talker English vowels. Similarly, Qin, Zhang, & Wang's (2010) study with 32 Mandarin learners of Cantonese found that participants with more precise pitch acuity (F0 = 100.07-178.17 Hz) benefited more from the intensive exposure to multi-talker Cantonese tones.

### **Implicit Acuity**

Implicit acuity concerns learners' ability to track a particular acoustic dimension on a *subconscious* level. Our research team has so far explored whether and to what degree auditory processing can predict the ultimate attainment of high-level L2 speech proficiency. To reach such an advanced stage of speech development, we assume that learners will need to have years of naturalistic and classroom learning experience. In addition, we assume that they will need explicit and implicit auditory processing abilities which allow them to maximize any learning opportunities, regardless of awareness. In our recent studies (Kachlicka, Saito, & Tierney, 2019; Saito, Sun, & Tierney, 2019; Saito, Kachlicka, Sun, & Tierney, 2020; Sun, Saito, & Tierney, 2021), we propose the idea of using electroencephalography (EEG) to measure how the brain tracks and reacts to the acoustic characteristics of sounds at a subcortical level (i.e., implicit acuity).

Among the many EEG paradigms in L2 speech research (e.g., Diaz et al., 2016 for a comprehensive overview), we have adopted the frequency following response (FFR) to study the subcortical auditory system (Coffey, Herholz, Chepesiuk, Baillet, & Zatorre, 2016). During FFR tasks, participants engage in a meaning-oriented activity (e.g., reading for pleasure, watching silent movies) while listening to a range of synthesized nonverbal sounds. As attention is not required in this task, FFR data can be assumed to reflect an unconscious sensitivity to certain aspects of acoustic signals (formants, pitch) without the contaminating influence of cognitive and affective states. There is a growing amount of research using FFR which has shown that those with more precise encoding of formants likely attain more advanced L2 segmental proficiency (e.g., Saito et al., 2019, 2020) and that those with more precise encoding of pitch gain more from pitch-based artificial language training (e.g., Chandrasekaran, Kraus, & Wong, 2012).

### **Empirical Evidence**

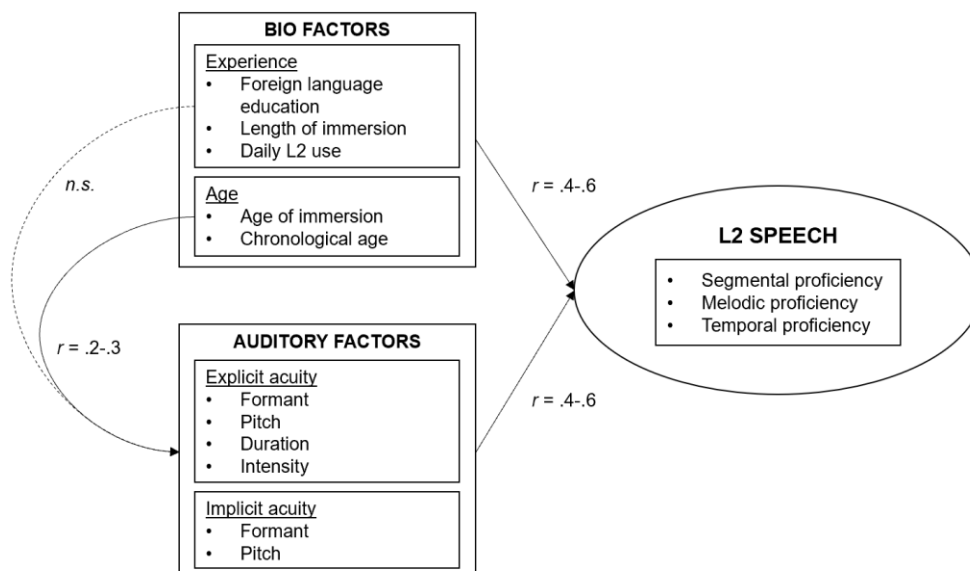
Our research team has conducted a series of cross-sectional and longitudinal projects to examine the complex relationships among auditory processing, experience, and L2 speech learning. We recruited more than 400 L2 speakers of English from Poland, Spain, China, Japan, and Vietnam who had studied in naturalistic and/or classroom conditions. Those participants

with any immersion experience (range <1 to 20 years) had arrived at an L2 country after the age of 17 (i.e., late bilinguals), assuming that they used L2 English with detectable L1-related accents. We measured participants' L2 comprehension and production proficiency via measures of segmentals, suprasegmentals, vocabulary, and morphosyntax. Next, we assessed their auditory processing profiles via behavioral and EEG measures. Finally, we surveyed their biographical backgrounds, gathering data on experience-related variables (length of foreign language education and residence, daily L1/L2 use) and age-related variables (chronological age, age of learning, and arrival).

The findings were published separately in several different papers between 2020 and 2022. Adopting cross-sectional or/and longitudinal designs, each paper linked various types of auditory processing (explicit, implicit) to different dimensions (segmentals, suprasegmentals, vocabulary, morphosyntax), modes (perception, production), and stages (early, mid, final) of L2 speech learning. By analyzing these studies as a group, it is possible to synthesize their findings in order to obtain *suggestive* patterns.

First and foremost, the results of multiple regression and mixed-effects modeling analyses showed that performance scores were *equally* associated with biographical and auditory processing factors. As visually summarized in Figure 1, half of the variance was explained by how much participants practiced a target language in a classroom setting, and how much they had been using the target language on a daily basis in immersion settings. The other half of the variance was accounted for by their auditory processing ability.

In terms of type of auditory processing, explicit auditory processing appeared to be important at every stage of adult L2 learning (e.g., Saito, Sun et al., 2020 for the longitudinal analyses of the first 1 year of immersion), while implicit auditory processing had stronger predictive power for experienced, long-term L2 residents (length of residence = 1-10 years; Kachlicka et al., 2019; Saito et al., 2019, 2020; cf. see Sun et al., 2021 for short-term residents with less than 1 year of length of residence). Interestingly, the effects of auditory processing are relatively weak among L2 learners in classroom settings (Saito, Sun et al., 2020; Saito, Suzukida, Sun, & Tierney, 2021; Saito et al., 2022). This is probably because auditory processing may be unrelated to the outcomes of classroom L2 speech learning wherein learners receive and process the limited amount of aural input (but see below the section of Different Types of Auditory Processing).



**Figure 1**

Summary of the suggestive relationship between auditory processing, biographical background, and L2 speech outcomes

Furthermore, in the context of 70 Japanese speakers of English with varied experience and proficiency levels, Saito et al. (in press-a) examined the extent to which auditory processing and cognitive abilities *interacted* to determine the rate of success in L2 speech proficiency. The results of the correlation analyses showed that all variables were equally related to L2 speech outcomes. More interestingly, the results of the factor analyses showed that auditory processing and explicit cognitive abilities (phonological short-term memory, executive functions, and declarative memory) were clustered into two different categories (see Table 3). Of course, the findings are tentative as they need to be replicated with L2 learners with different L1 backgrounds (e.g., Polish, Spanish, and Vietnamese). However, the study here at least hints at the possibility that auditory processing may be distinct from explicit cognitive abilities and instead related to implicit and procedural memory. The suggestions here support the view that the test of auditory processing may trigger implicit statistical learning of the distribution of stimuli across trials (combining the prior stimulus distribution and the acoustic representations of each incoming stimulus; Raviv, Ahissar, & Loewenstein, 2012; for a more detailed discussion on the role of implicit statistical learning in auditory processing, see Saito et al., in press-a).

**Table 3**

*Factor Analysis of Auditory Processing and Cognitive Abilities Presented Among 70 L2 Learners in Saito et al. (in press-a)*

	Factor 1: Explicit memory	Factor 2: Implicit audition
Auditory processing	-.071	<b>-.752</b>
Phonological short-term memory	<b>.627</b>	.271
Complex working memory	<b>.830</b>	.119

Declarative memory	<b>.695</b>	-.134
Procedural memory	.048	<b>.753</b>

Taken together, there are three main observations from the empirical research. First, auditory processing appears to be a relatively independent construct. Second, individual differences in auditory processing may serve as a moderate-to-strong determinant of post-pubertal L2 speech acquisition, especially if learners engage in a great deal of authentic, conversational auditory input on a daily basis. The first two observations led me to propose the last observation: that even adult L2 learners may draw on similar language learning mechanisms used for L1 acquisition, and that these have a lifelong impact on the rate and ultimate attainment of language learning throughout the lifespan (for a comprehensive summary of auditory processing in L1 and L2 acquisition research, see Saito et al., 2021).

### **Future Directions**

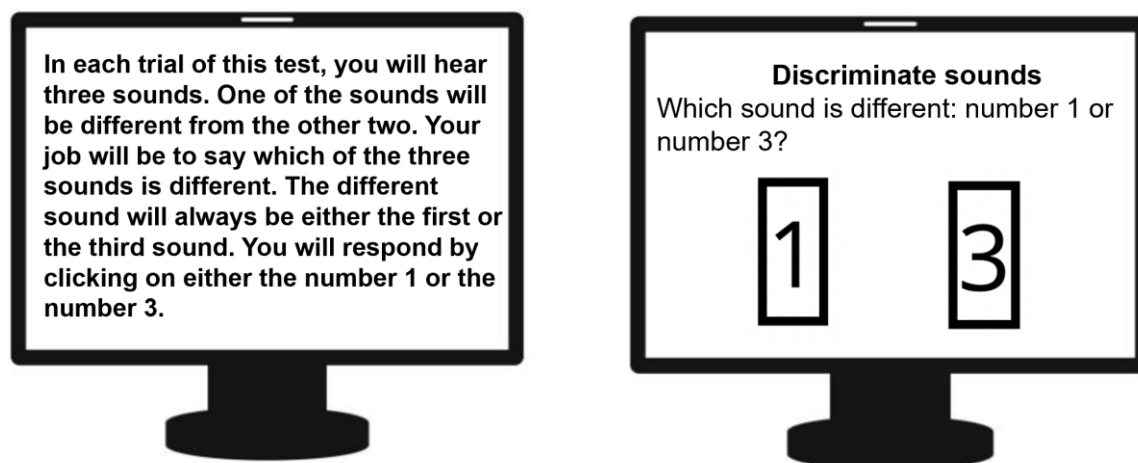
#### **Offline Test Development and Dissemination**

To facilitate follow-up studies on the role of auditory processing in L2 speech learning, our team has developed an open-source, freely available auditory processing test battery that researchers, students, and practitioners can use. The test comprises four subcomponents (formant discrimination, pitch, discrimination, duration discrimination, and amplitude rise discrimination) following an A × B discrimination task format (see Figure 2). The tasks adopt Levitt's (1971) adaptive procedure, wherein task difficulty decreases (i.e., the difference being wider) or increases (i.e., the difference being smaller) based on participants' trial-by-trial performance. Ultimately, the test allows us to measure the extent to which participants can perceive subtle differences in one of four different types of domain-general acoustic information: second formant (1500-1700 Hz), fundamental frequencies (300-360 Hz), stimulus duration (250-500 ms), and the timing of amplitude change (15-300 ms). Test materials and a user manual can be found on our team's website, where we also share and disseminate a range of research instruments and teaching materials, *Tools for Second Language Speech Research and Teaching*; <http://sla-speech-tools.com/> (Mora-Plaza, Saito, Suzukida, Dewaele, & Tierney, 2022).

Evidence for the reliability of these instruments was provided in a test-retest study with 100 L1 and L2 speakers (Saito & Tierney, forthcoming). The study found that the inter-class correlations among the different tasks could be considered “fair” to “good” ( $ICC_{(2,2)} = .4-.6$ ). This suggests that these behavioural measures can reliably tap into various dimensions of participants' supposedly stable perceptual acuity abilities (Moore, 2012). To further examine the source of individual variation among participants' auditory processing scores, future research could examine the auditory processing profiles of participants with varied biographical backgrounds (e.g., L1 vs. L2 vs. L3 speakers; classroom vs. immersion learners; tonal vs. non-tonal speakers; musicians vs. non-musicians). For instance, our tentative evidence suggests that auditory processing is relatively stable regardless of experience-related variables (e.g., length and intensity of immersion and foreign language education) but may be subject to the influence of age-related variables (e.g., Saito Kachlicka et al., 2020; Saito et al., 2022 for chronological age; Saito, Kachlicka et al., 2020 for age of arrival). Future studies on this topic will shed light on what characterizes the individual variation observed in explicit auditory processing ability.

**2A**

**2B**



**Figure 2**  
*Task Instruction (2A) and Onscreen Labels (2B)*

### **Enhancing Auditory Processing**

If auditory processing matters for L2 acquisition, one relevant question is, “Can it be enhanced via focused training?” In the L1 hearing literature, some studies have shown that a few hours of training can boost various dimensions of auditory processing among children with language disorders (Merzenich, et al., 1996 for temporal acuity; Micheyl, Delhommeau, Perrot, & Oxenham, 2006 for pitch acuity; Whiteford & Oxenham, 2018 for audio-motor integration). In turn, they can reach optimal auditory thresholds, and subsequently make the most of every input opportunity in their L1.

Following this line of work, our team’s current study examined whether domain-general auditory processing (i.e., precise representation of sounds) can be improved via focused online training and whether this affects speech learning (Saito et al., in press). 98 adult Japanese speakers were divided into two training groups targeting the acquisition of English [æ] and [ʌ]: an auditory training group and a phonetic training group. The auditory training group completed activities designed to improve their ability to use the second formant frequency (1200-1600 Hz) to discriminate between nonverbal sounds. The phonetic training group was taught to discriminate between English [æ] and [ʌ] contrasts using multi-talker speech stimuli. The results showed that the phonetic training group improved only their English [æ] and [ʌ] identification, while the auditory training group enhanced both auditory and phonetic skills. The results suggest that auditory acuity to key, domain-general acoustic cues (F2 = 1200-1600 Hz) anchors, triggers, and promotes speech learning on a domain-specific level (English [æ] vs. [ʌ]). The findings also suggest that auditory training could help remediate difficulties with L2 speech learning in some individuals with auditory deficits.

### **Different Types of Auditory Processing (Beyond Acuity)**

Thus far, auditory processing has been conceptualized as the ability to encode subtle acoustic characteristics of sounds (i.e., perceptual acuity). On a broader level, auditory processing can also comprise a range of neighboring abilities, such as attention to particular acoustic dimensions while ignoring others (i.e., auditory selective attention) and the use of acoustic information for motor action (i.e., audio-motor integration). There is emerging evidence that different types of auditory training are more or less relevant to different dimensions of L2 speech learning.

On the one hand, perceptual acuity and audio-motor integration appear to be good indices of successful L2 speech learning in *naturalistic* settings. Since such immersion experience can provide learners with ample L2 input and output opportunities, those with more precise acuity and integration can better encode the acoustic dimensions of new sounds and then integrate this information into their L2 system more efficiently and effectively. As a result, these learners can achieve more advanced L2 speech proficiency (e.g., Saito et al., 2022; Zheng et al., 2022).

On the other hand, the rate of learning success in *classroom* settings appears to be linked to audio-motor integration but *not* to perceptual acuity. In many English-as-a-Foreign-Language (EFL) classrooms, L2 learners typically learn the target language through decontextualized, production-based teaching methods (e.g., mechanical repetition and memorization of model pronunciation forms). Such learning environments do not provide an abundant amount of contextually rich, communicatively authentic input (Shintani, Ellis, & Li, 2013). Due to the asymmetry here (output > input), learners' audio-motor integration (but not perceptual acuity) has been found to impact the outcomes of classroom L2 speech learning (e.g., Saito et al., 2021 for Vietnamese EFL classrooms; Shao et al., 2022 for Chinese EFL classrooms).

### **Aptitude-Treatment Interaction**

In L2 morphosyntax learning, there is a well-researched hypothesis stating that learners with greater explicit aptitude will benefit more from explicit training, and those with greater implicit aptitude will benefit more from implicit training (for comprehensive reviews, see DeKeyser, 2012; Fu & Li, 2021). Following this line of thought, it would be intriguing to examine the extent to which auditory processing tests can be used as a diagnostic tool to provide profile-matched instructional approaches.

As reviewed earlier, it has been shown that learners with high-level explicit auditory processing benefit from explicit, language-focused speech training such as high variability phonetic training (e.g., Lengeris & Hazan, 2010; Qin et al., 2021). Few studies have examined the relationship between auditory processing (or any measure of aptitude for that matter) and the effectiveness of *incidental*, *implicit*, and *meaning-oriented* L2 speech training, arguably because scholars have exclusively used intentional, explicit and language-focused training to date. Though limited in number, some scholars have proposed using communicative focus on form (Lee & Lyster, 2016), task-based pronunciation training (Mora & Levkina, 2017), and phonological recasts (Saito, 2021) in this regard.

In accordance with the notion of incidental and multimodal auditory categorization learning in the field of cognitive psychology (Lim & Holt, 2011), our team has developed and tested the pedagogical potential of a video game-based target shooting game that aims to support segmental acquisition among Japanese learners of English (Saito et al., in press-b). In this game, participants are told that the faster targets are shot, the more points can be earned. Unknown to the participants, each target is accompanied by unique English consonants and vowel sounds. As such, participants are incidentally guided to use speech cues (L2 vowels and consonants) and acquire a series of novel foreign sounds as a by-product of playing the game. The findings of Saito et al. showed that participants' overall gains were similar to those of comparable explicit training (e.g., HVPT; overt identification of target contrasts followed by trial-by-trial feedback), but that the degree of improvement widely varied among individual participants. Follow-up studies are called for which investigate whether the effectiveness of this type of training is related to explicit and implicit auditory processing ability.

There is also a possibility that learners' degree of auditory precision in general (relatively strong or relatively poor) could help determine the extent to which they might benefit from

phonetic training (using speech stimuli) and auditory training (using non-speech stimuli). Provision of phonetic training alone could be sufficient for L2 learners with strong auditory processing skills as they are more capable of encoding the acoustic dimensions of new sounds and are likely to show larger gains when they receive various types of intensive L2 speech training (Lengeris & Hazan, 2010 for high variability phonetic training; Shao et al., 2022 for shadowing training; Sun et al., 2021 for 5 months of study abroad).

Conversely, such an approach (phonetic training only) could be confusing and/or have adverse effects when conducted with L2 learners with poorer auditory processing. Poor auditory processing prevents learners from detecting the novel acoustic characteristics of L2 speech while minimizing interference from their L1, extracting reliable acoustic cues (while ignoring irrelevant cues), and attaining robust L2 speech perception (e.g., pitch contour for the acquisition of lexical tones; Perrachione, Del Tufo, & Gabrieli, 2011).

As a remedial strategy, I propose that those with relatively low auditory processing may benefit from auditory training prior to phonetic training. During auditory training, learners are exposed to acoustically simple and monotonous nonspeech sounds that are manipulated along a single acoustic parameter. This can guide learners to focus on enhancing their sensitivity to the most useful dimensions of L2 speech (e.g., F2 = 1200–1600 Hz for English [æ] and [ʌ]; Saito et al., in press-c).

### **Auditory Processing and Different Aspects of L2 Learning**

In a broader sense, L2 speech proficiency concerns one's ability to access multiple dimensions of linguistic knowledge while comprehending and speaking language on a global level. Intuitively, it is unsurprising that auditory processing can explain some variances in the phonological aspects of L2 speech learning because the role of auditory input processing is most directly linked to segmental and suprasegmental acquisition. The question has now become: To what degree does auditory processing matter not only for the acquisition of lower-order linguistic information (phonology) but also the acquisition of higher-order linguistic information (vocabulary and grammar)? Auditory precision plays an important role in word segmentation (Norris, & McQueen, 2008) and the identification of word and phrase boundaries (Cutler & Butterfield, 1992). Further, auditory precision facilitates the detection of suffixes, inflection, and articles (Joanisse & Seidenberg, 1998) and word order (Penner et al., 2001). Since auditory processing is involved in every stage of L2 speech learning, future research can further explore how this ability *differentially* promotes the development of phonology, vocabulary, and grammar in a complementary fashion (for some emerging evidence, Kachlicka et al., 2019; Saito, Macmillan et al., in press-d).

### **Conclusion**

In this paper, I have introduced the auditory precision paradigm from L1 acquisition as a way to look at the complex mechanisms underlying adult L2 speech learning (i.e., Auditory Precision Hypothesis-L2). First and foremost, everyone can learn new sounds and achieve comprehensible, intelligible, communicatively adequate, and functional L2 oral proficiency as long as they practice the target language on a daily basis with a good level of motivation and willingness to communicate (Derwing & Munro, 2013). Here, the Auditory Precision Hypothesis-L2 is in line with major L2 speech learning theories in that both consider the quantity, quality, and intensity of experience as the crucial determinant of L2 speech learning (e.g., Flege & Bohn, 2021 for Speech Learning Model).

However, much individual variation has still been found in terms of the levels of attainment among highly experienced, regular, motivated, and functional L2 learners—some are



able to reach a stage of proficiency where they are almost indistinguishable from native speakers of the target language. These differences in learning outcomes exist not only because of the amount of time spent practicing the target language, but also because some learners are more cognitively and perceptually adept at making the most of every opportunity for input.

Consequently, this could lead to larger and more robust gains in the long run (Doughty, 2019).

An “auditory precision view” of L2 speech learning predicts that individuals with a good ear (i.e., precise auditory processing) are able to make the most of every input opportunity. That is, more precise auditory processing helps learners better capture the acoustic dimensions of L2 speech input (McAlister et al., 2001), adjust to new cue weighting patterns (Jasmin et al., 2021), develop new speech categories (or revise existing speech categories; Flege & Bohn, 2021), and continue to refine these categories to a near-nativelike level in the long run (Abrahamsson, 2012). The Auditory Precision Hypothesis-L2 assumes that domain-general and pre-categorical sound processing abilities govern language learning throughout the lifespan, and play a key role in late L2 speech learning (Mueller et al., 2012).

Given that auditory processing is fundamental to parsing L2 aural input, any lower-order problems will likely slow down other L2 speech learning processes, even if learners have a relatively strong working memory and high attentional control, receive ample input, and/or are motivated to practice the target language (Perrachione et al., 2011). Going forward, both researchers and practitioners are encouraged to carry out more auditory processing research that can provide insight into the different types of speech training participants may benefit from (e.g., explicit auditory processing for explicit speech training). In addition, more research is called for which explores how tests of auditory processing can be used to diagnose learners with relatively low-level auditory precision. This latter group of L2 learners may greatly benefit from auditory processing training, especially prior to L2 speech training and immersion experience. This will in turn ensure that all L2 learners can reduce the challenge of learning a new language despite any disadvantages they may have at the level of auditory processing.

## References

- Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*, 481-509. <https://doi.org/10.1017/S027226310808073X>
- Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology*, *83*(2), 111-130. [https://doi.org/10.1016/s0022-0965\(02\)00124-8](https://doi.org/10.1016/s0022-0965(02)00124-8)
- Bavin, E. L., Grayden, D. B., Scott, K., & Stefanakis, T. (2010). Testing auditory processing skills and their associations with language in 4—5-year-olds. *Language and Speech*, *53*(1), 31-47. <https://doi.org/10.1177/0023830909349151>
- Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception. In O. Bohn, & M. Munro (Eds.), *Language experience in second language speech learning: In honour of James Emil Flege* (pp. 13–34). John Benjamins.
- Brazil, D. (1997). *The communicative value of intonation in English book*. Cambridge University Press.
- Carroll, J. B., & Sapon, S. M. (1959). Modern language aptitude test. San Antonio, TX: Psychological Corporation.
- Casini, L., Pech-Georgel, C., & Ziegler, J. C. (2018). It's about time: revisiting temporal processing deficits in dyslexia. *Developmental Science*, *21*(2), e12530. <https://doi.org/10.1111/desc.12530>
- Chandrasekaran, B., Kraus, N., & Wong, P. C. (2012). Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology*, *107*(5), 1325-1336. <https://doi.org/10.1152/jn.00923.2011>
- Coffey, E. B., Herholz, S. C., Chepesiuk, A. M., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature Communications*, *7*(1), 1-11. <https://doi.org/10.1038/ncomms11070>
- 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., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, *40*, 63-72. <https://doi.org/10.1016/j.lindif.2015.04.005>
- 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>
- De Jong, N. H., & Mora, J. C. (2019). Does having good articulatory skills lead to more fluent speech in first and second languages? *Studies in Second Language Acquisition*, *41*, 227-239. <https://doi.org/10.1017/S0272263117000389>
- Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*(2), 163-185. <https://doi.org/10.1111/lang.12000>
- Diaz, B., Mitterer, H., Broersma, M., Escera, C., & Sebastian-Galles, N. (2016). Variability in L2 phonemic learning originates from speech-specific capabilities: An MMN study on

- late bilinguals. *Bilingualism: Language and Cognition*, 19(5), 955-970.  
<https://doi.org/10.1017/S1366728915000450>
- Doughty, C. J. (2019). Cognitive language aptitude. *Language learning*, 69, 101-126.  
<https://doi.org/10.1111/lang.12322>
- Eckman, F. (2004). From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition*, 26, 513-549.  
<https://doi.org/10.1017/S027226310404001X>
- Elfenbein, J. L., Small, A. M., & Davis, J. M. (1993). Developmental patterns of duration discrimination. *Journal of Speech, Language, and Hearing Research*, 36(4), 842-849.  
<https://doi.org/10.1044/jshr.3604.842>
- Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26, 551-585.  
<https://doi.org/10.1017/S0272263104040021>
- Field, J. (2008). Revising segmentation hypotheses in first and second language listening. *System*, 36(1), 35-51. <https://doi.org/10.1016/j.system.2007.10.003>
- Flege, J., & Bohn, O-S. (2021). The revised Speech Learning Model. In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress*. Cambridge University Press.
- Gabay, Y., Dick, F. K., Zevin, J. D., & Holt, L. L. (2015). Incidental auditory category learning. *Journal of Experimental Psychology: Human Perception and Performance*, 41, 1124. <http://dx.doi.org/10.1037/xhp0000073>
- Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the “natural order of L2 morpheme acquisition” in English: A meta-analysis of multiple determinants. *Language learning*, 51(1), 1-50. <https://doi.org/10.1111/1467-9922.00147>
- Gooch, D., Hulme, C., Nash, H. M., & Snowling, M. J. (2014). Comorbidities in preschool children at family risk of dyslexia. *Journal of Child Psychology and Psychiatry*, 55(3), 237-246. <https://doi.org/10.1111/jcpp.12139>
- Goswami, U., Fosker, T., Huss, M., Mead, N., & Szűcs, D. (2011). Rise time and formant transition duration in the discrimination of speech sounds: the Ba–Wa distinction in developmental dyslexia. *Developmental Science*, 14(1), 34-43.  
<https://doi.org/10.1111/j.1467-7687.2010.00955.x>
- Granena, G. (2019). Cognitive aptitudes and L2 speaking proficiency: Links between LLAMA and Hi-LAB. *Studies in Second Language Acquisition*, 41(2), 313-336.  
<https://doi.org/10.1017/S0272263118000256>
- Hamrick, P., Lum, J. A., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences*, 115(7), 1487-1492. <https://doi.org/10.1073/pnas.1713975115>
- Henry, L. A., Messer, D. J., & Nash, G. (2012). Phonological and visuospatial short-term memory in children with specific language impairment. *Journal of Cognitive Education and Psychology*, 11(1), 45-56. <https://doi.org/10.1177/0265659016655378>
- Hornickel, J., & Kraus, N. (2013). Unstable representation of sound: a biological marker of dyslexia. *Journal of Neuroscience*, 33(8), 3500-3504.  
<https://doi.org/10.1523/JNEUROSCI.4205-12.2013>
- 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>
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, B47-B57. [https://doi.org/10.1016/s0010-0277\(02\)00198-1](https://doi.org/10.1016/s0010-0277(02)00198-1)
- Isaacs, T., Trofimovich, P., & Foote, J. A. (2018). Developing a user-oriented second language comprehensibility scale for English-medium universities. *Language Testing*, 35(2), 193-216. <https://doi.org/10.1177/02655322177034>
- Jasmin, K., Dick, F., Holt, L. L., & Tierney, A. (2020). Tailored perception: Individuals' speech and music perception strategies fit their perceptual abilities. *Journal of Experimental Psychology: General*, 149(5), 914. <https://doi.org/10.1037/xge0000688>
- Jasmin, K., Sun, H., & Tierney, A. T. (2021). Effects of language experience on domain-general perceptual strategies. *Cognition*, 206, 104481.  
<https://doi.org/10.1016/j.cognition.2020.104481>
- 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>
- Kalashnikova, M., Goswami, U., & Burnham, D. (2019). Sensitivity to amplitude envelope rise time in infancy and vocabulary development at 3 years: A significant relationship. *Developmental science*, 22(6), e12836. <https://doi.org/10.1111/desc.12836>
- Koester, D., Gunter, T. C., Wagner, S., & Friederici, A. D. (2004). Morphosyntax, prosody, and linking elements: The auditory processing of German nominal compounds. *Journal of Cognitive Neuroscience*, 16, 1647–1668. <https://doi.org/10.1162/0898929042568541>
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5, 831–843. <https://doi.org/10.1038/nrn1533>
- 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(3), 530-566. <https://doi.org/10.1111/lang.12011>
- Lee, A. H., & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, 38, 35-64.  
<http://dx.doi.org/10.1017/S0272263115000194>
- Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, 128(6), 3757-3768.  
<http://dx.doi.org/10.1121/1.3506351>
- Levis, J. M. (2006). Pronunciation and the assessment of spoken language. In Hughes, R. (Ed.), *Spoken English, TESOL and applied linguistics: Challenges for theory and practice* (pp. 245–270). Palgrave Macmillan.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B), 467–477. <https://doi.org/10.1121/1.1912375>
- Lim, S. J., & Holt, L. L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science*, 35, 1390-1405.  
<https://doi.org/10.1111/j.1551-6709.2011.01192.x>

- McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of phonetics*, 30(2), 229-258. <https://doi.org/10.1006/jpho.2002.0174>
- McArthur, G. M., & Bishop, D. V. (2005). Speech and non-speech processing in people with specific language impairment: A behavioural and electrophysiological study. *Brain and Language*, 94(3), 260-273. <https://doi.org/10.1016/j.bandl.2005.01.002>
- Meara, P. (2005). LLAMA language aptitude tests: The manual. *Swansea: Lognostics*.
- Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal Processing Deficits of Language-Learning Impaired Children Ameliorated by Training. *Science*, 271(5245), 77–81. <https://doi.org/10.1126/science.271.5245.77>
- Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, 219(1), 36–47. <https://doi.org/10.1016/j.heares.2006.05.004>
- Moore, D. R. (2006). Auditory processing disorder (APD): Definition, diagnosis, neural basis, and intervention. *Audiological Medicine*, 4(1), 4-11. <https://doi.org/10.1080/16513860600568573>
- Mora, J. C., & Levkina, M. (2017). Task-based pronunciation teaching and research: Key issues and future directions. *Studies in Second Language Acquisition*, 39, 381-399. <https://doi.org/10.1017/S0272263117000183>
- Mora-Plaza, I., Saito, K., Suzukida, Y., Dewaele, J.-M., & Tierney, A. (2022). Tools for second language speech research and teaching. <http://sla-speech-tools.com/>
- Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, 35(4), 463–482. <https://doi.org/10.1093/applin/amu024>
- Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological review*, 115(2), 357. <https://doi.org/10.1037/0033-295X.115.2.357>
- Penner, Z., Weissenborn, J., & Wymann, K. (2001). On the prosody/ lexicon interface in learning word order. A study of normally developing and language impaired children. In J. Weissenborn, & B. Höhle (Eds.), *Approaches to bootstrapping: Phonological, lexical, syntactic, and neurophysiological aspects of early language acquisition* (pp. 269–294). John Benjamins.
- Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, 130(1), 461-472. <https://doi.org/10.1121/1.3593366>
- Protopapas, A. (2014). From temporal processing to developmental language disorders: Mind the gap. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), 20130090. <https://doi.org/10.1098/rstb.2013.0090>
- Qin, Z., Zhang, C., & Wang, W. S. Y. (2021). The effect of Mandarin listeners' musical and pitch aptitude on perceptual learning of Cantonese level-tones. *The Journal of the Acoustical Society of America*, 149, 435-446. <https://doi.org/10.1121/10.0003330>
- Raviv, O., Ahissar, M., & Loewenstein, Y. (2012). How recent history affects perception: The normative approach and its heuristic approximation. *PLoS Comput Biol*, 8(10), e1002731. <https://doi.org/10.1371/journal.pcbi.1002731>

- Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything? *Journal of Phonetics*, 31(3-4), 509–527. [https://doi.org/10.1016/S0095-4470\(03\)00046-9](https://doi.org/10.1016/S0095-4470(03)00046-9)
- Ruan, Y., & Saito, K. (forthcoming). Communicative focus on phonetic form revisited: Less precise auditory processing limits instructed L2 speech learning.
- Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., & Kraus, N. (2008). Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clinical Neurophysiology*, 119(8), 1720-1731. <https://doi.org/10.1016/j.clinph.2008.01.108>
- Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition*, 35(1), 1-29. <https://doi.org/10.1017/S0272263112000666>
- Saito, K. (2015). The role of age of acquisition in late second language oral proficiency attainment. *Studies in Second Language Acquisition*, 37, 713-743. <https://doi.org/10.1017/S0272263115000248>
- Saito, K. (2017). Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning*, 67, 665-693. <https://doi.org/10.1111/lang.12244>
- Saito, K. (2021). Corrective feedback and the development of L2 pronunciation. In H. Nassaji & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in language learning and teaching* (pp. 407-429). Cambridge University Press. <https://doi.org/10.1017/9781108589789.020>
- 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(3), 652-708. <https://doi.org/10.1111/lang.12345>
- Saito, K., Sun, H., & Tierney, A. (2019). Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental and suprasegmental sensitivity and performance via behavioural and neurophysiological measures. *Bilingualism: Language and Cognition*, 22(5), 1123-1140. <https://doi.org/10.1017/S1366728918000895>
- Saito, K., Sun, H., & Tierney, A. (2020). Domain-general auditory processing determines success in second language pronunciation learning in adulthood: A longitudinal study. *Applied Psycholinguistics*, 41(5), 1083-1112. <https://doi.org/10.1017/S0142716420000491>
- Saito, K., Sun, H., Kachlicka, M., Robert, J., Nakata, N., & Tierney, A. (2022). Domain-general auditory processing explains multiple dimensions of L2 acquisition in adulthood. *Studies in Second Language Acquisition*, 44(1), 57-86. <https://doi.org/10.1017/S0272263120000467>
- 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., Suzukida, Y., Tran, M., & Tierney, A. (2021). Domain-general auditory processing partially explains second language speech learning in classroom settings: A review and generalization study. *Language Learning*, 71(3), 669-715.  
<https://doi.org/10.1111/lang.12447>
- Saito, K., Cui, H., Suzukida, Y., Dardon, D., Suzuki, Y., Jeong, H., Revesz, A., Sugiura, M., & Tierney, A. (in press-a). Does domain-general auditory processing uniquely explain the outcomes of second language speech acquisition, even once cognitive and demographic variables are accounted for? *Bilingualism: Language and Cognition*. DOI:  
<https://doi.org/10.1017/S1366728922000153>
- Saito, K., Hanzawa, K., Petrova, K., Suzukida, Y., Kachlicka, M., & Tierney, A. (in press-b). Incidental and multimodal high variability phonetic training: Potential, limits, and future directions. *Language Learning*. DOI: <https://doi.org/10.1111/lang.12503>
- Saito, K., Petrova, K., Suzukida, Y., Kachlicka, M., & Tierney, A. (in press-c). Training auditory processing promotes second language speech acquisition. *Journal of Experimental Psychology: Human Perception and Performance*. DOI:  
<https://doi.org/10.1037/xhp0001042.supp>
- Saito, K., Macmillan, K., Kroeger, S., Magne, V., Takizawa, K., Kachlicka, M., & Tierney, A. (in press-d). Roles of domain-general auditory processing in spoken second-language vocabulary attainment in adulthood. *Applied Psycholinguistics*. DOI:  
<https://doi.org/10.1017/S0142716422000029>
- Silbert, N. H., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., & Tare, M. (2015). Non-native phonemic discrimination, phonological short-term memory, and word learning. *Journal of Phonetics*, 50, 99-119. <https://doi.org/10.1016/j.wocn.2015.03.001>
- Skehan, P. (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). John Benjamins.  
<https://doi.org/10.1075/bpa.3.02ske>
- Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2015). Stability and plasticity of auditory brainstem function across the lifespan. *Cerebral Cortex*, 25(6), 1415–1426.  
<https://doi.org/10.1093/cercor/bht311>
- Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, 17, 675-681.  
<https://doi.org/10.1111/j.1467-9280.2006.0176>
- Snowling, M. J., Gooch, D., McArthur, G., & Hulme, C. (2018). Language skills, but not frequency discrimination, predict reading skills in children at risk of dyslexia. *Psychological Science*, 29(8), 1270-1282.  
<https://doi.org/10.1177/0956797618763090>
- Sun, H., Saito, K., & Tierney, A. (2021). A longitudinal investigation of explicit and implicit auditory processing in L2 segmental and suprasegmental acquisition. *Studies in Second Language Acquisition*, 43(3), 551-573. <https://doi.org/10.1017/S0272263120000649>
- Thompson, E. C., White-Schwoch, T., Tierney, A., & Kraus, N. (2015). Beat synchronization across the lifespan: Intersection of development and musical experience. *PLoS One*, 10(6), e0128839. <https://doi.org/10.1371/journal.pone.0128839>
- Thompson, N. C., Cranford, J. L., & Hoyer, E. (1999). Brief-tone frequency discrimination by children. *Journal of Speech, Language, and Hearing Research*, 42(5), 1061-1068.  
<https://doi.org/10.1044/jslhr.4205.1061>

- Tierney, A., Gomez, J. C., Fedele, O., & Kirkham, N. Z. (2021). Reading ability in children relates to rhythm perception across modalities. *Journal of Experimental Child Psychology*, *210*, 105196. <https://doi.org/10.1016/j.jecp.2021.105196>
- Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in second language acquisition*, *28*, 1-30. <https://doi.org/10.1017/S0272263106060013>
- Trofimovich P, Kennedy S, and Foote JA (2015) Variables affecting L2 pronunciation development. In: Reed M and Levis J (eds) The handbook of English pronunciation. Malden, MA: Wiley, pp. 353-73. <https://doi.org/10.1002/9781118346952.ch20>
- Wen, Z. E., & Skehan, P. (2021). Stages of Acquisition and the P/E Model of Working Memory: Complementary or contrasting approaches to foreign language aptitude? *Annual Review of Applied Linguistics*, *41*, 6-24. <https://doi.org/10.1017/S0267190521000015>
- Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Applied Psycholinguistics*, *39*, 703-728. <https://doi.org/10.1017/S0142716418000152>
- Whiteford, K. L., & Oxenham, A. J. (2018). Learning for pitch and melody discrimination in congenital amusia. *Cortex*, *103*, 164–178. <https://doi.org/10.1016/j.cortex.2018.03.012>
- Zhang, Y.-X., Moore, D. R., Guiraud, J., Molloy, K., Yan, T.-T., & Amitay, S. (2016). Auditory Discrimination Learning: Role of Working Memory. *PLOS ONE*, *11*(1), e0147320. <https://doi.org/10.1371/journal.pone.0147320>
- Zheng, C., Saito, K., & Tierney, A. (2022). Successful second language pronunciation learning is linked to domain-general auditory processing rather than music aptitude. *Second Language Research*, *38*(3), 477-497. <https://doi.org/10.1177/0267658320978493>