Auditory Processing as Perceptual, Cognitive, and Motoric Abilities Underlying Successful Second Language Acquisition: Interaction Model

Kazuya Saito[1], Magdalena Kachlicka[1, 2], Yui Suzukida[1, 3], Ingrid Mora-Plaza[1, 4],
Yaoyao Ruan[1, 5], and Adam Tierney[2]

[1] Institute of Education, University College London
[2] School of Psychological Sciences, Birkbeck, University of London
[3] School of Medical Science, Juntendo University
[4] Department of Modern Languages and Literatures and English Studies, University of Barcelona
[5] Department of Education, University of Oxford

A growing amount of attention has been given to examining the domain-general auditory processing of individual acoustic dimensions as a key driving force for adult L2 acquisition. Whereas auditory processing has traditionally been conceptualized as a bottom-up and encapsulated phenomenon, the interaction model (Kraus & Banai, 2007) proposes auditory processing as a set of perceptual, cognitive, and motoric abilities—the perception of acoustic details (acuity), the selection of relevant and irrelevant dimensions (attention), and the conversion of audio input into motor action (integration). To test this hypothesis, we examined the relationship between each component and the L2 outcomes of 102 adult Chinese speakers of English who varied in age, experience, and working memory background. According to the results of the statistical analyses, (a) the tests scores tapped into essentially distinct components of auditory processing (acuity, attention, and integration), and (b) these components played an equal role in explaining various aspects of L2 learning (phonology, morphosyntax) with large effects, even after biographical background and working memory were controlled for.

Public Significance Statement
Everyone has a unique skill set when it comes to processing general characteristics of sounds, such as pitch and duration. An emerging theory proposes that individual differences in how well we can perceive sound can impact how quickly, effectively, and easily we learn our first and second languages. To date, auditory processing has traditionally been considered as a bottom-up, automatic, and isolated phenomenon. Following the interaction view, however, we argue that auditory processing can be considered a multidimensional phenomenon that encompasses not only the perception of acoustic characteristics (perceptual proficiency), but also the direction of attention towards specific acoustical elements (cognitive proficiency) and the transformation of auditory information into motor output (motoric proficiency). Indeed, recent studies have introduced new concepts of auditory processing that extend beyond mere acuity to include aspects such as attention and integration. In our study involving 102 Chinese English learners, we empirically examined the extent to which a model incorporating not just perceptual, but also cognitive and motoric facets of auditory processing could account for additional variance in language learning outcomes. Our findings align with an expanding body of evidence suggesting that auditory processing, which includes perceptual, cognitive, and motoric components, plays a pivotal role in language learning throughout the lifespan.

Correspondence concerning this article should be addressed to Kazuya Saito, Institute of Education, University College London, 20 Bedford Way, London WC1H0AL, United Kingdom. Email: k.saito@ucl.ac.uk

The outcomes of second language (L2) acquisition in adulthood are subject to a great deal of individual variation, with some learners achieving highly advanced, near-native proficiency and others experiencing considerable learning difficulties. Over the past 50 years, the mechanisms underlying complex L2 learning have been extensively researched in relation to the quantity, quality, and timing of immersion experience (e.g., Abrahamsson & Hyltenstam, 2009 for age of arrival; Jia & Aaronson, 2003 for types of interlocutors; Trofimovich & Baker, 2006 for length of residence [LOR]; for a comprehensive overview; see Flege & Bohn, 2021). At the same time, there is ample research evidence showing that even highly motivated and regular L2 users differ substantially in terms of learning speed and ultimate attainment, suggesting that adult L2 learning may not only be driven by experience-related factors but also tied to perceptual and cognitive factors within learners, that is, aptitude (Doughty, 2019). A range of frameworks have been established that feature a set of cognitive abilities specific to foreign language learning (e.g., Modern Language Aptitude Test for phonemic coding, grammatical sensitivity, inductive language learning, and rote memorization; Carroll & Sapon, 1959).

Recently, a growing number of scholars have examined domain-general auditory skills as a crucial component of aptitude, broadly defined as one's perceptual abilities to encode individual acoustic dimensions of sounds at a precategorical level (e.g., pitch, formants, duration, and amplitude). The auditory precision account of first language (L1) learning (Goswami, 2015) proposes that auditory processing serves as one of the initial abilities that learners use to parse auditory input for linguistic processing, and that individual differences in this ability impact all aspects of L1 learning and impairment. It has been shown that learners' auditory processing can be a key determinant of L2 learning outcomes (e.g., Mueller et al., 2012). To date, auditory processing has traditionally been assessed via a psychoacoustic discrimination task in which researchers examine how small a difference participants can perceive in the spectral and temporal content of sounds (i.e., perceptual acuity). Although we have obtained mounting evidence in support of the link between auditory processing and L2 acquisition, the contribution of auditory processing abilities beyond acuity remains unclear (cf., Snowling et al., 2018).

According to the interaction view (Kraus & Banai, 2007), auditory processing can be conceptualized as a multifaceted phenomenon wherein perceptual, cognitive, and motoric abilities are reciprocally interwoven at multiple levels. Specifically, auditory processing covers not only the perception of acoustic details (perceptual acuity) but also the direction of attention to particular acoustic dimensions (attentional control) and the conversion of audio information into motor action (audio-motor integration). All the auditory skills (i.e., acuity, attentional control, and audio-motor integration) are required for perception of sound in real-world tasks, such as learning a musical passage, identifying an environmental sound, or learning to produce a certain speech sound. However, little is known about the extent to which we can separately assess the different components of auditory processing and how they uniquely contribute to the outcomes of language learning.

In the current investigation, we propose a set of behavioral outcome measures to tap into the perceptual, cognitive, and motoric components of auditory processing, which we predict to be largely independent of each other, and investigate associations between each of these and L2 learning outcomes among $N = 102$ adult Chinese speakers of English with varied experience, proficiency, and working memory capacities. The process and product of L2 learning can be characterized by various dimensions (phonology, vocabulary, grammar, discourse), modes (perception, production), and contexts (naturalistic, classroom). For the sake of comparability to similar existing work (e.g., Kachlicka et al., 2019; Saito, Macmillan, et al., 2022), this study primarily examines the roles of auditory processing in two extensively studied areas of naturalistic L2 learning—specifically, phonological perception (measured through a vowel and prosody identification task) and morphosyntactic comprehension (measured by a grammaticality judgement task).

Domain-General Auditory Processing and Language Learning

Individuals differ widely in terms of the way they encode spectral and temporal information (Surprenant & Watson, 2001). In the field of cognitive psychology, there is an influential paradigm which states that although acoustic signals are used differently in different domains (e.g., language, music, emotion, and environmental sounds), auditory perception serves as an anchor across these domains. Similar perceptual processes are activated, for example, when we listen to someone speaking and someone playing an instrument (Kraus & Banai, 2007). In the context of language learning, toddlers orchestrate the acoustic information available in aural input to detect the statistical distribution of phonetic categories (Werker, 2018) while using prosodic cues to identify word and phrase boundaries (Cutler & Butterfield, 1992), encode syntactic structures (Jusczyk et al., 1992), and fill in morphological details (Joanisse & Seidenberg, 1998).

According to the auditory precision account of child language learning, individual differences in auditory processing are associated with lower-order phonetic processing outcomes (e.g., speech-in-noise perception) as well as higher-order linguistic skills (e.g., vocabulary and morphosyntax; Anvari et al., 2002; Bavin et al., 2010; Boets et al., 2008; Talcott et al., 2000; Tierney et al., 2021). Kalashnikova et al. (2019) provided longitudinal evidence regarding how auditory processing influences L1 vocabulary development over the first 3 years of life. Toddlers with difficulties discriminating certain dimensions of sounds are more likely to demonstrate slower phonological, lexical, and morphosyntactic processing later in development, which can in turn lead to more global language problems such as dyslexia (Casini et al., 2018 for duration; Goswami et al., 2011 for amplitude rise time; McArthur & Bishop, 2005 for fundamental frequency; but see Rosen, 2003; Schulte-Körne & Bruder, 2010 for their counterarguments regarding the causal relationship between auditory processing and L1 impairment).

Scholars have also begun to examine the role of auditory processing in adult L2 learning (Mueller et al., 2012). It has been argued that

individual differences in auditory processing can be even more consequential in L2 acquisition than in L1 acquisition (Saito & Tierney, 2022). In L1 acquisition, learners have ample input opportunities during the development of their phonetic categories (e.g., by the time they reach puberty, at approximately 14- to 15-years-old; Walley & Flege, 1999). Thanks to intensive exposure to language on a daily basis, auditory deficits can be mitigated (Rosen, 2003) and/or learners can learn how to use different perceptual strategies to achieve successful speech perception (e.g., Jasmin et al., 2020 for the case of people with difficulties processing pitch [amusics] using duration rather than pitch cues for the perception of structural features in speech and music). In contrast, input in L2 classrooms is generally insufficient, likely foreign-accented, and typically limited to several hours of language-focused instruction per week (Muñoz, 2014). Even when learners choose to study abroad or live in a country where they are immersed in their L2, opportunities for interaction with various interlocutors may still be limited and are susceptible to their willingness to communicate (Derwing & Munro, 2013).

Furthermore, unlike L1 acquisition, postpubertal L2 acquisition takes place in a linguistic space where the L1 system is fully developed. The acoustic analyses of incoming input are most likely affected by learners' already-automatized acoustic representations for L1 speech perception (McAllister et al., 2002). To acquire new speech sounds, L2 learners must adjust their existing cue weighting patterns (e.g., Chinese speakers need to reduce their reliance on pitch information and use both pitch and duration information to perceive English prosody; Jasmin et al., 2021). In some cases, L2 learners must establish new acoustic representations to process relatively new cues that are not actively exploited in their L1 system (e.g., Japanese speakers need to use third formant (F3) variation to perceive English [r] and [l]; Iverson et al., 2003). If L2 learners possess an imprecise auditory processing ability, they may struggle to analyze the acoustic properties of L2 input and how they diverge from those of L1 input. Consequently, these learners may persist in utilizing L1 perception strategies and fail to acquire more native-like L2 perception strategies, even after prolonged exposure to L2 input (Perrachione et al., 2011; Ruan & Saito, 2023).

To date, cross-sectional investigations have shown that auditory processing relates to various aspects of L2 speech proficiency (e.g., Kachlicka et al., 2019 for phonology; Saito, Macmillan, et al., 2022 for lexicogrammar). Longitudinal evidence has indicated that those with more precise auditory processing demonstrate gains when they immerse themselves in an L2 speaking environment (e.g., Saito et al., 2020 for L2 speech production; Sun et al., 2021 for L2 speech perception) and when they receive intensive phonetic training (e.g., Lengeris & Hazan, 2010 for English vowels; Qin et al., 2021 for Cantonese lexical tones; but see Brekelmans et al., 2022).

Auditory Processing as Perceptual, Cognitive, and Motoric Abilities

To date, most existing literature on auditory skills and language learning solely concerns perceptual acuity, that is, the ability to encode the acoustic details of sounds. To measure this precategorical, domain-general ability, scholars have commonly adopted a psychoacoustic task (Surprenant & Watson, 2001). Participants listen to and then discriminate a series of nonverbal, artificially synthesized sounds which are identical except for one particular target acoustic dimension. If they can hear smaller differences in these sounds,

they are considered to have more precise auditory processing. Given that behavioral tasks of this kind inevitably tap into a range of other cognitive abilities, some have argued that the unique contribution of auditory processing (operationalized as perceptual acuity) to language learning needs to be reexamined by measuring, comparing, and controlling for neighboring abilities (Snowling et al., 2018).

As a theoretical model of auditory skills, therefore, the interaction perspective (Kraus & Banai, 2007) has argued that sound perception can be conceptualized a set of perceptual, cognitive, and motoric phenomena that extends beyond acuity. Under this view, the way low levels of the auditory system encode sounds in a context-specific manner can be mediated in a top-down fashion. This departs from a traditional view that auditory processing is a bottom-up, automatic, and encapsulated phenomenon wherein the physical properties of sounds are encoded. Rather, the interaction view has stressed that auditory processing is a dynamic, integrated system as a result of "the intricate anatomical and functional connections between auditory and other brain areas and between cortical and subcortical areas within the auditory system" (p. 105). Building on this model, auditory processing consists of perceptual, cognitive, and motoric components. This composite ability concerns not only how precisely learners can perceive the details of a particular acoustic dimension (perceptual levels), but also the extent to which they can direct attention to one dimension while ignoring others (cognitive levels) and convert such information into motor action (motoric levels).

Attentional Control

Speech conveys acoustically complex information that can be perceived by attending to a range of different acoustic cues. With a view of accurate and prompt speech categorization, L1 listeners must not only perceive such acoustic cues but also identify which of these are most reliable and deserving of greater weight—for example, F3 information for English [r] and [l] (Espy-Wilson, 1993); spectral (rather than durational) cues for the English tense-lax vowel contrasts (e.g., English [i] and [ɪ]; Hillenbrand et al., 2000); pitch, duration, and amplitude for English stress (Lieberman, 1960). Certain scholars have claimed that individuals differ in their capacities to direct their attention to certain acoustic dimensions while ignoring others on a domain-general, precategorical level (Holt et al., 2018). Listeners rely on different dimensional weighting strategies to achieve the same speech percepts (e.g., Idemaru et al., 2012; Kong & Edwards, 2016). These cue weighting patterns can be shaped by domain-specific experience but are generalizable across different domains. For instance, Mandarin Chinese speakers tend to make greater use of pitch information and less use of duration information not only during their identification of phrase boundaries in English but also during music beat perception (Jasmin et al., 2021).

With respect to L2 speech learning, it has been argued that acquisition is made more difficult because learners process L2 input using their L1 segmentation patterns—for example, Japanese listeners overrely on second formant [F2] information for English [r] and [l] (Iverson et al., 2003); Chinese and Spanish listeners use duration cues for English vowel sounds (Escudero & Boersma, 2004; Flege et al., 1997; Liu et al., 2014); and Chinese and Vietnamese listeners depend more on pitch information and less on other information during their categorization of English stress (Nguyễn et al., 2008;

Yu & Andruski, 2010; Y. Zhang & Francis, 2010). Here, we hypothesize that the ability to direct attention to acoustic dimensions could be a key factor for successful L2 learning. Those with more precise dimension-selective attention can adjust their cue weighting patterns with more flexibility. As a result, they may develop new phonetic, lexical, and morphosyntactic categories more effectively and efficiently, and demonstrate more advanced L2 proficiency in the long run (Kim et al., 2018).

Intervention studies have attempted to guide L2 learners to attend to relevant cues via acoustic manipulation. Such training has been found to greatly help reset cue weightings and lead to more gains than mere exposure to L2 input—for example, in the acquisition of English [i] and [ɪ] among Spanish learners (Kondaurova & Francis, 2010) and Chinese learners (X. Zhang et al., 2021), and in the acquisition of English [r] and [l] among Japanese learners (Iverson et al., 2005). The extant literature suggests that cognitive individual differences in attentional control may be associated with L2 phonetic learning (Darcy et al., 2015) and L2 morphosyntax learning (Ellis, 2006). However, there is mixed evidence for the predictive power of selective attention when L2 learners engage in phonetic training (Mora-Plaza, Saito, et al., 2022; Mora-Plaza, Ortega, & Mora, 2022 vs. Ghaffarvand Mokari & Werner, 2018).

Notably, most existing studies have assessed selective attention to domain-specific information in particular modalities (e.g., recognition of words and sentences produced by two different talkers at the same time; for a methodological discussion, Humes et al., 2006). However, such domain-specific linguistic measurements of selective attention (such as speech-in-speech perception) are problematic in L2 populations because performance can be driven in part by the ability to perceive the input in the first place, regardless of selective attention ability. Additionally, while previous research has focused on the general capacity to direct attention to one of two different signals (attending to one speaker and disregarding another), some scholars have refined selective attention as the more nuanced capability to direct attention to various components of a single signal (Holt et al., 2018). The literature is scant in regard to how individuals differ in their attention to domain-general acoustic parameters within a single signal (e.g., pitch, formants, duration) and how such individual variation (i.e., dimension-selective attention) relates to various aspects of L2 acquisition (phonology and morphosyntax). The current study was designed to address these concerns.

Audio-Motor Integration

Those who are able to precisely perceive acoustic details may still have difficulty proceduralizing such information while learning to perceive and produce speech. The motor elements of speech processing encompass various autonomous processes such as articulatory planning, phonatory control, and neuromuscular execution (Guenther, 2006). Several influential models argue that motor systems underpin both speech perception and production in a complementary fashion. The motor theory (Liberman & Mattingly, 1985) posits that the process of speech perception is primarily driven by encoding the gestures made in the vocal tract during production, rather than by decoding the acoustic properties of speech. That is, the perception of speech can be facilitated by listeners' simulation of the necessary motor actions required for its production. The dual-stream model (Hickok et al., 2011) proposes two key neural

pathways interact to shape speech processing. The dorsal stream is primarily responsible for the sensorimotor mapping of sounds to articulatory representations. As such, listeners can form a plan for the motor actions necessary to produce the perceived speech. In contrast, the ventral stream focuses on speech comprehension by mapping speech sounds to their meanings, aiding the lexical and syntactic aspects of language.

There is substantial empirical evidence highlighting the role of audio-motor integration in speech perception and production. Neuroimaging studies have illustrated that areas typically associated with motor planning, such as the supplementary motor area or the premotor cortex, can be activated when participants engage in speech perception without any associated movement. For example, activation was observed in the region associated with tongue control and lip movement when participants listened to the speech sound /t/ (Pulvermüller et al., 2006). Similarly, the perception of nonverbal rhythms activates cortical and subcortical motor areas, such as the basal ganglia and supplementary motor area (Grahn, 2012; Grahn & Brett, 2007). Thus, another auditory processing ability that neighbors acuity, auditory-motor integration, involves tracking a single acoustic dimension across a sound stream, remembering the pattern, and converting it into immediate motor action.

L1 listeners efficiently encode the articulatory representations of sounds (i.e., feedforward mechanisms) and adjust their speech articulation while monitoring their own speech (i.e., feedback mechanisms), although there is much variation among individuals (e.g., those with dysarthria; Simmonds et al., 2011). Unsurprisingly, these mechanisms significantly slow down among L2 listeners, arguably because they often access the articulatory characteristics of new incoming inputs through existing L1 system representations (for a comprehensive view of the articulatory account of L1 and L2 speech learning, see Best & Tyler, 2007). Alternatively, it could be due to L2 listeners' probable difficulty in monitoring and adjusting their own speech, resulting from a misalignment between their own speech and that of native counterparts (Trofimovich et al., 2016). There is neuroimaging evidence of striatal plasticity showing that after brief training, adult listeners can perceive and produce novel speech sounds using their L1 motor sequences and developing similar sequences necessary for the novel sounds (Simmonds et al., 2014). Given the vital role of audio-motor integration in language learning, it is reasonable to assume that individual differences in this skill could explain the various degrees of success in L2 learning (Saito, Suzukida, et al., 2021).

While the motor processes involved in speech can be complex and multitiered (Guenther, 2006), we focus on one specific aspect of audio-motor integration in our current study. It is defined as the capacity to establish consistent alignment between the sequence order of one's movements and the sequence order of the sound stimuli (Tierney & Kraus, 2014) and measured via one's ability to reproduce target sound sequences (Flaugnacco et al., 2014).

For example, participants use a piano-like keyboard with five notes to play back melodic sequences which differ in fundamental frequencies (F0; e.g., Saito, Suzukida, et al., 2021 for a melody reproduction task) and/or use a drum to repeat back rhythm sequences (Tierney et al., 2017). These tasks are assumed to reveal not only how well participants can perceive but also how well they can replicate and predict a sequence of sounds. Tracking an entire sound stream (rather than one sound) across time requires spectral and temporal processing on a slow time scale (i.e., more than 1 s;

see the details of the reproduction task in the Method section). Such slow auditory processing can stimulate complex interactions between auditory regions and motor planning regions in the brain (Patel & Iversen, 2014).

The ability to remember, replicate, and predict the broad patterns of sound sequences is instrumental to detecting the prosodic patterns of speech. This skill enables listeners to segment auditory input into word units, conduct detailed phonological analyses, and achieve more reliable word recognition (Jusczyk & Aslin, 1995). The use of prosodic cues has been found to play a significant role in helping listeners pay attention to the morphosyntactic properties of language. In the context of L2 acquisition, the difficulty of learning is not only associated with the conceptual complexity of grammatical structures but also with their perceptual saliency, including factors such as voicing, pitch contour, sonority, and rhythm (Goldschneider & DeKeyser, 2001).

In our current study, to investigate the roles of audio-motor integration in L2 learning, we employed melody and rhythm reproduction tasks (for details, see the Method section). Both tasks were designed to tap into participants' abilities to recognize and reproduce sound sequences. While there are obvious differences between the precise movements required by the melody and rhythm reproduction tasks (replicating melodies and rhythms) and those involved in speech production (e.g., coordinating tongue configurations and movements), these tests could potentially assess broader mechanisms in audio-motor integration—that is, planning sequential movements intended for producing sound sequences.

In the L1 acquisition literature, some evidence has shown that audio-motor integration has been measured via domain-general tasks (e.g., rhythm reproduction) and linked to phonological skills and reading abilities (Flaugnacco et al., 2014) as well as grammar competence (Gordon et al., 2015). Though studies are limited in number, there is emerging evidence that audio-motor integration may play a key role in determining the degree of L2 success when learners study a target language via various types of focused training (e.g., Brekelmans et al., 2022 for perception training; M. Li & DeKeyser, 2017 for production training; Saito, Suzukida, et al., 2021 for foreign language experience). We have yet to know about the extent to which audio-motor integration can explain the outcomes of L2 learning outcomes in naturalistic and immersive settings.

Motivation for Current Study

A growing amount of evidence has shown that perceptual acuity could serve as a bottleneck for both L1 and L2 learning, but the link between auditory processing and language learning involves the influence of other neighboring abilities (Snowling et al., 2018). In accordance with the interaction view of auditory processing as a multifaceted phenomenon (Kraus & Banai, 2007), we test a three-factor model of auditory processing characterized by three distinct, perceptual, cognitive, and motoric abilities—the perception of acoustic details (acuity), the direction of attention towards relevant and away from irrelevant dimensions (attention), and the conversion of audio input into motor action (integration). By designing a set of behavioral tasks to tap into these abilities, we are the first to attempt to use the perceptual, cognitive, and motoric model of auditory processing and investigate how each component can help explain L2 learning outcomes.

While L2 learning is a complex phenomenon encompassing various dimensions (phonology, grammar), modes (perception, production), and contexts (naturalistic, classroom), existing literature is exclusively concerned with the roles of auditory processing in L2 phonology (the acquisition of segmentals and suprasegmentals). The finding of a medium-to-strong association ($R^2 = .15$ to $.35$; Kachlicka et al., 2019) between these is rather unsurprising given the deeply interwoven connection between auditory and phonological processing (acoustic information is used to develop phonetic categories). Some scholars have pointed out that auditory processing may also play a role in L2 morphosyntax learning. In the case of L2 English, auditory processing abilities may be especially helpful in the acquisition of relatively difficult morphosyntactic features, such as those with fewer phonemes, low syllabicity, and low sonority (Goldschneider & DeKeyser, 2001). Prosodic sensitivities are claimed to play a critical role in the acquisition of complex morphology (e.g., inflection), syntax (e.g., word order), and semantics (e.g., articles; Goad & White, 2019). Although some work has explored the relationship between auditory processing and L2 morphosyntax learning, resulting in small-to-medium effects ($R^2 = .05$ to $.15$; e.g., Saito, Macmillan, et al., 2022), the evidence is exclusively limited to the acuity dimension of auditory processing. It remains unclear whether, to what degree, and how the composite model of acuity, attention, and integration can increase the predictive power of auditory processing for both phonological and morphosyntactic aspects of L2 learning.

The current study aims to unravel whether, to what degree, and how the composite model of acuity, attention, and integration can increase the predictive power of auditory processing for both phonological and morphosyntactic aspects of L2 learning. The results of this study will be contrasted with those of a previous study, which highlighted the roles of perceptual acuity in L2 phonology and morphosyntax (e.g., $R^2 = .15$ to $.35$ in Kachlicka et al., 2019; $R^2 = .05$ to $.15$ in Saito, Macmillan, et al., 2022). In this investigation, a total of 102 Chinese learners of English with a wide range of L2 experience and proficiency levels were recruited in the United Kingdom. First, they took a battery of auditory processing tests which were assumed to tap into acuity, attention, and integration. Then, they completed both speech perception and grammaticality judgement tasks (to index their L2 phonological and morphosyntactic proficiency). Latent variables underlying test scores were examined via an exploratory factor analysis, and the relationship between these variables and other affecting factors—biographical background and working memory—was also investigated. To determine the unique contribution of auditory processing, we examined the link between auditory processing scores and L2 learning outcomes with the effect of biographical background and working memory statistically factored out. Two research questions were formulated as follows:

1. To what extent are the perceptual, cognitive, and motoric components of auditory processing associated with participants' biographical backgrounds and working memory?
2. To what extent do acuity, attention, and integration explain independent variance in both lower- and higher-order aspects of L2 learning (phonology and morphosyntax)?

As for R1 the relatively stable nature of auditory processing (Hornickel et al., 2012) suggests that it may be free of the influence

of experience-related factors (e.g., LOR and daily L2 use). Thus, it was predicted that the link between auditory processing and experience would be marginal. From a methodological perspective, scholars have discussed the extent to which the tasks of auditory processing could tap into one's individual differences in perceptual skills because any behavioral measures can inevitably involve other associated cognitive abilities such as executive function and working and long-term memory (Ahissar et al., 2006). Consequently, researchers have begun to incorporate both auditory processing and other cognitive tasks in their studies, aiming to scrutinize the connection between audition and language, while accounting for other cognitive factors (Snowling et al., 2018). In the current investigation, we compared participants' auditory processing scores (the discrimination task for acuity, the repetition task for attention, and the reproduction task for integration) with their performance in the forward and backward digit span tasks both of which are designed to measure individual differences in working memory. Our hypothesis posited that working memory plays a crucial role in the completion of each auditory processing task, thereby exhibiting substantial correlations with participants' acuity, attention, and integration scores.

This particular aspect of cognitive abilities (i.e., working memory) was chosen for the following reasons. First, working memory encapsulates a variety of cognitive functions necessary to accomplish the tasks created for our study, aimed at measuring perceptual, cognitive, and motor facets of auditory processing. Participants would be required to utilize their phonological memory to hold acoustic information from the stimuli, facilitating their ability to discern differences (discrimination task), focus on specific acoustic dimensions while ignoring others (repetition task), and duplicate heard melodic and rhythmic sequences (reproduction task). Incorporating working memory enables us to isolate the effects of auditory processing on L2 learning outcomes. Second, working memory is among the most extensively researched cognitive abilities related to L2 acquisition (the main focus of our study) across various contexts (e.g., Linck et al., 2013). Therefore, by contrasting the predictive capacities of auditory processing and working memory, we can assess the relative significance of auditory processing in L2 learning outcomes. Third, according to S. Li's (2016) L2 aptitude framework, the two working memory tasks utilized in this study (i.e., forward vs. backward digit span) are commonly used and believed to target overlapping yet fundamentally different types of memory abilities. The former represents phonological short-term memory (i.e., the ability to store information in the phonological loop), while the latter embodies executive function working memory (i.e., the capacity to manipulate stored information). We contend that the use of these two memory tasks covers a broad spectrum of cognitive abilities that may intersect with or diverge from those utilized in the auditory processing tasks. Finally, as the digit span tasks involve retaining numerical numbers (devoid of any audio information), the results may demonstrate the domain generality of participants' cognitive memory abilities. This, in turn, can assist in controlling for the auditory specificity of the outcomes derived from the auditory processing tasks.

As for R2, previous literature has shown that L2 learners' individual differences in perceptual acuity are moderately predictive of L2 phonology outcomes ($R^2 = .15$–.25; Kachlicka et al., 2019) and weakly associated with L2 morphosyntax outcomes ($R^2 = .05$ to .15; Saito, MacMillan, et al., 2022). Given that our test battery

incorporates two additional components of auditory processing (attention and integration), it was predicted (a) that the association between audition and acquisition would be stronger in the current study ($R^2 > .30$); (b) that such auditory processing effects would be more pronounced in phonology than morphosyntax; and (c) that the strength of these correlations would remain significant even after participants' biographical backgrounds and working memory were statistically factored out.

Method

Transparency and Openness

Following the Transparency and Openness Promotion guidelines set by the Center for Open Science (Nosek et al., 2017), we report how the participants were recruited and screened, how the sample size was determined and justified by the power analyses, and how the auditory processing and L2 measures were developed. For future replication efforts, all materials have been deposited on the L2 Speech Tools platform, a resource for L2 researchers and educators (Mora-Plaza et al., 2022; Mora-Plaza, Ortega, & Mora, 2022; https://sla-speech-tools.com). Additionally, the materials are shared as Open Materials on the GORILLA online psychology experiment platform (Anwyl-Irvine et al., 2020; https://app.gorilla.sc/openmaterials/497080). The raw data from this project is available for review on Open Science Framework (https://osf.io/s4vkz/?view_only=25b22feb0e4142d4a3191ae96617a2b4).

Participants

Given that some adult L2 learners have little access to their target language on a daily basis (resulting in little learning), efforts were made to recruit regular users of L2 English in immersive settings. In Winter 2021, an electronic flyer was disseminated across a range of online communities and social media platforms across the United Kingdom. The flyer explicitly explained the content of the project and specified the conditions necessary for participation: (a) Participants had to speak Mandarin Chinese as a native language from birth onward (to control for L1 effects); (b) they had to use L2 English as a main language of communication either at home or work (to avoid those who used only their L1); (c) they had to have arrived in the United Kingdom after the age of 16 (to limit the study to adult L2 learners); and (d) they must have experienced at least 0.5 years of immersion (allowing us to focus on mildly to highly experienced L2 users). This particular LOR range was chosen because the predictive power of auditory processing has been most clearly observed among this population (mid-to-long-term L2 residents; for a comprehensive review, Saito, Suzukida, et al., 2021).

Data Collection

Due to the global pandemic (as of Winter 2021), all data collection took place via GORILLA (Anwyl-Irvine et al., 2020). The experiment comprised three broad tasks in the following sequence: (a) auditory processing tasks (discrimination, attention, and integration); (b) L2 tests (phonology, and morphosyntax); and (c) a comprehensive biographical questionnaire (age and experience). The entire session lasted for approximately 45 min. To ensure that participants met the necessary conditions for the project and had the facilities to complete the entire experiment on their own without any

major interruptions (a computer meeting specification requirements and stable Internet access), a range of precautionary measures were taken. All 150+ participants who expressed interest in the study were individually contacted by a researcher (a native speaker of Mandarin), who ensured that they were L1 Mandarin users (based on the interview) and they had normal hearing (based on the questionnaire). The participants who passed the first screening were given a link to the Gorilla platform where they were asked to complete a quick sound check using their own headsets and two brief working memory tasks. Only after we confirmed that there were no technical problems (i.e., all the sound samples were perceptible with their headsets and all the working memory responses were correctly recorded), did the remaining participants ($n = 110$) proceed with the main data collection phase. Participants' working memory scores also served as another selection criterion. Individuals who failed to recall three numeral strings in the forward digit span task were deemed to lack sufficient concentration for this type of online experiment and were therefore excluded. In our data set, however, all participants displayed abilities exceeding the cutoff point (i.e., maintaining 4+ numbers). Each participant was given a Gorilla link along with a particular time slot (date and time) to start and complete the experiment. They were asked to complete the task in one sitting. Using Gorilla's progress tracking function (which displays the timing of task completion), their compliance with this instruction and performance was monitored remotely via Gorilla. Data from a total of 102 participants who completed all the tasks were used for final analyses.

Power Analyses

The current investigation analyzed the relationships between three facets of auditory processing (perceptual, cognitive, and motoric abilities) and L2 proficiency (phonology and morphosyntax), while controlling for biographic variables such as LOR, age of acquisition, daily L2 use, prior foreign language instruction, and music education. As outlined in the Results section, a series of simulation analyses were performed based on the prior studies (Kachlicka et al., 2019; Saito, Sun, et al., 2022) to determine and confirm the adequacy of the sample size of the current study ($n = 102$).

Auditory Processing Measures

The perceptual, cognitive, and motoric components of auditory processing (acuity, attention, and integration) were measured via discrimination, repetition detection, and reproduction tasks, respectively. As previously reported (Saito & Tierney, 2022), the test–retest reliability of these test formats is "fair" to "excellent" (interclass correlations = .4 to .8).

Perceptual Acuity

The ability to perceive the details of particular acoustic information (i.e., perceptual acuity) can be measured via a discrimination task (Surprenant & Watson, 2001). We prepared a set of synthesized stimuli with very simple acoustic characteristics (e.g., completely flat fundamental frequencies, formant contours, and harmonic spectrums) that listeners would not perceive as speech. For each of the three subtests (formants, pitch, and amplitude rise time), the stimuli were identical apart from the target acoustic dimension.

The task had three subtests—formant, pitch, and amplitude rise time discrimination. For the formant subtest, a total of 101 complex tones were created (one standard stimulus [Level 0] plus 100 comparison stimuli [Levels 1 to 100]). Each sample had a duration of 500 ms. Two linear 5-ms amplitude ramps were inserted at the beginning and end of each stimulus. The fundamental frequency was set to 100 Hz with harmonics up to 3,000 Hz. Three formants were inserted at 500, 1,500, and 2,500 Hz, using a parallel formant filter bank (Smith, 2007). The second formant (F2) of the standard stimulus was set to 1,500 Hz, and the comparison stimuli ranged from 1,502 to 1,700 Hz in 100 steps of approximately 2 Hz. For the pitch and amplitude rise time subtests, a total of 101 four-harmonic complex tones were prepared. A 5-ms linear ramp was inserted at the beginning and end. In the pitch subtest, the fundamental frequency was set to 330 Hz for the standard stimulus and ranged from 330.3 to 360 Hz for the comparison stimuli in increments of 0.3 Hz. In the amplitude rise time subtest, the length of the first amplitude ramp, which was set to 15 ms for the standard stimulus, changed from 10 to 300 ms.

In each trial, participants listened to three synthesized stimuli. While the second stimulus remained constant, only the first or last stimulus could vary. Participants were then required to identify the sound that differed from the other two by pressing either the number "1" or "3" on a computer screen.

Based on Levitt's (1971) adaptive threshold procedure, the size of the difference varied from trial to trial depending on participants' task performance. The tests started from the midpoint of the comparison stimuli (Level 50) and changed with a step size of 10. That is, when an incorrect response was made, the difficulty of the task decreased by increasing the difference between stimuli by 10 steps, and when three consecutive correct responses were provided, the task difficulty increased by decreasing the difference between stimuli by 10 steps. After the first reversal of direction, the step size changed from 10 to 5 and then from 5 to 1. The tests stopped either after 70 trials or eight reversals. Participants' auditory processing score was determined by the location of the final reversal. The scores indicated how small a difference (between the standard and comparison stimuli) participants could perceive. Consequently, lower scores suggested a higher precision in perceptual acuity among participants.

For audio stimuli and a demo task, see L2 Speech Tools (Mora-Plaza et al., 2022; https://sla-speech-tools.com/) and Gorilla Open Materials (Anwyl-Irvine et al., 2020; https://app.gorilla.sc/open materials/497080).

Dimension-Selective Attention

To date, selective attention has been primarily measured on the level of tracking different sound sources (e.g., attending to key words embedded in a sentence while ignoring other sentences produced at the same time; Humes et al., 2006). Employing an analogous design, we measured the ability to attend to a single target acoustic dimension (i.e., dimension-selective attention) while ignoring a distractor dimension within a single sound stream. This dimension-selective attention paradigm involved a repetition detection task where participants are asked to attend to changes in one of the acoustic dimensions and report repetitions within the attended stream (Holt et al., 2018; Symons et al., 2021).

Each trial consisted of presentation of a single sequence of 12 complex tones, consisting of a fundamental frequency and 50

harmonics. Each tone was 100 ms in duration and tones were separated by 400 ms of silence, for a total sequence length of 6 s. Each tone began with a linear amplitude ramp, and each tone featured a single constant formant created using a parallel formant filter bank (Smith, 2007). Tones could take on three combinations of fundamental frequency (pitch), formant, and amplitude rise time (the length of the amplitude ramp at the beginning of the stimulus). The three possible values were, for pitch, 100, 109, and 118 Hz; for formant, 2,040, 2,425, and 2,884 Hz, and for rise time, 5, 50, and 95 Hz.

For each trial, there was an attended dimension, an unattended dimension, and a dimension that did not vary. The dimension that did not vary was always set to the middle value of the continuum for that dimension (109, 2,425 Hz, and 50 ms for the pitch, formant, and amplitude rise time dimensions, respectively). The attended dimension and unattended dimensions varied between the highest and lowest values for the associated continua. The attended and unattended dimensions each varied at steady rates, either every two tones (1 Hz) or every three tones (0.667 Hz). However, on half of the trials, there was a "repetition": a point where the expected change in the dimension did not happen. Participants' task was to indicate, after each trial, whether they heard a repetition in the attended dimension, ignoring any potential repetitions in the unattended dimension. Responses were made by clicking buttons marked "yes" or "no" on the screen after each trial. Feedback (correct or incorrect) was given after each trial.

For a given attention condition, trials were equally split regarding assignment of the unattended and unvarying dimensions. For example, for half of the attention to pitch trials, formant varied, while amplitude rise time did not vary, while for the remaining half, amplitude rise time varied, but formant did not. Sixteen trials were presented for each attention condition, for a total of 48 trials. Performance was calculated as percent correct. Two versions of the task were created with identical stimuli but varying attention instructions so that stimuli across attention conditions were identical, varying only in the focus of attention.

Higher scores achieved by participants in the repetition task were indicative of their enhanced ability to focus on a specific sound dimension while disregarding others. Therefore, these higher scores demonstrated greater attentional capabilities.

For a visual summary of the repetition detection task, audio stimuli, and a demo task, see online supplemental material S1, L2 Speech Tools (Mora-Plaza et al., 2022; https://sla-speech-tools.com/), and Gorilla Open Materials (Anwyl-Irvine et al., 2020; https://app.gorilla.sc/openmaterials/497080).

Audio-Motor Integration

Following the procedure used in Tierney et al. (2017), the participants completed two different reproduction tasks, where they were asked to listen to a series of easily perceptible, nonverbal sounds which differed in pitch (melody) or nonpitch information (rhythm), remember the sequence, and reproduce it on a keyboard or by pressing buttons displayed on the screen with a mouse.

In the melody reproduction task, a total of 10 melodies were created, each of which consisted of a sequence of seven notes. The duration of each note was 300 ms, with a 50-ms cosine ramp at the beginning and end of the note to avoid the perception of transients. Each note was drawn from one of five six-harmonic complex tones.

The amplitude was held constant across harmonics, and the fundamental frequencies in these tones were set to 220, 246.9, 277.2, 311.1, and 329.6 Hz (representing the first five notes of the major scale, respectively). Each melody began on the third note of the scale (277.2 Hz) and was followed by either one note higher on the scale (311.1 Hz) or one note lower on the scale (246.9 Hz). When the melody reached the lowest (220 Hz) or highest note on the scale (329.6 Hz), the next note in the sequence was either the same pitch (220, 329.6 Hz) or one note higher or lower (246.9 or 311.1 Hz). This pseudorandomized process repeated up to the seventh note.

First, participants were introduced to five buttons, each representing a single musical note. To minimize the influence of prior relevant experience (e.g., piano training), these buttons were displayed in a vertical (rather than horizontal) manner. The relative height of each button corresponded to the fundamental frequency it represented (higher positions indicated higher fundamental frequencies). Participants were encouraged to try clicking the buttons to hear what type of tone each button produced. After they were familiarized with the tones, they proceeded to the main task. Unlike in working memory tests, where sound and letter sequences are displayed only once, participants heard each melody three times before being asked to repeat it, reducing the memory load required to complete the task. This task feature should also ease encoding.

Given that the main objective of the test was to examine how participants could adjust to a new motor task (i.e., hitting buttons in response to sound prompts), they were not given an opportunity to practice. The accuracy ratio of the first seven button presses was calculated as a percentage out of 100. Each note was scored as a 1 when it was identical to the target note and scored as a 0 when it differed from the target note. The analysis included seven first button presses only as melodies had seven notes. As such, accuracy was calculated for those first seven notes only in both rhythm and melody reproduction tasks. Participants' averaged performance across all 10 melodies served as their melodic Integration score.

In the rhythm reproduction task, a total of 10 rhythmic patterns were prepared based on Povel and Essens's (1985) notion of strongly and weakly metrical sequences. Each sequence consisted of 150-ms conga drum hits (retrieved from www.freesound.org). The first five sequences were strongly metrical, including more drum hits on the first and third beats than the remaining five weakly metrical sequences. The total duration of each rhythm pattern was 3.2 s. Participants listened to each rhythmic sequence three times and then reproduced the rhythm by pressing the space key as if they were beating a drum (i.e., pressing the button when there was a drum hit and pausing when there was silence). To calculate participants' accuracy rate, the interpress times were quantized at the first ten 200-ms interval points (200, 400, 600, 800, 1,000, 1,200, 1,400, 1,600, 1,800, and 2,000 ms). The presence of a drum hit or a rest was calculated at every interval point and compared to the sequences of hits and rests in the stimulus. The resulting ratio of correct hits and rests served as the rhythm integration score. Higher scores in the reproduction task were indicative of participants' higher proficiency in reproducing melodic and rhythmic sequences.

For a visual summary of the reproduction task, audio stimuli, and a demo task, see the online supplemental material S2, L2 Speech Tools (Mora-Plaza et al., 2022; sla-speech-tools.com/),

and Gorilla Open Materials (Anwyl-Irvine et al., 2020; https://app .gorilla.sc/openmaterials/497080).

Working Memory Measures

The behavioral assessment of auditory processing inevitably taps into a range of memory and executive functions (Snowling et al., 2018). To examine the degree of overlap between auditory processing scores (measured via discrimination, sequence detection, and reproduction tasks) and other cognitive abilities, participants' working memory was measured as a covariate variable. In Baddeley's (2000) influential framework, working memory comprises the phonological loop (how much information can be stored) and the central executive (how much information can be processed for cognitive operations). Following Olsthoorn et al. (2014), these two aspects of working memory were assessed via forward and backward digit span tasks, respectively. During both tasks, participants were asked to remember a series of digits, recall the digits in the order they were presented (for forward span) or in reverse order (for backward span), and submit the response in a provided space on the screen with their keyboard. Length of the digits increased from three to 11 digits with two trials at each length. Each digit was displayed on a computer screen for 500 ms. Participants' memory score for each task was determined by the highest number of digits in the series for which they provided correct responses at both trials (summarized in Table 1). For the task materials used in the current study, L2 Speech Tools (Mora-Plaza et al., 2022; Mora-Plaza, Ortega, & Mora, 2022; https://sla-speech-tools.com/) and Gorilla

Open Materials (Anwyl-Irvine et al., 2020; https://app.gorilla.sc/ openmaterials/497080).

Second-Language Measures

Phonology

In the current study, participants' L2 speech perception ability was measured via a two-alternative forced-choice identification test, which has been widely used in L2 speech literature (Sakai & Moorman, 2018) as an ecologically valid way to mirror the present state of L2 phonological proficiency (Flege, 1993). To cover a range of difficult segmental and suprasegmental features for L2 listeners, a total of 46 phonological contrasts were initially selected, comprising tense-lax vowel contrasts ($n = 13$; e.g., /i/ vs. /ɪ/, /u/ vs. /ʊ/, /æ/ vs. /ɛ/), consonant voicing contrasts ($n = 13$; e.g., /g/ vs. /k/, /d/ vs. /t/), and contrastive focus ($n = 20$; e.g., READ books vs. read BOOKS). As word frequency likely affects L2 learners' perception performance (Flege et al., 1996), Vocab Profiler (Cobb, 2012) was consulted to confirm that the target items were limited to frequent words, that is, the first 4,000 word families in British National Corpus Word Lists. All stimuli were produced by a male native speaker of Southern British English. In this test, participants listened to a word or sequence of words and chose the word or phrase which best matched what they heard from two options shown on the screen by pressing the keys "1" (left) or "2" (right). Results of our pilot studies conducted with a wide range of L2 speakers in London revealed that participants' perception of vowels and contrastive focus varied widely, while their consonant perception was largely at ceiling.

Table 1

Biographical Information, Auditory Processing, Working Memory, and L2 Learning Outcomes of Participants

				95% CI	
Biographical information	*M*	*SD*	Range	Bottom	Upper
A. Biographical backgrounds					
Age (years)	25.2	3.8	19–36	24.5	26.0
LOR (years)	3.0	2.2	0.5–10	2.6	3.5
Age of arrival (years)	21.6	4.1	16–36	20.8	22.4
Current L2 use at home (%)	29.2	28.0	0–100	23.7	34.70
Current L2 use at work (%)	78.3	21.2	10–100	74.1	82.5
Current L2 use at social (%)	33.9	27.7	0–100	28.4	39.3
Length of (EFL) training prior to arrival (years)	12.5	3.2	5–22	11.9	13.1
Music training	Yes ($n = 49$), no ($n = 53$)				
Gender	88 females, 14 males				
B. Auditory processing					
Formant discrimination (acuity)[a]	22.1	12.9	2.3–67.3	19.6	24.7
Pitch discrimination (acuity)[a]	8.8	6.3	2.3–41.0	7.5	10.1
Rise time discrimination (acuity)[a]	24.1	21.8	2.3–89.6	19.9	28.5
Formant detection (attention)[b]	69.7	18.0	18.7–100	66.1	73.2
Pitch detection (attention)[b]	67.8	15.8	25.0–100	64.6	70.9
Rise time detection (attention)[b]	67.8	19.2	18.7–100	64.0	71.6
Melody reproduction (integration)[b]	71.1	26.3	14.2–100	65.5	76.0
Rhythm reproduction (integration)[b]	73.1	11.6	50.7–100	70.6	75.2
C. Working memory					
Forward digit span	8.6	1.4	4–11	8.3	8.9
Backward digit span	5.1	1.8	3–11	4.7	5.5
D. L2 learning outcomes					
Phonology (%)	82.5	11.8	51.5–100	80.2	84.9
Morphosyntax (%)	68.2	11.6	44.3–94.9	65.9	70.5

Note. L2 = second language; CI = confidence interval; LOR = length of residence; EFL = English-as-a-foreign-language.
[a] Smaller acuity scores indicate more sensitivities to a specific acoustic dimension. [b] Larger attention and integration scores reflect superior processing abilities of specific acoustic dimensions.

Thus, the final test in the current study included only vowels and contrastive focus (a total of 36 stimuli). Participants first engaged in the identification of vowel contrasts[1] and then the identification of contrastive focus. Within each block, the stimuli were played in a randomized order. The accuracy ratio scores (0% to 100%) were automatically calculated based on the total number of correct responses.

Morphosyntax

To measure participants' L2 morphosyntax proficiency, a timed grammaticality judgement task was adopted (Plonsky et al., 2020). There remains some debate over whether and how much these test scores probe into L2 learners' implicit morphosyntax knowledge—knowledge acquired without explicit awareness. However, a consensus has emerged that timed grammaticality judgement scores reflect the degree of automatization in L2 grammar competence (Suzuki & DeKeyser, 2017). A total of 68 short sentences (five to 12 words) were adopted from Godfroid et al. (2015). They covered a total of 17 morphosyntactic structures in English (e.g., plurality, tense, articles). For each structure, there were four sentences, two of which were free of linguistic errors and two of which included an incorrect use of the target structure. Based on the results of the pilot work by Godfroid et al., a specific time limit (1,800–6,240 ms) was given for participants to read and decide whether the sentence was syntactically acceptable by clicking the response boxes provided on the computer screen. This time limit was determined by L1 English speakers' performance. The accuracy ratio scores (0%–100%) were automatically calculated based on the number of sentences that were correctly identified.

Results

Components of Auditory Processing

The first objective of the statistical analyses was to find any broad patterns underlying the outcomes of this auditory processing battery. In the current study, the acuity, attention, and integration components of auditory processing were assessed via a series of discrimination, repetition detection, and reproduction tasks. According to the results of normality tests (Kolmogorov–Smirnov test), participants' reproduction scores were comparable to normal distribution ($D = .066, .062, p = .777, .803$), but their discrimination and repetition detection scores significantly differed from normal distribution ($D = .113$ to $.286, p < .001$). Thus, the latter scores were transformed via a log 10 function. An exploratory factor analysis was conducted with Varimax rotation. Loewen and Gonulal's (2015) field-specific recommendations were used to determine the number of groupings. To include the largest amount of variance in the participants' auditory processing abilities, the Jolliffe criterion (.7; explaining 82.0%) rather than the Kaiser criterion (1.0; explaining 53.27%) was used. The factorability of the entire data set was adequate according to Bartlett's test of sphericity ($\chi^2 = 188.152, p < .001$). The Kaiser–Meyer–Olkin measure of sampling adequacy (.811) was considered to be "meritorious" according to the benchmark set by Kaiser and Rice (1974). A "five-factor" solution was suggested with an eigenvalue beyond .7, accounting for 57.539% of the variance in auditory processing measures.

In light of Hair et al.'s (1998) guidelines, .4 was set as the cutoff value to identify the "practically" significant factor loadings. All the factor loadings are summarized in Table 2. Factor 1 was labeled as "attention" as it captured all the recognition measures which were designed to tap into dimension-selective attention. Factor 2 was labeled as "integration" as it encompassed both reproduction tasks for audio-motor integration. Factors 3 to 5 were associated with each of the three discrimination tasks and were thus termed "pitch acuity," "formant acuity," and "temporal acuity," respectively. The results suggested that the auditory processing test battery represented three dimension-specific acuity abilities, that is, the ability to perceive three different acoustic dimensions (pitch, formants, and amplitude rise time acuity); and two broad abilities, that is, the ability to attend to a single dimension while ignoring others (selective attention) and the ability to convert audio information into motor action (audio-motor integration). To minimize multicollinearity problems, the resulting factor scores were used for the rest of the statistical analyses.

Auditory Processing, Experience, and Working Memory

The next analyses were performed to explore the extent to which the five different aspects of auditory processing (attention, integration, pitch acuity, formant acuity, and temporal acuity) related to biographical background (LOR, age of arrival, current L2 use, amount of English-as-a-foreign-language training, and amount of music training) and to working memory (forward and backward digit span). The results of nonparametric Spearman correlation analyses are summarized in the online supplemental material S3. In most instances, auditory processing was not significantly associated with experience-related variables ($p > .05$). However, a weak correlation was found between integration and music training ($r = .299, p = .002$). While working memory was not clearly related to the acuity aspects of auditory processing ($p > .05$), a small-to-medium amount of overlap was found between working memory and the attention and integration aspects of auditory processing ($r = .320–.349, p \leq .001$).

Auditory Processing and L2 Learning Outcomes

The final objective of the statistical analyses was to examine the extent to which participants' auditory processing scores (acuity, attention, and integration) uniquely explained their L2 proficiency (phonology and morphosyntax). For the comparability of the current investigations to the prior study (Kachlicka et al., 2019; Saito, Sun, et al., 2022), the same statistical procedure, that is, linear mixed effects regression analyses, was adopted to examine the relationship between two different types of L2 learning outcomes (phonology and morphosyntax) and auditory processing profiles via the R statistical environment (Version 4.3.1; R Core Team, 2023). To construct models (MODELs 1–4; see below), the lme 4 package was used (Bates et al., 2021). In each model, participants' phonology and morphosyntax scores were used as dependent variables (DVs) reflecting two different facets of L2 learning outcomes. Fixed effects included participants' auditory processing scores (acuity, attention, integration) and types of L2 learning outcomes (phonology, morphosyntax).

---

[1] The participants completed the same vowel identification task three consecutive times. The results of their first attempt were used for the main analyses; those of the other repetitions will be reported in a different venue (with different research objectives).

Table 2

Summary of a Five-Factor Solution Based on a Factor Analysis of Acuity, Attention, and Integration Scores

Auditory processing task type	F1: attention	F2: integration	F3: pitch acuity	F4: formant acuity	F5: temporal acuity
Cumulative %	38.622	53.296	64.544	74.804	83.822
A. Acuity					
Pitch discrimination	−.244	−.114	**.920**	.100	.061
Formant discrimination	−.045	−.095	.094	**.987**	.070
Rise time discrimination	−.153	−.126	.062	.073	**.974**
B. Attention					
Pitch recognition	**.875**	.041	−.129	−.067	−.036
Formant recognition	**.805**	.202	−.126	−.013	−.154
Rise time recognition	**.839**	.191	−.099	−.004	−.071
C. Integration					
Melody reproduction	.047	**.795**	−.313	−.107	−.108
Rhythm reproduction	.326	**.790**	.117	−.028	−.067

Note. All loadings > .4 were highlighted in bold; acuity scores reveal the subtlety of differences perceptible to participants, with lower scores indicating more precise acuity. In contrast, attention and integration scores correlate positively with enhanced cognitive and motor skills, with higher scores representing greater capacities. F1 = first format; F2 = second format.

As noted in the literature review section, and as demonstrated in previous studies, our hypothesis was auditory processing could differentially predict the two different dimensions of L2 outcomes (phonology vs. morphosyntax). Specifically, we predicted that the differential effects of auditory processing would be more clearly observed in phonology (Kachlicka et al., 2019) than in morphosyntax (Saito, Macmillan, et al., 2022). Therefore, it was crucial not just to include test type, but also its interaction effects with auditory processing. For the rest of the analyses, therefore, we constructed the following model:

If interaction effects (i.e., auditory_processing: test_type) were statistically significant ($p < .05$), we further examined the simple slopes of auditory processing (continuous variables) at each level of test using the emmeans package (Lenth & Lenth, 2018). To that end, the following R code was used: emtrends (MODEL ~ test_type, var = "auditory processing").

Simulation Analyses

To evaluate the adequacy of our sample size for the linear mixed-effects regression analyses, simulation analyses were performed based on the two prior data sets: a sample of 40 Polish learners of English from Kachlicka et al. (2019) and a sample of 39 Spanish learners of English from Saito, Sun, et al. (2022). Each data set was analyzed using the SIMR package (Green & MacLeod, 2016). Dependent variables were participants' phonology and morphosyntax scores, which were assessed in relation to the interaction of four predictor variables: test type (phonology vs. morphosyntax tests), pitch acuity (pitch discrimination), formant acuity (formant discrimination), and temporal acuity (risetime discrimination). The model was then extended to include up to 200 participants and each model simulation was repeated 1,000 times. This approach allowed us to explore the necessary participant count to achieve a satisfactory statistical power (i.e., 80%, 90%).

In the Polish data set ($n = 40$), the regression model (DV ~ pitch_acuity × test_type + formant_acuity × test_type + temporal_acuity × test_type + (1|ID)) identified pitch acuity as a significant predictor ($b = -0.2946$, $SE = 0.1138$, $t = -2.558$, $p = .012$, $R^2_{conditional} = .687$, $R^2_{marginal} = .348$). However, the power analysis yielded 74.00% [64.27, 82.26], indicating that the original sample size ($n = 40$) was slightly below the lower end of the satisfactory power level, 80%. Upon extending the data set to 200 participants, the powerCurve function suggested that a sample size of 47 would yield a power of 90.00% [82.38, 95.10], and a sample size of 69 would yield a power of 94.00% [87.40, 97.77] (see Figure 1a).

For the Spanish data set ($n = 39$), the same regression model also identified pitch acuity as a significant predictor ($b = -0.20369$, $SE = 0.09608$, $t = -2.120$, $p = .038$, $R^2_{conditional} = .658$, $R^2_{marginal} = .233$). The power for this data set was insufficient at 50.00% [18.71, 81.29]. With an extended data set of 200 participants, the powerCurve function suggested a sample size of 91 for a power of 86.00% [77.63, 92.13], and a sample size of 112 for a power of 94.00% [87.40, 97.77] (see Figure 1b). In light of the varying degrees of conservativeness in the data sets (Polish being more conservative and Spanish less so), our simulations suggest a minimum sample size of 50 is needed to achieve 80% statistical power, and over 90 for 90% power. Therefore, the sample size of the current investigation ($n = 102$) seems sufficient for obtaining acceptable statistical power.

Main Analyses: Acuity_Only Versus Composite Models

As outlined in the Predictions section, our initial hypothesis proposed that while acuity factors alone could moderately explain the outcomes of both phonology and morphosyntax ($R^2_{marginal} = .348$ in Kachlicka et al., 2019; $R^2_{marginal} = .233$ in Saito, Sun, et al., 2022), the inclusion of attention and integration factors could account for additional variances. Therefore, we compared the null model (MODEL0) against the full model (MODEL1):

- MODEL0: DV ~ pitch_acuity × test_type + formant_acuity × test_type + temporal_acuity × test_type + (1|ID)

We hypothesized that the full model would be significantly different from the null model ($p < .05$) and have larger effect sizes than the previous studies—Kachlicka et al. (2019) ($R^2_{marginal} = .348$) and Saito, Sun, et al. (2022) ($R^2_{marginal} = .233$).

As demonstrated in Table 3, MODEL0 identified significant relationships between acuity (pitch, temporal) and L2 learning outcomes (phonology, morphosyntax), with medium to large effect sizes ($R^2_{marginal} = .385$). MODEL1 not only recognized acuity (pitch, temporal) as significant, but also added attention and integration as significant predictors of L2 learning outcomes. This

Figure 1

Results of Simulation Analyses Based on Previous Studies: n = *40 for Polish Learners of English From* Kachlicka et al. (2019) *(See 1A);* n = *39 for Spanish Learners of English From* Saito, Sun et al. (2022) *(See 1B)*
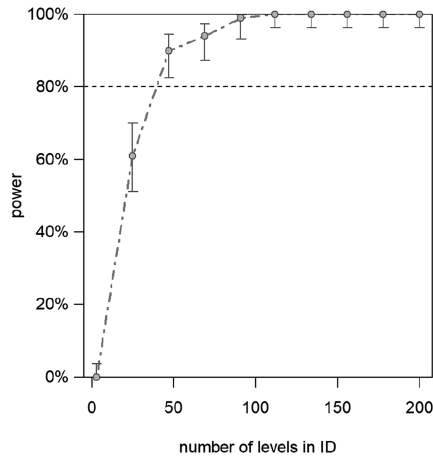
1A Polish   **1B Spanish**
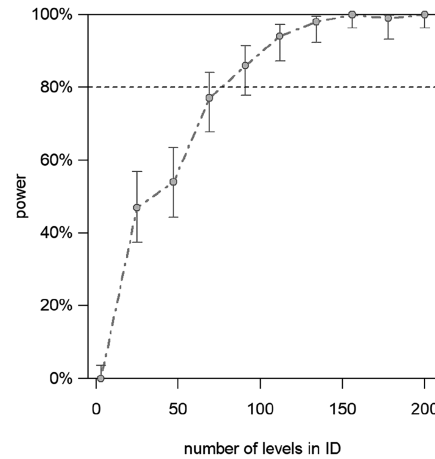
Note. The simulations suggest that to achieve 80% power, a sample size exceeding 50 is needed, and for 90% power, a sample size exceeding 90 is required. ID = identification number.

comprehensive auditory processing model, encompassing acuity, attention, and integration, explained a notably larger portion of the variance ($R^2_{\text{marginal}} = .519$). While both models highlighted a significant main effect of test type ($M_{\text{phonology}} = 82.5\%$ [80.9, 84.9] vs. $M_{\text{morphosyntax}} = 68.2\%$ [65.9, 70.5]), none of the interaction effects were statistically significant ($p > .05$). This implies that the impacts of auditory processing could be consistent across different task conditions (phonology, morphosyntax). As illustrated in Figure 2, the correlation coefficients between participants' auditory processing abilities (pitch acuity, temporal acuity, attention, integration) and their average proficiency scores (phonology, morphosyntax) ranged from small to medium ($r = -.202–.427$).

In order to test the hypothesis that the full model (MODEL1) would account for a greater proportion of the variance in the DV compared to the null model (MODEL0), a model comparison was conducted using the anova function in R. The analysis of variance (ANOVA) test revealed a significant difference between the two models, $\chi^2(2) = 43.354$, $p < .001$. This indicates that the addition of the attention and integration predictors in MODEL1 significantly improved the model's fit compared to MODEL0. Furthermore, the full model (MODEL1) yielded a lower Akaike information criterion (AIC; 1,551.1 vs. 1,515.7) and Bayesian information criterion (BIC; 1,584.3 vs. 1,562.2), further suggesting a better fit to the data than MODEL0. Therefore, the results supported our hypothesis that MODEL1 would provide a significantly better explanation of the data compared to MODEL0.

Main Analyses: Auditory Processing Versus Experience Variables

In our next hypothesis, we posited that the predictive power of auditory processing factors—including acuity, attention, and integration—on L2 learning outcomes would remain significant even

after accounting for the influence of experience-related factors. These experience-related factors comprise age of acquisition, LOR, current L2 use, length of formal English training, and music training. Strong correlations were found between participants' chronological age and age of arrival, $r = .819$, $p < .001$, confidence interval (CI) [0.742, 0.874]. To avoid the multicollinearity problem, only the latter was entered as a predictor into the regression model. To index participants' current L2 use as a single predictor, three thematically overlapping categories—participants' L2 use in home, work, and social settings—were averaged. As a result, no clear evidence of multicollinearity was observed (variance inflation factor = 1.040 to 1.440).

To investigate the roles of experience in L2 learning, a preliminary analysis was conducted by constructing another mixed effects regression model (i.e., experience-only model). Here, participants' L2 scores were used as DVs relative to the main and interaction effects of test_type (phonology, morphosyntax) and experience-related factors as independent variables (age of acquisition, LOR, current L2 use, length of formal English training, and music training). The experience-only model accounted for a relatively small amount of variance in L2 learning ($R^2_{\text{conditional}} = .610$, $R^2_{\text{marginal}} = .343$). As demonstrated in the previous literature (Slevc & Miyake, 2006), music training was identified as a significant predictor, $b = 5.177$, $SE = 2.457$, $t = 2.107$, $p = .036$.

To examine the composite effects of auditory processing (acuity, attention, and integration) on L2 learning, with experience-related variables controlled for, MODEL2 was constructed using the residual DV scores from which the experience-related variables (age of acquisition, LOR, current L2 use, length of formal English training, and music training) were factored out. As indicated in Table 3, MODEL2 remained strongly predictive of L2 learning outcomes even after accounting for participants' individual differences in L2 learning experience ($R^2_{\text{conditional}} = .576$, $R^2_{\text{marginal}} = .493$).

Table 3

Summary of Mixed-Effects Modeling Analyses of Auditory Processing and L2 Phonological and Morphosyntactic Proficiency

	Fixed effects	Estimate (b)[a]	SE	t	p	Random effects	Variances	SD	$R^2_{conditional}$	$R^2_{mariginal}$
MODEL0	Intercept	82.590	1.077	76.646	<.001	Participant ID	41.93	6.475	.602	.385
	Pitch acuity	−3.814	1.088	−3.505	<.001*	Residual	76.51	8.747		
	Test type	−14.341	1.224	−11.709	<.001*					
	Formant acuity	−1.330	1.088	−1.223	.223					
	Temporal acuity	−2.682	1.088	−2.465	.014*					
	Pitch: test type	−0.610	1.237	−0.493	.622					
	Formant: test type	0.754	1.237	0.61	.543					
	Temporal: test type	1.295	1.237	1.047	.297					
MODEL1	Intercept	82.590	0.9547	86.513	<.001*	Participant ID	16.82	4.102	.606	.519
	Pitch acuity	−3.814	0.9642	−3.957	<.001*	Residual	76.14	8.726		
	Test type	−14.341	1.2218	−11.738	<.001*					
	Formant acuity	−1.330	0.9642	−1.380	.169					
	Temporal acuity	−2.682	0.9642	−2.782	.005*					
	Attention	2.514	0.9642	2.608	.009*					
	Integration	5.238	0.9642	5.433	<.001*					
	Pitch: test type	−0.610	1.234	−0.494	.622151					
	Formant: test type	0.754	1.234	0.611	.542616					
	Temporal: test type	1.295	1.234	1.050	.296356					
	Attention: test type	0.435	1.234	0.353	.725003					
	Integration: test type	−1.893	1.234	−1.534	.128229					
MODEL2[b]	Intercept	7.170	0.945	7.588	<.001*	Participant ID	14.96	3.868	.576	.493
	Pitch acuity	−3.544	0.954	−3.713	<.001*	Residual	76.14	8.726		
	Test type	−14.341	1.221	−11.738	<.001*					
	Formant acuity	−1.200	0.954	−1.257	.210					
	Temporal acuity	−2.514	0.954	−2.634	.009*					
	Attention	2.016	0.954	2.113	.035*					
	Integration	4.382	0.954	4.591	<.001*					
	Pitch: test type	−0.610	1.234	−0.494	.622					
	Formant: test type	0.754	1.234	0.611	.542					
	Temporal: test type	1.295	1.234	1.050	.296					
	Attention: test type	0.435	1.234	0.353	.725					
	Integration: test type	−1.893	1.234	−1.534	.128					
MODEL3[c]	Intercept	7.170	0.9685	7.404	<.001*	Participant ID	19.54	4.420	.573	.463
	Pitch acuity	−3.511	0.9781	−3.59	<.002*	Residual	76.14	8.726		
	Test type	−14.341	1.2218	−11.738	<.001*					
	Formant acuity	−1.076	0.9781	−1.101	.270					
	Temporal acuity	−2.775	0.9781	−2.838	.005*					
	Attention	0.888	0.9781	0.909	.364					
	Integration	3.946	0.9781	4.034	<.001*					
	Pitch: test type	−0.610	1.234	−0.494	.622					
	Formant: test type	0.754	1.234	0.611	.542					
	Temporal: test type	1.295	1.234	1.050	.296					
	Attention: test type	0.435	1.234	0.353	.725					
	Integration: test type	−1.893	1.234	−1.534	.128					

Note. L2 = second language; DV = dependent variable; EFL = English-as-a-foreign-language. [a] Negative correlation expected between auditory processing and L2 outcomes in acuity (lower scores = more precise acuity); positive correlation expected for attention and integration (higher scores = greater cognitive/motor abilities). [b] For DV represented by residual scores, with the effects of experience-related variables accounted for and removed (age of acquisition, length of residence, daily L2 use, length of training, and music training). [c] For DV represented by residual scores, with the effects of working memory accounted for and removed (forward span and backward span). * p < .05.

As demonstrated in our first hypothesis, the composite model of auditory processing (MODEL1) exhibited a significantly stronger predictive power than the acuity-only model (MODEL0). Our second hypothesis posited that a significant difference would continue to exist between the composite and the acuity-only models, even after factoring out experience-related variables. To evaluate this, MODEL2 (the composite model excluding the experience effects) was compared again with MODEL0 (the acuity-only model). To align with MODEL2, where the residual scores (DV_residuals_experience) were used as DVs, MODEL0 was revised as MODEL0_experience, using the residual scores as DVs.

Results demonstrated that MODEL2 provided a significantly better fit to the data than MODEL0_experience, $\chi^2(2) = 31.679$, $p < .001$. The AIC and BIC values for MODEL2 were also lower (AIC = 1,512.2, BIC = 1,558.7) compared to MODEL0 (AIC = 1,535.9, BIC = 1,569.1), indicating that MODEL2 was more parsimonious. This in turn suggests that adding two cognitive components of auditory processing (attention, integration) significantly strengthened the predictive power of the acuity-only model for L2 learning outcomes, even after the potential effects of the relevant experience-related factors were removed.

Figure 2

Correlations Between L2 Proficiency (y Axis) and Factor Auditory Processing Scores (x Axis)
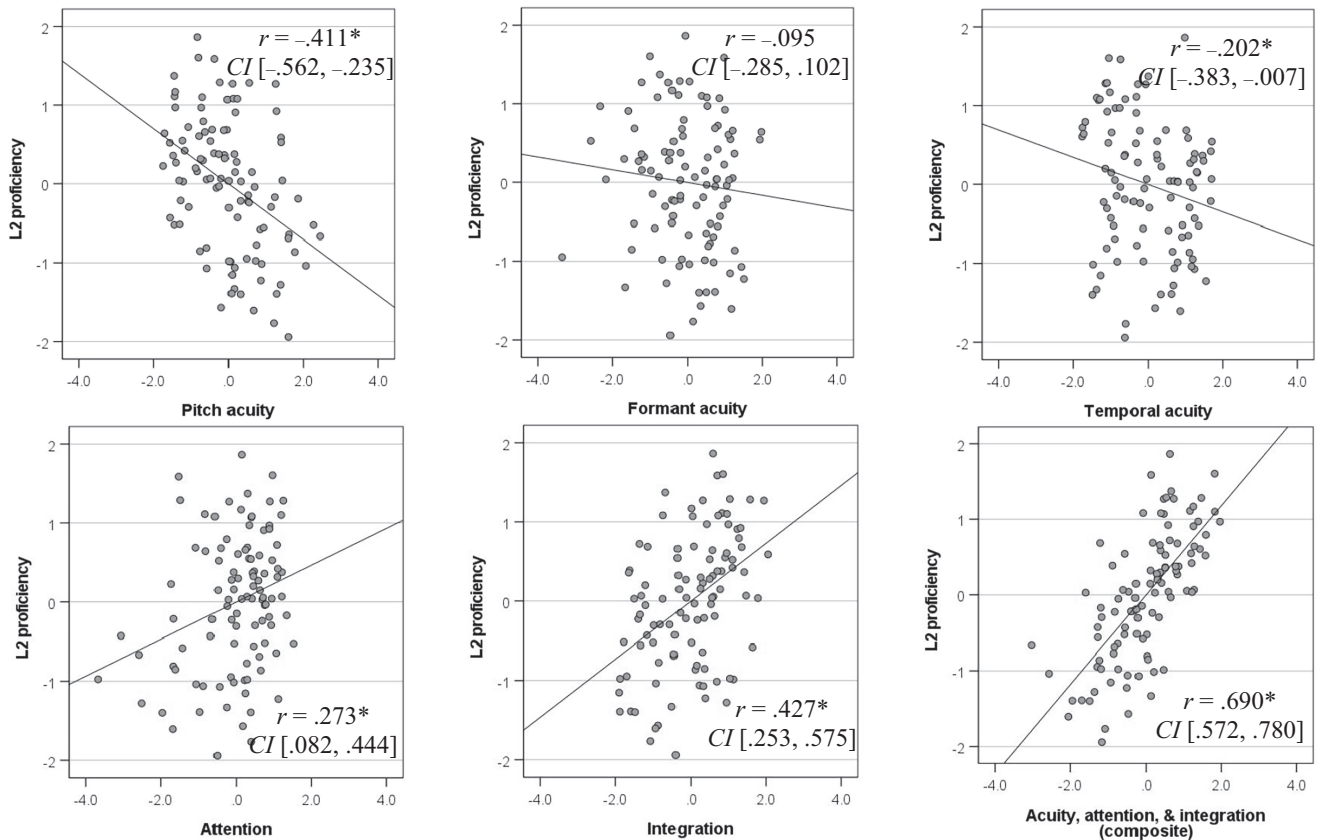
Note. The L2 proficiency scores were calculated by standardizing and averaging participants' phonology and morphosyntactic scores. The composite auditory processing scores (presented in right bottom) were generated as standardized predicted values based on the regression model comprising acuity, attention, and integration as predictors of L2 proficiency. CI = confidence interval.

Main Analyses: Auditory Processing Versus Working Memory Variables

Model 3 was constructed to investigate the roles of acuity, attention, and integration in L2 learning outcomes when their overlap with working memory was statistically controlled for. Although the two working memory test scores (forward and backward digit span) were weakly correlated ($r = .282$, $p = .004$, CI [0.093, 0.452]), multicollinearity problems were not identified among the predictors entered into the model (variance inflation factor = 1.009 to 1.345).

As discussed earlier, the relationship between working memory and L2 learning behaviors has been extensively explored (e.g., Linck et al., 2013). To confirm the effects of working memory in our data set, we conducted a preliminary analysis by building another mixed effects regression model. This model used participants' L2 scores as DVs, with forward and backward span scores serving as predictors. Unsurprisingly, the results demonstrated a significant predictive power of the backward span ($b = 0.136$, $SE = 0.043$, $t = 3.122$, $p = .002$) and a marginally significant power of the forward span ($b = 0.109$, $SE = 0.058$, $t = 1.877$, $p = .063$). The working memory model produced medium-to-large effects

comparable to those of auditory processing ($R^2_{\text{conditional}} = .599$, $R^2_{\text{marginal}} = .351$).

To examine the independent contribution of auditory processing to L2 learning outcomes, the residual DV scores were generated by factoring out the combined effects of the working memory variables (forward and backward digit span). MODEL3 was constructed below.

As summarized in Table 3, MODEL3 remained strongly predictive of L2 learning outcomes even when participants' working memory profiles were statistically controlled for ($R^2_{\text{conditional}} = .573$, $R^2_{\text{marginal}} = .463$). Notably, the predictive power of attention failed to reach statistical significance ($p = .364$). In fact, participants' attention scores were moderately correlated with forward digit span ($r = .349$ [0.165, 0.510]) and backward digit span ($r = .332$ [0.146, 0.495]; see the online supplemental material S3). The findings suggests that (a) working memory and attention, both of which are classified as domain-general cognitive abilities, share some variances in their interrelations with L2 learning outcomes and (b) acuity and integration exist as constructs separate from working memory.

Our final hypothesis posited that the initially tested predictive power of auditory processing for L2 learning outcomes (i.e., MODEL1 for

acuity, attention, and integration > MODEL0 for acuity only) would continue to remain significant even after adjusting for the effects of working memory (i.e., MODEL3 for acuity, attention, and integration minus working memory). To test this, we compared MODEL3 to MODEL0 using the anova function in R. Considering that residual scores were used in MODEL3 (DV_residuals_wm), the same residual scores were employed as DVs for MODEL0 (i.e., MODEL0_wm).

Results of the ANOVA showed that MODEL3 provided a significantly better fit for the data than MODEL0, $\chi^2(2) = 19.899$, $p < .001$. Additionally, the AIC and BIC values for MODEL3 (AIC = 1,520.7; BIC = 1,565.7) were lower than those for MODEL0_wm (AIC = 1,532.5; BIC = 1,567.1). Thus, even when the influence of working memory was statistically accounted for, the components of auditory processing—acuity, attention, and integration—sustained their significant predictive power for L2 learning outcomes. This underscores the independent contribution of these auditory processing factors (acuity and integration in particular) to L2 proficiency, further suggesting that their role in language acquisition is not merely a byproduct of their relationship with working memory abilities.

All the R codes and relevant data for simulation, mixed effects regression, and model comparison analyses are provided in the online supplemental material S4.

Discussion

Recently, there has been a growing amount of research showing that domain-general auditory processing serves as a foundation of language learning throughout the lifespan (Mueller et al., 2012) and thus explains some of the variances in adult L2 learning outcomes (Saito & Tierney, 2022). Extending previous research which has operationalized auditory processing as participants' ability to discriminate small acoustic differences at sensory levels, the interaction model has proposed a reconceptualization of auditory processing as a set of perceptual, cognitive, and motoric abilities (Kraus & Banai, 2007). Under this model, auditory processing includes not only the ability to notice acoustic details but also the ability to attend to relevant acoustic dimensions while ignoring irrelevant dimensions (attentional control) and the ability to convert acoustic information into motor action (audio-motor integration). In the current study, hypotheses were tested with 102 Chinese learners of English in the United Kingdom. We first assessed the perceptual, cognitive, and motoric aspects of auditory processing via three behavioral tasks (discrimination, repetition detection, and reproduction). Subsequently, we examined these three constructs in relation to participants' biographical backgrounds and working memory. Finally, we explored the link between all of these factors (auditory processing, biographical background, and working memory) and the phonological and morphosyntactic aspects of L2 proficiency.

The results of factor analyses showed that participants' performance in the auditory processing test battery appeared to tap into three dimension-specific abilities to perceive pitch, formant, and temporal details (pitch, formant, and temporal acuity) and two dimension-general abilities to direct selective attention to individual acoustic parameters (attention) and convert the perceived acoustic information into motor action (audio-motor integration). The correlation analyses further showed that participants' attention and integration scores demonstrated a significant overlap with working memory ($r = .320$ to $.349$, $p \le .001$), while the link between their

acuity scores and working memory did not reach statistical significance ($p > .05$). The influence of biographical background (length of training, length of immersion, and age) on auditory processing was minor, suggesting that auditory processing, like other aptitudes (Doughty, 2019), is a rather stable trait, which is unlikely to change dramatically over time.

On the one hand, as in the previous literature in L1 contexts (Surprenant & Watson, 2001) and L2 contexts (Saito & Tierney, 2022), our findings support a view that measurement results for the more perceptual aspect of auditory processing (acuity) differ greatly depending on the type of acoustic information perceived, with some listeners being sensitive to spectral information (pitch and formant) and others to temporal information (duration and amplitude rise time). Contrary to previous arguments (Snowling et al., 2018; but see Saito, Haining, et al., 2022), however, our findings indicate that the acuity aspect of auditory processing is not closely related to other cognitive abilities, such as phonological short-term and executive working memory.

On the other hand, measurements of the more cognitive and motoric aspects of auditory processing (attention and integration) are comparable across different acoustic dimensions. This supports a view that individuals can possess essentially distinct, dimension-general auditory processing abilities at higher-order levels, such as attention (Holt et al., 2018) and integration (Flaugnacco et al., 2014), regardless of their varied sensives to each acoustic dimension at lower-order levels (pitch, formant, and temporal acuity). Importantly, given that attention and integration tasks necessarily involved the perception of acoustic details, performance on these tasks may have been affected by other cognitive abilities (such as working memory) to some degree.

To further examine the complex relationship between acuity, attention, and integration, we explored their unique contribution to L2 learning outcomes. The results of linear mixed-effects analyses demonstrated that the perceptual, cognitive, and motoric model of auditory processing demonstrated large predictive power for L2 proficiency ($R^2_{\text{marginal}} = .519$) and remained significant even after all the relevant biographical background and working memory variables were controlled for ($R^2_{\text{marginal}} = .463$, $.493$, respectively). The patterns reported here not only concur with the extant literature which has noted a small-to-medium link between acuity and L2 learning ($R^2_{\text{marginal}} = .348$ in Kachlicka et al., 2019; $R^2_{\text{marginal}} = .233$ in Saito, Macmillan, et al., 2022) but also lend empirical support to our hypothesis that the inclusion of neighboring abilities (attention and integration) can explain additional variance in language learning outcomes (Tierney & Kraus, 2014). Our argument here concurs with ongoing claims that auditory processing, comprised of perceptual, cognitive, and motoric components, serves as a bottleneck for language learning (Kraus & Banai, 2007).

As auditory processing is more directly related to the phonological aspects of language learning than to the morphosyntactic ones, we hypothesized that the predictive power of auditory processing would be stronger for L2 phonological outcomes than for morphosyntactic outcomes. However, our results did not reveal any significant interaction effects in any context, suggesting that auditory processing plays an equally important role in both the lower and higher-order dimensions of L2 acquisition. While we acknowledge the methodological discussions surrounding the extent to which the L2 tasks used in this study (identification, judgments) can truly capture L2 phonological and morphosyntactic proficiency

(e.g., Saito & Plonsky, 2019), the main effects (as opposed to interaction effects) of auditory processing in our study provide the first empirical support for the hypothesis regarding the pivotal role of auditory processing in morphosyntax (Goad & White, 2019). Future longitudinal studies should delve deeper into how auditory processing differentially impacts the acquisition of various morphosyntactic features, considering their varied prosodic profiles and conceptual complexities (e.g., Henry, 2023).

One might question to what degree the contribution of auditory processing might overlap with other domain-general abilities (Snowling et al., 2018). Interestingly, while the composite model of auditory processing accounted for a large amount of variance in L2 learning outcomes ($R^2_{\mathrm{marginal}} = .519$), the findings from the aforementioned factor analyses revealed some overlaps between auditory processing and working memory, even though no audio stimuli were involved in the latter tasks ($r = .320$ to $.349$, $p < .001$). Importantly, the outcomes from the mixed effects modeling regression analyses showed that whereas the predictive power of the composite model remained significant even after accounting for the effects of working memory ($R^2_{\mathrm{marginal}} = .463$), the role of dimension-selective attention in L2 became nonsignificant ($p = .364$). This suggests that working memory and attention could be interrelated as similar domain-general cognitive abilities; that they may interact closely with each other to influence L2 learning outcomes; and that their relationship to L2 learning differs from that of acuity and integration.

In summary, we propose (a) that at least two of the auditory processing effects observed in this study (acuity, integration) could be auditory-specific (distinguishable from other related cognitive abilities) but (b) that one's ability to attend to domain-general acoustic dimensions may involve other cognitive abilities (e.g., working memory). Although these findings lend empirical support to the view that auditory processing can uniquely contribute to language learning throughout the lifespan (Goswami, 2015), further research is needed to precisely identify the mechanisms that underlie the measurement of dimension-selective attention and our proposed task formats (repetition detection).

Interestingly, although most of the auditory processing in this study measures significantly predicted L2 outcomes, formant acuity did not achieve statistical significance in any context. This could be arguably due to the fact that we assessed L2 learning outcomes through the composite tests, which included vowel and prosody identification and grammaticality judgments. While formant acuity is believed to be directly relevant to L2 segmental accuracy (Saito et al., 2020), the relevance of this specific acoustic cue—quantified as sensitivity to F2 variation between 1,500 and 1,700 Hz—may not extend to all areas of L2 learning, such as morphosyntactic acquisition. Conversely, evidence suggests that L2 learners lean on prosodic information like pitch and amplitude, given their integral roles in English prosody (Trofimovich & Baker, 2006) and morphosyntax (Goldschneider & DeKeyser, 2001). The results from this study suggest a dimension-specific link between auditory processing and L2 learning outcomes, underscoring the idea that learners' sensitivities to certain acoustic dimensions may bear more relevance to L2 learning than others. For the dimension-specific relationship between F2 and F3 variation and the acquisition of English [r] and [l] by Japanese speakers, see Saito, Kachlicka, et al. (2022).

To close, we would like to provide a range of future directions that researchers can pursue to unravel the relationship between auditory processing and L2 speech learning. First and foremost, if we take the stance that auditory processing alone can explain a relatively large amount of the variance in L2 learning outcomes ($R^2_{\mathrm{marginal}} = .519$), it is reasonable to wonder if the provision of focused training could enhance auditory processing, and, as a result, impact the degree of success in L2 acquisition. In L1 acquisition research, a few hours of auditory processing training have been found to enhance auditory processing abilities; whether such training could ultimately benefit language learning, however, remains unclear (Merzenich, et al., 1996 for temporal acuity; Micheyl et al., 2006 for pitch acuity). In a recent study, three hours of formant acuity training significantly boosted Japanese listeners' auditory processing and improved their L2 English vowel perception abilities with small-to-medium effects (Saito, Petrova, et al., 2022). Notably, the existing literature in both L1 and L2 acquisition has been exclusively concerned with the training of perceptual acuity (participants are guided to enhance their abilities to perceive acoustic details via an AXB discrimination task with immediate feedback). It would be interesting if we could expand the scope of auditory processing training to include other aspects of auditory processing, such as attention and integration.

Another promising research direction concerns aptitude-treatment interaction (DeKeyser, 2012). It has been argued that individuals with certain perceptual–cognitive abilities can benefit more from certain types of L2 learning. Thus, aptitude measures can be used as a diagnostic tool for identifying profile-matched training methods. In L2 grammar literature, for example, there is some evidence that those with greater explicit aptitude (e.g., working memory, grammatical sensitivity) are more likely to show gains when they engage in explicit training (e.g., S. Li, 2016 for a meta-analysis), and those with greater implicit aptitude (e.g., procedural memory) can benefit more from implicit training (e.g., Yilmaz & Granena, 2021; for a comprehensive review, Wen & Skehan, 2021). In terms of auditory processing (the main focus of this article), previous studies have demonstrated that those with more precise acuity can benefit from input-based phonetic training (Lengeris & Hazan, 2010). Future studies can further explore what type of training can most benefit L2 learners with greater attentional control (e.g., training with and without noise; cf., Mora et al., 2022) and audio-motor integration (e.g., output-based training; cf., Shao et al., 2023).

The use of a composite model of a"Iit'ry processing (acuity, attention, and integration) in future research could greatly inform the revised view aptitude-treatment interaction espoused by proponents of the Auditory Precision Hypothesis (Perrachione et al., 2011; Ruan & Saito, 2023; Saito & Tierney, 2022). In this view, both strengths and weaknesses in aptitude can be used to provide a detailed set of profile-matched training recommendations. Stronger aptitude profiles are better suited to different types of phonetic training, while weaker aptitude profiles may be better suited to various types of auditory training designed to mitigate or prevent any detrimental effects of aptitude deficits on L2 learning. Different types of auditory processing tests can be used to diagnose strong as well as weak auditory processing abilities for the purpose of providing the most appropriate training.

We are aware of the limitations in terms of the generalizability of the conclusions drawn from this study. It is important to replicate the suggestive patterns based on experienced Chinese English learners in the United Kingdom in various L1 and L2 contexts. Past research indicates that L2 speech learning can be influenced by both auditory and visual input. For instance, L1 English listeners utilize both

acoustic information (e.g., F3 variation) and lip rounding to perceive English [r] (Kawase et al., 2014). Considering that an individual's visual processing abilities must explain some variance in L2 speech perception, it would be compelling to investigate how a spectrum of perceptual–cognitive abilities beyond auditory processing contribute to successful L2 speech learning.

Conclusion

Adopting the interaction view which reconceptualizes domain-general auditory processing as a combination of perceptual, cognitive, and motoric abilities (Kraus & Banai, 2007), we examined and confirmed the independent relationship between the three broad abilities to perceive acoustic details (acuity), to selectively focus on specific acoustic dimensions (attention), and to convert sound sequences into motor actions (integration). Among 102 Chinese English speakers, not only were their scores clustered into three dimension-specific latent variables (formant, pitch, and rise time acuity) and two dimension-general latent variables (attention, integration), but they also differentially predicted L2 learning outcomes. These findings remained same when accounting for participants' experience-related variables (e.g., age of acquisition, LOR, and music training) and working memory profiles (forward and digit span), though some overlaps were noted between attention and working memory as both are cognitive abilities. The findings suggest that future researchers should consider auditory processing as a multilayered phenomenon and thus approach the connection between audition and language learning at three different levels (perceptual, cognitive, and motoric).

References

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

Ahissar, M., Lubin, Y., Putter-Katz, H., & Banai, K. (2006). Dyslexia and the failure to form a perceptual anchor. *Nature Neuroscience*, *9*(12), 1558–1564. https://doi.org/10.1038/nn1800

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

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2021). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

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.

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

Brekelmans, G., Lavan, N., Saito, H., Clayards, M., & Wonnacott, E. (2022). Does high variability training improve the learning of non-native phoneme contrasts over low variability training? A replication. *Journal of Memory and Language*, *126*, Article 104352. https://doi.org/10.1016/j.jml.2022.104352

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

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

Cobb, T. (2012). *The Compleat Lexical Tutor*. https://www.lextutor.ca/vp/

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*(2), 218–236. 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

DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, *62*(S2), 189–200. https://doi.org/10.1111/j.1467-9922.2012.00712.x

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

Doughty, C. J. (2019). Cognitive language aptitude. *Language Learning*, *69*(S1), 101–126. https://doi.org/10.1111/lang.12322

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*(2), 164–194. https://doi.org/10.1093/applin/aml015

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), 551–585. https://doi.org/10.1017/S0272263104040021

Espy-Wilson, C. (1993). Acoustic measures for linguistic features distinguishing the semi-vowels/w y r l/in American English. *Journal of the Acoustical Society of America*, *92*(2), 736–757. https://doi.org/10.1121/1.403998

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

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America*, *93*(3), 1589–1608. https://doi.org/10.1121/1.406818

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470. https://doi.org/10.1006/jpho.1997.0052

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /L/. *The Journal of the Acoustical Society of America*, *99*(2), 1161–1173. https://doi.org/10.1121/1.414884

Ghaffarvand Mokari, P., & Werner, S. (2018). On the role of cognitive abilities in second language vowel learning. *Language and Speech*, *62*(2), 260–280. https://doi.org/10.1177/0023830918764517

Goad, H., & White, L. (2019). Prosodic effects on L2 grammars. *Linguistic Approaches to Bilingualism*, *9*(6), 769–808. https://doi.org/10.1075/lab.19043.goa

Godfroid, A., Loewen, S., Jung, S., Park, J.-H., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye–movement patterns. *Studies in Second Language Acquisition*, *37*(2), 269–297. https://doi.org/10.1017/S0272263114000850

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

Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & Devin McAuley, J. (2015). Musical rhythm discrimination explains individual differences in grammar skills in children. *Developmental Science*, *18*(4), 635–644. https://doi.org/10.1111/desc.12230

Goswami, U. (2015). Sensory theories of developmental dyslexia: Three challenges for research. *Nature Reviews Neuroscience*, *16*(1), 43–54. https://doi.org/10.1038/nrn3836

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

Grahn, J. A. (2012). Neural mechanisms of rhythm perception: Current findings and future perspectives. *Topics in Cognitive Science*, *4*(4), 585–606. https://doi.org/10.1111/j.1756-8765.2012.01213.x

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, *19*(5), 893–906. https://doi.org/10.1162/jocn.2007.19.5.893

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, *39*(5), 350–365. https://doi.org/10.1016/j.jcomdis.2006.06.013

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Prentice Hall.

Henry, N. (2023). The additive use of prosody and morphosyntax in L2 German. *Studies in Second Language Acquisition*, *45*(2), 348–369. https://doi.org/10.1017/S0272263122000092

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, *69*(3), 407–422. https://doi.org/10.1016/j.neuron.2011.01.019

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, *108*(6), 3013–3022. https://doi.org/10.1121/1.1323463

Holt, L. L., Tierney, A. T., Guerra, G., Laffere, A., & Dick, F. (2018). Dimension-selective attention as a possible driver of dynamic, context-dependent re-weighting in speech processing. *Hearing Research*, *366*, 50–64. https://doi.org/10.1016/j.heares.2018.06.014

Hornickel, J., Knowles, E., & Kraus, N. (2012). Test–retest consistency of speech-evoked auditory brainstem responses in typically-developing children. *Hearing Research*, *284*(1–2), 52–58. https://doi.org/10.1016/j.heares.2011.12.005

Humes, L. E., Lee, J. H., & Coughlin, M. P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *The Journal of the Acoustical Society of America*, *120*(5), 2926–2937. https://doi.org/10.1121/1.2354070

Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *The Journal of the Acoustical Society of America*, *132*(6), 3950–3964. https://doi.org/10.1121/1.4765076

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/–/L/ to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267–3278. https://doi.org/10.1121/1.2062307

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47–B57. https://doi.org/10.1016/S0010-0277(02)00198-1

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), 914–934. 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*, Article 104481. https://doi.org/10.1016/j.cognition.2020.104481

Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, *24*(1), 131–161. https://doi.org/10.1017/S0142716403000079

Joanisse, M. F., & Seidenberg, M. S. (1998). Specific language impairment: A deficit in grammar or processing? *Trends in Cognitive Sciences*, *2*(7), 240–247. https://doi.org/10.1016/S1364-6613(98)01186-3

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23. https://doi.org/10.1006/cogp.1995.1010

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*(2), 252–293. https://doi.org/10.1016/0010-0285(92)90009-Q

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

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement*, *34*(1), 111–117. https://doi.org/10.1177/001316447403400115

Kalashnikova, M., Goswami, U., & Burnham, D. (2019). Delayed development of phonological constancy in toddlers at family risk for dyslexia. *Infant Behavior and Development*, *57*, Article 101327. https://doi.org/10.1016/j.infbeh.2019.101327

Kawase, S., Hannah, B., & Wang, Y. (2014). The influence of visual speech information on the intelligibility of English consonants produced by non-native speakers. *The Journal of the Acoustical Society of America*, *136*(3), 1352–1362. https://doi.org/10.1121/1.4892770

Kim, D., Clayards, M., & Goad, H. (2018). A longitudinal study of individual differences in the acquisition of new vowel contrasts. *Journal of Phonetics*, *67*, 1–20. https://doi.org/10.1016/j.wocn.2017.11.003

Kondaurova, M. V., & Francis, A. L. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods. *Journal of Phonetics*, *38*(4), 569–587. https://doi.org/10.1016/j.wocn.2010.08.003

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*, 40–57. https://doi.org/10.1016/j.wocn.2016.08.006

Kraus, N., & Banai, K. (2007). Auditory-processing malleability: Focus on language and music. *Current Directions in Psychological Science*, *16*(2), 105–110. https://doi.org/10.1111/j.1467-8721.2007.00485.x

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

Lenth, R., & Lenth, M. R. (2018). Package 'lsmeans'. *The American Statistician*, *34*(4), 216–221. https://doi.org/https://cran.r-hub.io/web/packages/lsmeans/lsmeans.pdf

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

Li, M., & DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 mandarin tone-word learning. *Studies in Second Language Acquisition*, *39*(4), 593–620. https://doi.org/10.1017/S0272263116000358

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, *38*(4), 801–842. https://doi.org/10.1017/S027226311500042X

Liberman, A. M., & Mattingly, I. G. (1985). The Motor Theory of Speech Perception revised. *Cognition*, *21*(1), 1–36. https://doi.org/10.1016/0010-0277(85)90021-6

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, *32*(4), 451–454. https://doi.org/10.1121/1.1908095

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

Liu, C., Jin, S.-H., & Chen, C.-T. (2014). Durations of American English vowels by native and non-native speakers: Acoustic analyses and perceptual effects. *Language and Speech*, *57*(2), 238–253. https://doi.org/10.1177/0023830913507692

Loewen, S., & Gonulal, T. (2015). Exploratory factor analysis and principal components analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 182–212). Routledge. https://doi.org/10.4324/9781315870908-9

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. M. (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

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*(1–2), 36–47. https://doi.org/10.1016/j.heares.2006.05.004

Mora, J. C., Ortega, M., Mora-Plaza, I., & Aliaga-García, C. (2022). Training the pronunciation of L2 vowels under different conditions: The use of non-lexical materials and masking noise. *Phonetica*, *79*(1), 1–43. https://doi.org/10.1515/phon-2022-2018

Mora-Plaza, I., Ortega, M., & Mora, J. C. (2022). High-variability phonetic training under different conditions: Individual differences in auditory attention control. In V. Sardegna & A. Jarosz (Eds.), *Theoretical and practical developments in English speech assessment, research, and training* (pp. 241–260). Springer.

Mora-Plaza, I., Saito, K., Suzukida, Y., Dewaele, J.-M., & Tierney, A. (2022). *Tools for second language speech research and teaching*. https://sla-speech-tools.com.

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

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

Nguyễn, T. A.-T., Ingram, C. L. J., & Pensalfini, J. R. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics*, *36*(1), 158–190. https://doi.org/10.1016/j.wocn.2007.09.001

Nosek, B. A., Shaw, L. C., Errington, T. M., Pfeiffer, N., Mellor, D. T., Brooks, R. E., III, Rice, A., & Litherland, D. M. (2017, March 6). *Center for open science: Strategic plan*. https://doi.org/10.31219/osf.io/x2w9h

Olsthoorn, N. M., Andringa, S., & Hulstijn, J. H. (2014). Visual and auditory digit-span performance in native and non-native speakers. *International Journal of Bilingualism*, *18*(6), 663–673. https://doi.org/10.1177/1367006912466314

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: The action simulation for auditory prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, *8*, Article 57. https://doi.org/10.3389/fnsys.2014.00057

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (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

Plonsky, L., Marsden, E., Crowther, D., Gass, S. M., & Spinner, P. (2020). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research*, *36*(4), 583–621. https://doi.org/10.1177/0267658319828413

Povel, D.-J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, *2*(4), 411–440. https://doi.org/10.2307/40285

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, *103*(20), 7865–7870. https://doi.org/10.1073/pnas.0509989103

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*(1), 435–446. https://doi.org/10.1121/10.0003330

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://doi.org/https://cran.r-project.org/

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

Ruan, Y., & Saito, K. (2023). Domain-general auditory difficulties limit instructed second language speech learning: Communicative focus on phonetic form revisited. *System*, *114*, Article 103020. https://doi.org/10.1016/j.system.2023.103020

Saito, K., Haining, C., Suzukida, Y., Dardon, D., Suzuki, Y., Jeong, H., Révész, A., Sugiura, M., & Tierney, A. (2022). 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*, *25*(5), 856–868. https://doi.org/10.1017/S1366728922000153

Saito, K., Kachlicka, M., Suzukida, Y., Petrova, K., Lee, B., & Tierney, A. (2022). Auditory precision hypothesis-L2: Dimension-specific relationships between auditory processing and second language segmental learning in adulthood. *Cognition*, *229*, Article 105236. https://doi.org/10.1016/j.cognition.2022.105236

Saito, K., Macmillan, K., Kroeger, S., Magne, V., Takizawa, K., Kachlicka, M., & Tierney, A. (2022). Roles of domain-general auditory processing in spoken second-language vocabulary attainment in adulthood. *Applied Psycholinguistics*, *43*(3), 581–606. https://doi.org/10.1017/S0142716422000029

Saito, K., Petrova, K., Kachlicka, M., Suzukida, Y., & Tierney, A. (2022). Training auditory processing promotes second language speech acquisition. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(12), 1410–1426. https://doi.org/10.1037/xhp0001042

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., Kachlicka, M., Alayo, J. R. C., Nakata, T., & 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., Sun, H., & Tierney, A. (2020). Domain-general auditory processing as a perceptual-cognitive anchor of L2 pronunciation learning in adulthood: A longitudinal study. *Applied Psycholinguistics*, *41*(5), 1083–1112. https://doi.org/10.1017/S0142716420000491

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., & Tierney, A. (2022). Domain-general auditory processing as a conceptual and measurement framework for second language speech learning aptitude: A test-retest reliability study. *Studies in Second Language Acquisition*, 1–25. https://doi.org/10.1017/S027226312200047X

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

Schulte-Körne, G., & Bruder, J. (2010). Clinical neurophysiology of visual and auditory processing in dyslexia: A review. *Clinical Neurophysiology*, *121*(11), 1794–1809. https://doi.org/10.1016/j.clinph.2010.04.028

Shao, Y., Saito, K., & Tierney, A. (2023). How does having a good ear promote instructed second language pronunciation development? Roles of domain-general auditory processing in choral repetition training. *TESOL Quarterly*, *57*(1), 33–63. https://doi.org/10.1002/tesq.3120

Simmonds, A. J., Leech, R., Iverson, P., & Wise, R. J. (2014). The response of the anterior striatum during adult human vocal learning. *Journal of Neurophysiology*, *112*(4), 792–801. https://doi.org/10.1152/jn.00901.2013

Simmonds, A. J., Wise, R. J., & Leech, R. (2011). Two tongues, one brain: Imaging bilingual speech production. *Frontiers in Psychology*, *2*, Article 166. https://doi.org/10.3389/fpsyg.2011.00166

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, *17*(8), 675–681. https://doi.org/10.1111/j.1467-9280.2006.01765.x

Smith, J. O. (2007). *Introduction to digital filters: With audio applications* (Vol. 2) [Published online]. 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*(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

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

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*(4), 747–790. https://doi.org/10.1111/lang.12241

Symons, A. E., Dick, F., & Tierney, A. T. (2021). Dimension-selective attention and dimensional salience modulate cortical tracking of acoustic dimensions. *NeuroImage*, *244*, Article 118544. https://doi.org/10.1016/j.neuroimage.2021.118544

Talcott, J. B., Witton, C., McLean, M. F., Hansen, P. C., Rees, A., Green, G. G. R., & Stein, J. F. (2000). Dynamic sensory sensitivity and children's word decoding skills. *Proceedings of the National Academy of Sciences*, *97*(6), 2952–2957. https://doi.org/10.1073/pnas.040546597

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*, Article 105196. https://doi.org/10.1016/j.jecp.2021.105196

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

Tierney, A., White-Schwoch, T., MacLean, J., & Kraus, N. (2017). Individual differences in rhythm skills: Links with neural consistency and linguistic ability. *Journal of Cognitive Neuroscience*, *29*(5), 855–868. https://doi.org/10.1162/jocn_a_01092

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

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition*, *19*(1), 122–140. https://doi.org/10.1017/S1366728914000832

Walley, A. C., & Flege, J. E. (1999). Effect of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics*, *27*(3), 307–332. https://doi.org/10.1006/jpho.1999.0098

Wen, Z., & 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*(4), 703–728. https://doi.org/10.1017/S0142716418000152

Yilmaz, Y., & Granena, G. (2021). Implicitness and explicitness in cognitive abilities and corrective feedback: A double dissociation? *Studies in Second Language Acquisition*, *43*(3), 523–550. https://doi.org/10.1017/S0272263120000601

Yu, V. Y., & Andruski, J. E. (2010). A cross-language study of perception of lexical stress in English. *Journal of Psycholinguistic Research*, *39*(4), 323–344. https://doi.org/10.1007/s10936-009-9142-2

Zhang, X., Cheng, B., Qin, D., & Zhang, Y. (2021). Is talker variability a critical component of effective phonetic training for nonnative speech? *Journal of Phonetics*, *87*, Article 101071. https://doi.org/10.1016/j.wocn.2021.101071

Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics*, *38*(2), 260–271. https://doi.org/10.1016/j.wocn.2009.11.002