Mandarin speakers' English lexical stress acquisition: a cross-sectional and longitudinal perspective

Katya Petrova

Thesis submitted for the degree of Doctor of Philosophy

UCL Culture, Communication and Media,

Institute of Education

2024

Declaration

I, Katya Petrova, confirm that the work presented in my thesis is my own.
Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Adult L2 speech learning is characterized by large individual variability in the rate and level of ultimate attainment. Growing evidence suggests that individual differences in learners' abilities to process acoustic information may be linked to variability in acquisition success for a broad range of language outcomes. Additionally, many studies have highlighted the crucial role that suprasegmentals play in L2 learner success, but much is still unknown about L2 suprasegmental acquisition. To this end, the current dissertation focused on the cross-sectional and longitudinal investigation of L2 lexical stress acquisition by inexperienced native Mandarin learners of English.

In Experiment 1, I explored the link between individual differences in domain-general auditory processing abilities and learners' lexical stress processing cross-sectionally. Participants were assessed on a battery of auditory processing measures and their lexical stress perception scores and cue weights were also collected. In Experiment 2 I designed and assessed the effectiveness of a novel prosodic training paradigm by adapting the high variability phonetic training (HVPT) approach to the teaching of L2 lexical stress. Mandarin learners engaged in 6 sessions of training and were retested on their lexical stress processing post-training. In Experiment 2 I also asked if individual differences in participants' auditory processing at Time 1 would explain variability in how much they benefited from the prosodic training at Time 2.

Individual differences in auditory processing abilities were linked to lexical stress perception accuracy at Time 1 and were associated with learners' cue weighting strategies. However, auditory processing abilities were not predictive of learning gains from the HVPT training, which was found to be effective for all learners and resulted in significant improvements in lexical stress perception. These findings have important theoretical implications for understanding the factors affecting L2 speech learning success and informing the development of real-world suprasegmental teaching approaches.

Impact Statement

This dissertation explores various aspects of lexical stress acquisition by Mandarin learners of English. It aims to fill a gap in the literature by bringing more research attention to L2 suprasegmental acquisition through 1) exploring the potential factors that might influence the attainment of prosodic accuracy, and through 2) developing an effective training paradigm for teaching non-native prosodic contrasts. This project advances the theoretical understanding of the role of learner factors such as individual differences in auditory processing abilities in determining L2 speech outcomes. Additionally, the project provides new avenues for research into L2 speech training and offers a promising new tool that can complement existing L2 teaching approaches.

Firstly, evidence to date has indicated that individual differences in the way learners process acoustic information may be related to L2 acquisition success. However, there are still open questions about the precise role of auditory processing in the acquisition of specific non-native contrasts. The results reported in this thesis lend empirical support to the view that differences in learners' auditory processing abilities are associated with lexical stress performance (measured in terms of perception accuracy and cue weighting strategies). Nonetheless, auditory processing was not predictive of how much learners benefitted from the phonetic training. The findings reveal a complex relationship between individual differences in learners' domain-general auditory abilities and L2 learning. Taken in combination with previous studies which have reported an effect of auditory processing in language learning (both in naturalistic settings, and modulating learning gains from training), the findings from the current dissertation suggest that the role of auditory processing abilities in language acquisition may depend on other factors, such as the stage and context of learning, as well as the non-native contrasts that are taught. In this sense, the project makes recommendations for future research that can attempt to isolate the contexts and specific instances when training learners' auditory processing can have a positive impact on their language acquisition.

Another important contribution of this dissertation is the development of a short-term prosodic training paradigm for the teaching of L2 lexical stress. Here, I adapted the well-established high variability phonetic training (HVPT) technique used primarily in segmental teaching, to train the lexical stress of native Mandarin learners of English. The training, characterized by high variability input (featuring multiple talkers and phonetic contexts), was highly effective at improving lexical stress perception and it did so for all learners irrespective of their auditory processing aptitude. These positive results bear direct significance for both research theory and language education. On the one hand, the successful application of the HVPT paradigm to the prosodic domain presented in this project, opens up new avenues for research with language learning populations known to experience persistent difficulties when acquiring contrastive stress. On a practical level, the current paradigm can be easily adapted to real-world education contexts, and it can contribute to existing L2 speech interventions that will optimize non-native prosodic acquisition.

Acknowledgements

This dissertation has been an incredible journey—one that has transformed my life in more ways than I could have imagined. Throughout this process, I have grown both professionally and personally, and I am deeply grateful to those who have supported me along the way.

First and foremost, I would like to express my deepest gratitude to my exceptional supervisors, Dr. Kazuya Saito and Dr. Adam Tierney. Their unwavering support, guidance, and invaluable feedback over the years have been instrumental in shaping my growth as a researcher. I am especially grateful for their mentorship and belief in me, which have been a constant source of motivation, helping me persevere through every challenge and milestone.

I also extend my heartfelt thanks to Chaoqun Zheng for translating all testing materials for both experiments featured in this dissertation. Beyond her translation work, I am grateful for her invaluable assistance with participant recruitment in China. Despite the geographic distance, her dedication made the data collection for both experiments possible, and I am truly grateful for her support.

I would also like to express my appreciation to the Leverhulme Trust for sponsoring this research project.

Finally, I am deeply thankful to my parents, family, and friends for their constant belief in me and for always being there to offer words of support and encouragement. Thank you, too, for celebrating every win with me along the way! And above all, I give thanks to God for leading the way.

Table of Contents

Chapter 1 General introduction

1.1 Introduction

One of the challenges most frequently associated with adult second language learning relates to the acquisition of the target language phonological system. Regardless of the language competence attained in other domains, the presence of a perceptible foreign accent is an all-too familiar marker of the non-native speech produced by adult language learners. Normally, speech productions exhibit substantial variation in their degree of foreign accentedness but for the most part learners deviate from the target L2 phonetic realizations in predictable ways shaped by their native language experience.

A wealth of literature exists dealing with the adult acquisition of non-native speech contrasts. However, most of it has centred around segmental categories, and while challenges in this area can adversely affect both perception and pronunciation abilities, sometimes remaining frustratingly resistant to change (for instance, the acquisition of the English /r-l/ contrast by Japanese speakers; Bradlow et al., 1997; Flege et al., 1996; Ingvalson et al., 2012), mastering the phonology of an L2 involves other less widely explored but equally important aspects. One such aspect concerns how the target language encodes prosodic information and the structure of its suprasegmental space.

The significance of successfully acquiring the suprasegmentals of a new language becomes apparent if we examine the evidence accumulated in the area of prosodic production. While complete communication breakdown as a direct result of prosodic inaccuracies is a rare occurrence, stress production accuracy for instance, is essential for L2 speech intelligibility and comprehensibility (Field, 2005). What is more, unskillful manipulation of the acoustic cues associated with lexical stress identity is considered a major contributor to the perception of non-native accent in adults (Zhang et al., 2008). In fact, there is growing evidence suggesting that errors in prosodic realization more so than segmental inaccuracies, can have a negative impact on speakers' perceived accent, comprehensibility ratings and intelligibility judgements (Anderson-Hsieh et al., 1992; Gallego, 1990; Isaacs & Trofimovich,

2012; Kang, 2010; Kang et al., 2010; Moyer, 1999; Munro & Derwing, 1995; Pennington & Richards, 1986). In a study examining the pronunciation attainment of English L2 learners of German, Moyer (1999) showed that successful production performance (that of learners rated closer to the native-like range), was highly correlated with self-reported suprasegmental phonological training. Consistent with this finding, other researchers in the SLA field have shown that more comprehensive pronunciation teaching approaches that integrate suprasegmental components lead to greater improvements on measures of comprehensibility and fluency compared to instructional approaches with a more limited scope (i.e. segmental-only training) (Derwing et al., 1998; Derwing & Rossiter, 2003; Zhang & Yuan, 2020).

Notably, studies carrying out assessments of speech samples obtained under high-stakes conditions, for instance, those elicited as part of language proficiency exams, have emphasized the central role played by suprasegmental features in non-native proficiency and comprehensibility ratings (Choi & Kang, 2023; Kang et al., 2010). In fact, computer modelling using a range of suprasegmental features distilled into four broad groups (prominence, filled pause, speech rate, and intonation) has managed to predict actual oral proficiency scores received in a Cambridge proficiency exam with a high degree of accuracy (Kang & Johnson, 2018). These findings underscore the significance of suprasegmental attainment and invite a reappraisal of the prevailing research priorities in the areas of L2 speech processing and phonological acquisition. A shift in focus from the more narrow, segmental acquisition of non-native categories, to a more holistic teaching approach with equal emphasis on achieving both segmental and prosodic accuracy in the second language is key to promoting communicative success.

Another important aspect of the L2 language learning experience which should receive more attention is the great deal of variability in acquisition outcomes even after experience and input-related factors have been taken into account (Abrahamsson & Hyltenstam, 2008; Birdsong, 2007; Bongaerts et al., 1997). Researchers have proposed that learner-specific factors related to individual differences in perceptual-cognitive abilities may play a role in the rate and degree of L2 attainment (DeKeyser, 2000; Ghaffarvand & Werner, 2019; Granena & Long, 2013; Miyake & Friedman, 1998; Révész, 2012; O'Brien et al., 2007).

Recently researchers have advanced another set of domain-general abilities that may have an effect on L2 learning outcomes – namely, domain-general auditory processing abilities, or the degree of precision with which learners can perceive, encode, and extract information for acoustic sources (Kachlicka et al., 2019; Saito, 2023).

With the need for better understanding of L2 suprasegmental acquisition in mind, the overarching goal of the current dissertation was to investigate the acquisition of English lexical stress by native speakers of Mandarin Chinese. Mandarin is a tonal language using a combination of cues applied at the syllable level to differentiate meaning (Chao, 1968; Duanmu, 2007). English, on the other hand, is an intonation language where changes in meaning are conveyed by contrastive word-level stress (Beckman, 1986; Fry, 1958). Mandarin learners of English have well-documented difficulties with English prosodic acquisition ranging from assigning the correct lexical stress to manipulating its acoustic dimensions in a native-like manner (Archibald, 1997; Chen et al., 2001; Hung, 1993; Juffs, 1990; Zhang et al., 2008; Wang, 2008).

The purpose of the research in the present dissertation was two-fold. Firstly, in Experiment 1 I aimed to explore the auditory processing variables which could have a bearing on Mandarin speakers' acquisition of English lexical stress. Considering that multiple acoustic cues signal lexical stress in English (Beckman, 1986; Bolinger, 1961; Fry, 1955, 1958, 1965), I examined if individual differences in participants' domain-general auditory processing abilities can explain variability in their lexical stress performance (see Kachlicka et al., 2019; Lengeris & Hazan, 2010; Saito et al., 2020a; Saito et al., 2022a; Sun et al., 2021 for research linking individual differences in auditory processing measures and L2 acquisition success).

Next, in a longitudinal study (Experiment 2), I investigated if targeted perceptual training can aid Mandarin speakers' lexical stress acquisition. To this end, I designed and tested a new prosodic training paradigm for teaching participants to identify and discriminate English lexical stress. The paradigm was based on the high variability phonetic training approach (Logan et al., 1991; Lively et al., 1993) which here was adapted to the prosodic context.

The next section of this chapter presents the theoretical framework motivating the two studies. It reviews the main research findings on second language speech acquisition, including individual differences in learner outcomes as well as the potential role of domain-general auditory processing abilities in explaining this variability. Next, I discuss the prosodic structures of English and Mandarin and the potential sources of difficulty for Mandarin learners of English. This is followed by a brief introduction of the HVPT method which served as the basis for developing the training protocol used in this project. Finally, I end the chapter by defining the scope of the dissertation, outlining its main research goals and presenting the key measures collected for both experiments.

1.2. Literature review

1.2.1 The adult second language experience

Adult learners' abilities to discriminate and produce new sounds in their L2 vary substantially and they are largely constrained by the existing phonetic representations acquired through native language experience (Archibald, 1998; Best, 1995; Flege, 1995a). Evidence indicates that infants are initially born with the innate ability to discriminate phonetic contrasts even when they are not encountered in their linguistic environment (Werker et al., 1981). Within the first year of life, however, this window of heightened perceptual sensitivity sees a gradual decline (see Polka & Werker, 1994 for a developmental trajectory of non-native vowel perception; and Werker & Tees, 1984 for non-native consonant perception), with perceptual abilities becoming increasingly tuned to discriminating the phonetic categories experienced in the native language environment (Kuhl et al., 1992; Werker, 1989). This perceptual reorganization, sometimes termed perceptual warping (Kuhl, 2000), leads to the establishment of language-specific speech perception strategies tailored to searching for and processing the acoustic dimensions most relevant for disambiguating L1 phonetic categories. At the same time, more marginally informative dimensions are underattended (Iverson et al., 2003). This frequently becomes a source of difficulty when processing non-native phonemic contrasts, especially when acoustic cues carry different informational loads in the L2.

There is ample evidence documenting the challenges adult learners experience with the acquisition of new speech categories at the segmental level for both vowel (Best et al., 2003; Flege et al., 1997; Flege & MacKay, 2004; Fox et al., 1995; Munro et al., 1996) and consonant contrasts (Best & Strange, 1992; Flege et al., 1995c; Flege & Hillenbrand, 1986; Flege et al., 1996). A well-studied case of persistent difficulties with the perception and production of L2 contrasts, is the acquisition of the English /r/ and /l/ phonetic categories by native Japanese speakers (Goto, 1971; Iverson et al., 2003; Miyawaki et al., 1975; Saito, 2013; Strange & Dittmann, 1984). The English /r/-/l/ contrast is perceptually assimilated into a single Japanese tap category making the discrimination between the contrast especially challenging (Best & Strange, 1992; Guion et al., 2000a). Research has shown that the English /r-l/ contrast is conveyed by changes in the F3 onset frequency, the transition duration of the first formant, F1, and the F2 onset frequency (Hattori & Iverson, 2009; Strange, 1988), with F3 considered the most diagnostic cue for native English speakers (Epsy-Wilson, 1992; Miyawaki et al., 1975). However, due to the perceptual similarity between the English /r-l/ contrast and the Japanese tap category, native Japanese learners of English have been consistently documented to up-weight reliance on F2 and down-weight F3 both in their perceptions and productions (Iverson et al., 2003; Iverson et al., 2005), likely because of interference from their native language where the excessive variability of F3 makes it an unreliable correlate of the Japanese flap (Miyawaki et al., 1975). Consequently, Japanese learners frequently fail to perform near native English levels even after extensive natural immersion (Aoyama et al., 2004; Gordon et al., 2001), and evidence has suggested their cue weighting strategies in relation to the third formant, F3, do not change as a function of length of residence or age of arrival in the immersion environment (Ingvalson et al., 2011).

The above examples illustrate the significant challenges that L2 learners face when acquiring new perceptual patterns later in life. These stem from the fact that the speech signal is highly redundant and linguistic contrasts are conveyed by multiple acoustic dimensions (Winter, 2014) which don't carry the same attentional weights (Fear et al., 1995; Lisker, 1986) with some dimensions signalling category membership more reliably than others (Holt

& Lotto, 2006). Bearing these considerations in mind, the most immediate priority for learners when encountering the phonological system of a new language is perceptual retuning, i.e., learning how to exploit acoustic cues to the degree and in the manner that they are utilized by native speakers. This could involve two distinct but by no means trivial processes. Depending on the perceptual strategies established in response to native language experience, non-native listeners might be tasked with learning to direct their attention to changes in acoustic dimensions they might be accustomed to disregarding (Francis & Nusbaum, 2002). Alternatively, the acquisition process might require them to ignore changes along a dimension that is perceptually salient and heavily attended to in their native language but might not be effective at signalling a category in the target language (as is the case with F2 rather than F3 for Japanese learners acquiring the English /r-l/ contrast; Ingvalson et al., 2012).

When learning a new language, however, we must contend not only with the acquisition of novel sounds at the segmental level, but also with unfamiliar intonational patterns, at the suprasegmental level. Indeed, adult L2 learners have been shown to experience challenges with the acquisition of non-native L2 suprasegmental accuracy in the instances when the native and target languages have different prosodic systems or employ a different set of acoustic cues to signal prosodic contrasts (Guion et al., 2000b; Pittam & Ingram, 1992; Scuffil, 1982). The literature has documented issues concerning various aspects of L2 suprasegmental acquisition covering both perception and production abilities. For instance, the acquisition of tonal properties in target L2 tonal languages has been shown to largely depend on learners' experience with pitch use in their native language (Francis et al., 2008; Hallé et al., 2004; Leather, 1987; Wang et al., 1999; Wayland & Guion, 2004), with native tonal language experience having a facilitatory effect, compared to experience with languages where pitch does not have a phonemic role.

Speech acquisition models

Several theoretical accounts have been put forth to explain some of the difficulties associated with the acquisition of non-native speech contrasts by adult learners.

The Speech Learning Model (SLM) proposed by Flege (1995a) with its more recent update, the revised SLM (SLM-r) (Flege & Bohn, 2021), link success in L2 speech perception to the degree of similarity between the native and target language phonological systems. According to this model, non-native segmental categories that are too similar to existing native categories are viewed as sources of interference. By contrast, adult learners can acquire new sounds when the degree of similarity between the linguistic categories in the two languages is small (i.e., the likelihood of L1 perceptual interference is regarded as minimal). The SLM-r maintains that the capacity to learn novel categories remains intact through our lifetime and it further recognizes individual differences in auditory processing and cognitive abilities as contributing factors to the large variability observed in L2 speech acquisition.

Best's Perceptual Assimilation model (Best, 1995; Best et al., 2001; Best & Tyler, 2007; Tyler, 2019), on the other hand, is grounded in theories of articulatory phonology and the model predicts improved perception of a non-native category when its articulatory properties are sufficiently similar to (and can be assimilated into), an existing native category.

Lastly, the Native Language Magnet model proposed by Kuhl (Iverson & Kuhl, 1994; Kuhl, 1991, 2000), is based on infant first language acquisition and it follows a developmental trajectory of perceptual abilities going from language-universal to language-specific as a function of exposure to the native language. According to this model, native language experience acts as a filter and can affect perceptual performance when we encounter new languages as adults (Kuhl et al., 1992).

For the most part, these models have sought to explain the underlying constraints on second language speech acquisition documented in instances of vowel and consonant learning. As such, they were developed with segmental phonology in mind. However, referring to these models can also be of theoretical benefit when formulating predictions about L2 prosodic acquisition as some evidence of similarities between L2 segmental and suprasegmental acquisition have been proposed, with segmental and suprasegmental learning both happening gradually over extended periods of time, requiring considerable L2 experience, and most importantly, with performance varying depending on the specific

contrasts acquired (Trofimovich & Baker, 2006). For instance, studies have advanced the theory that native language background and more specifically prosodic typology, can exercise a negative impact on the acquisition of L2 prosodic attributes that are dissimilar to those present in the native language (Archibald, 1992, 1993; Nguyen & Ingram, 2005; Shen, 1990).

More recently, models related to how attention is allocated among different acoustic dimensions making up the speech signal have highlighted the dynamic nature of perceptual cue weighting, with listeners keeping track of the informativeness of acoustic dimensions and adjusting attentional weights based on prior linguistic experience (Francis & Nusbaum, 2002; Gordon et al., 1993; Holt et al., 2018). By this view, acoustic dimensions which have held high informational load for learners in the past (for instance due to their native language experience), become more perceptually salient (i.e., they have a greater tendency to capture auditory attention), and consequently also receive higher perceptual weightings. These mechanisms shaping perception, however, may become problematic when unfamiliar patterns of cue weightings are encountered for new phonetic contrasts, for example when learning a second language. Then the successful acquisition of the new language phonology would sometimes require the overriding of existing cue weighting strategies (Ingvalson et al., 2012), and in many cases learning to shift auditory attention away from highly salient cues to dimensions that are important for reliably signalling the new phonetic contrast (Francis & Nusbaum, 2002).

An ideal case in point to illustrate this model are the perceptual strategies of native Mandarin speakers. Mandarin Chinese is a tonal language where pitch is a crucial cue for determining the meaning of words (Gandour, 1978; Howie, 1976). Similarly, native Mandarin speakers have been shown to rely predominantly on pitch (and doing so significantly more than native English speakers do) when processing English prosody (Archibald, 1997; Chen et al., 2001; Hung, 1993; Juffs, 1990; Zhang et al., 2008). In an experiment designed to test Mandarin speakers' domain-general cue weighting strategies, Jasmin et al. (2021) assessed participants' cue reliance for both speech and non-speech stimuli. They found that Mandarin speakers recently arrived in the UK, weighted pitch significantly more when categorizing both phrase boundaries and musical sequences. Highly

pertinent for the attentional theories of speech perception, the authors also tasked participants with categorizing English speech stimuli by attending to one of two dimensions (pitch or amplitude) and simultaneously ignoring irrelevant changes along the other dimension. The study showed Mandarin listeners were unable to direct their attention to amplitude and ignore pitch changes when explicitly required by the task. In a follow-up experiment with Mandarin long-term residents living in the UK, this pitch-biased categorization behaviour was shown to remain unchanged even after extensive immersion experience, with Mandarin speakers still unable to pull their attention away from pitch and direct it towards amplitude (Petrova et al., 2023). Significantly for L2 acquisition, however, the study showed that the cue weighting strategies of these more experienced Mandarin learners in the case of categorizing English prosody (phrase boundaries) had undergone a shift towards greater reliance on durational information, in effect showing evidence of developing more nativelike cue weighting strategies. These findings support dimensional salience theories by demonstrating how extensive experience with a tonal language shapes the attentional salience of acoustic dimensions which can have an impact on second language speech processing. It also seems that, at least in certain cases, subsequent extensive experience with a second language can lead to the development of more native-like cue weighting strategies in that target language even if auditory attention to highly salient cues remains unchanged (Petrova et al., 2023).

1.2.2 Prosody and lexical stress

Prosody can be defined as the vocal modulations of speech typically associated with suprasegmental phonology (Wennerstrom, 2001). It is broadly subdivided into affective prosody, relating to the expression of emotional states, and linguistic prosody (Monrad-Krohn, 1947), associated with suprasegmental features such as lexical stress (e.g. CONtract vs conTRACT; Gay, 1978), and contrastive sentence focus (it was HER vs it WAS her; Ladd & Morton, 1997), among others. Prosodic information is commonly conveyed by changes in pitch, duration, intensity and spectral quality (Fear et al., 1995; Fonagy, 1978; Jasmin et al., 2023), and it is realized over larger stretches of the speech signal extending beyond segmental phonology, to include syllables, words and phrases (Crystal, 2009; Lehiste, 1970).

Lexical stress, for instance, relates to syllable prominence within a word with different syllables receiving different degrees of prominence (Kager, 2007). As such lexical stress is considered a relative property judged based on the perception of emphasis on individual syllables (Liberman, 1975), often instantiated by the opposition of weak versus strong syllables within the same phonological word (Fear et al., 1995).

Prosodic information in general, and lexical stress, in particular, play a fundamental role in speech recognition. For instance, English is characterized by the existence of robustly uneven distributional regularities of stress patterns. In a corpus study of the English language Cutler & Carter (1987) found that the number of lexical words with a trochaic stress pattern (first-syllable stressed) was three times as large as that of words beginning with weak syllables. In practical terms, about 90% of content words begin with a stressed syllable. Unsurprisingly, studies have shown that native English speakers are acutely aware of this distributional regularity and show a strong preference for interpreting stress patterns as strong-weak (Cutler & Butterfield, 1992) whereby the detection of a strong syllable in the continuous speech signal triggers segmentation (Cutler & Norris, 1988), pointing to a stress-based mechanism for lexical access and speech segmentation (Mattys & Samuel, 1997, 2000). Adult learners of a second language, therefore, must also be aware of the lexical stress regularities inherent in their target L2 in order to develop appropriate segmentation strategies that will allow them to successfully parse L2 speech input into meaningful units.

However, languages differ with respect to the employment of prosodic features such as lexical stress patterns, and they also differ in the number and weighting of acoustic cues serving to disambiguate stress contrasts (Gordon & Roettger, 2017; Grzegorz & Briony, 1999; Morton & Jassem, 1965). As a result, prior experience with the native prosodic system can affect L2 acquisition outcomes either facilitating (Wayland & Guion, 2004) or interfering (Wang et al., 1999) with the target L2 prosodic feature.

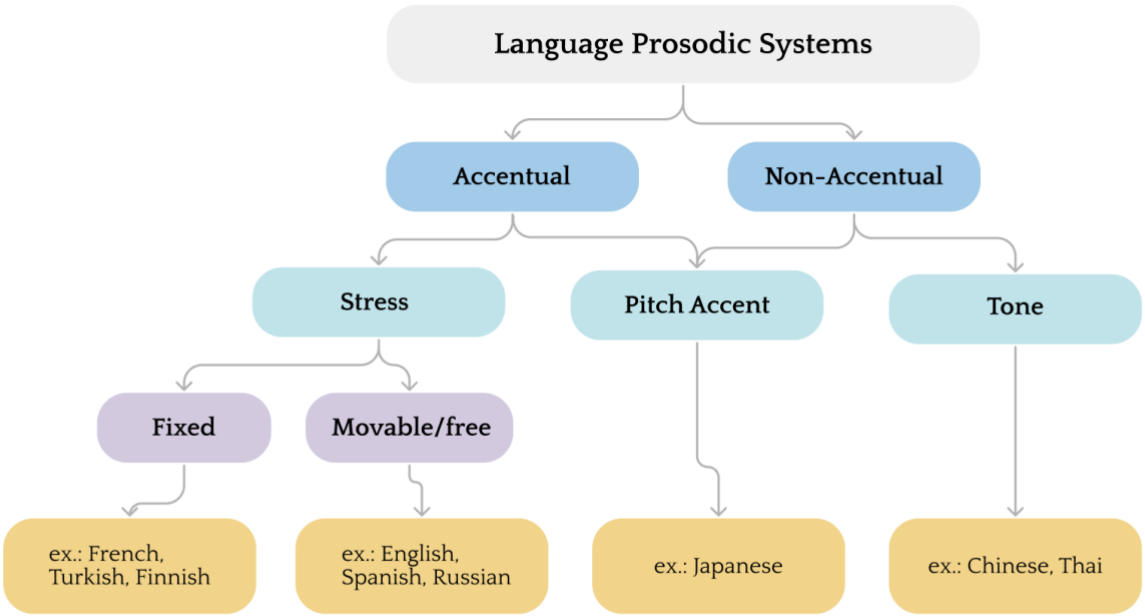
L2 lexical stress perception

The world's languages can be broadly categorized into accentual or non-accentual (tone) languages (see Figure 1 adapted from Archibald, 1997). Accentual languages utilize pitch (among other perceptual dimensions) to mark stress accent, while tonal languages use pitch

phonemically. There is a third type of prosodic systems, pitch-accent, which crosses over into both accentual and non-accentual types. Stress languages also differ according to how flexible they are with respect to stress placement. For instance, in English and Russian, minimal pairs of words exist that differ only in their stress placement. In these "free" stress languages, the assignment of stress is flexible and it can fall on different syllables within the word. In contrast, positionally fixed-stress languages (such as French, Turkish, Finnish) are governed by formal rules for stress placement, whereby word stress always falls on the same syllable (but see Peperkamp, 2004, for some exceptions).

Figure 1

Prosodic systems categorization (adapted from Archibald, 1997)



One source of difficulty with the perception of L2 suprasegmentals can be traced back to native language experience and the presence, absence, or specific acoustic expression of prosodic structures in the L1. For instance, numerous studies investigating the stress perception abilities of speakers with non-contrastive stress systems have documented

systematic difficulties in the acquisition of L2 stress (Dupoux et al., 1997, 2001; Lukyanchenko, et al., 2011; Peperkamp & Dupoux, 2008), warranting the term "stress deafness" (Dupoux et al., 1997). The Lexical Parameter Setting hypothesis has been proposed to account for these findings (Dupoux et al., 2008; Peperkamp & Dupoux, 2008; Peperkamp, 2004). It views learners' persistent difficulties with the acquisition of L2 lexical stress as a processing challenge stemming from the presence or absence of a *stress parameter* (a mechanism responsible for the encoding of stress). According to this hypothesis, depending on the stress regularity existing in their native language, infants either activate a "stress parameter" when this is needed, or they fail to do so if their L1 does not use stress in a contrastive way.

Cross-language transfer at the suprasegmental level is also observed between languages having different prosodic systems. For instance, some languages do not utilize word-level stress in either a strictly phonological sense, or to convey differences in word meanings. They belong to two broad categories: tone languages (such as Mandarin Chinese where tones are applied at the level of the syllable to distinguish lexical items; Chao, 1968), and pitch-accent languages (e.g., Japanese, classified as an accentual tone language (Grice, 2001)). For native English adult learners of tonal languages, for instance, the acquisition of lexical tones is particularly challenging (Wang et al., 1999) largely due to differences in the use and function of pitch across the two languages (Chen, 1974; White, 1981). While pitch is crucial for distinguishing meaning in Mandarin (Chao, 1968; Duanmu, 2007), it plays a less prominent role in English by conveying prosodic structures such as lexical stress and phrase boundaries (Fear et al., 1995; Mattys, 2000; Streeter, 1978). The converse is also correct, with native Mandarin learners of English showing evidence of interference from their native tonal system in the acquisition of English suprasegmentals (Archibald, 1997; Chen et al., 2001; Hung, 1993; Juffs, 1990).

English lexical stress

English is a stress language which means that syllables differ in their prominence within a word. This prosodic feature while central to comprehension is lexically contrastive only in a limited number of minimal word pairs denoting lexical category (i.e., SUBject vs subJECT). English lexical stress is acoustically correlated with pitch (fundamental frequency, F0),

duration, intensity, and vowel (or spectral) quality (Beckman, 1986; Bolinger, 1961; Fry, 1955, 1958, 1965; Lieberman, 1960; Sluijter et al., 1997; Sluijter & van Heuven, 1996). Stressed syllables tend to be characterized by higher pitch, greater intensity, longer duration, and full, unreduced vowels (Goffman & Malin, 1999). In contrast, unstressed syllables are normally instantiated by lower pitch, weaker intensity, shorter length, and the presence of vowel reduction.

The dimension most frequently argued to carry the most perceptual weight in signalling lexical stress is pitch (Beckman, 1986; Bolinger, 1958; Fry, 1958; Lieberman, 1960; Morton & Jassem, 1965). In his 1958 study, Fry examined listeners' stress perception in a task involving modified natural recordings of the minimal stress pair /SUBject – subJECT/, in which the fundamental frequency was manipulated in conjunction with duration. The experiment revealed that F0 differences between the first and second syllable (largely irrespective of the size of the frequency change), affected stress judgements at each level of the duration cue manipulation. However, a closer examination of the available literature reveals a lack of a clear consensus on the relative ranking of the four acoustic cues in lexical stress perception (Beckman, 1986; Bolinger, 1958; Fry, 1958; , Sluijter & van Heuven, 1996). One possible source of difficulty with teasing apart the relative contributions of the different acoustic dimensions arises from the existence of complex relationships between these cues at both the segmental and suprasegmental levels.

At the suprasegmental level, in any given communicative context (Eady & Cooper, 1986), the fundamental frequency makeup of a syllable, can be affected by phrase- or sentential-level prominence superimposed on the word-level prominence. For instance, a specific word can receive additional accentual prominence within the larger context of the sentence (in effect confounding accent and stress, Sluijter & van Heuven, 1996). Given this context sensitivity and the ensuing difficulty in measuring the relative contribution of pitch to lexical stress perception, some scholars have suggested that pitch is in fact the least useful cue for supporting reliable stress judgements (Beckman & Edwards, 1994; Sluijter & van Heuven, 1996).

Other authors have highlighted the significant role of syllable duration in marking word stress distinctions (Adams & Munro, 1978; Okobi, 2006). Taylor (1981), for instance, argued that duration was second in importance only to pitch, which seems to be supported by the existence of a natural speech regularity in English. The analysis of a large vocabulary database reveals that longer vowels and vowel diphthongs tend to be stressed in English in about 60% of the cases, and shorter vowels are stressed only 35% of the times (Guion et al., 2003). However, here too coarticulation factors are important to consider given that durational information tends to vary as a function of adjacent consonants, and vowel types (Chen, 1970; House, 1961; Peterson & Lehiste, 1960). Another cue signalling lexical stress that has invited less controversy, is intensity. Syllable loudness is generally recognized as a less reliable cue when judging stressed versus unstressed syllables and consequently it carries less perceptual strength in comparison to pitch and duration (Mattys, 2000; Morton & Jassem, 1965; van Heuven & Menert, 1996).

Last but not least, a particularly useful cue to English lexical stress perception, is vowel quality (Beckman, 1986; Beckman & Edwards, 1994; Cutler, 1986; Fear et al., 1995; Fry, 1965). In essence, vowel quality refers to the formant structure (F1 and F2 centralization) of stressed and unstressed vowels in a word. Reduced vowels found in unstressed syllables tend to undergo F1-F2 centralization compared to vowels in stressed positions (Gay, 1978; Rosner & Pickering, 1994). This acoustic parameter and its place in the cue weighting hierarchy, have been investigated extensively. A study examining the relative importance of the four acoustic dimensions to stress in both speaker perception and production ranked the spectral characteristics of vowels as secondary only to F0 (Howell, 1993). Similarly, in an experiment with synthesized non-words Rietveld & Koopmans-van Beinum (1987) manipulated the vowel spectral characteristics of tri-syllabic words, while variability along the F0, duration and intensity dimensions was held constant across all syllables. Their findings revealed that the lack of vowel reduction alone was sufficient to shift perception with respect to stress placement, and all other cues held constant, listeners tended to classify non-spectrally reduced syllables as stressed syllables. Even research into L2 speech acquisition has highlighted the importance of using vowel reduction in a systematic manner to reduce the perception of foreign accent (Flege & Bohn, 1989; Fokes et al., 1984; Hammond, 1986).

Mandarin tones

In Mandarin Chinese changes in lexical tone realized at the level of individual syllables are used to convey differences in meaning (Chao, 1968; Duanmu, 2007). In terms of prosodic expression, Mandarin is markedly different to intonation languages. While no segmental changes are involved, a tonal change can result in a change of meaning. As a starting point, Mandarin Chinese has 4 full lexical tones (tones that convey lexical distinctions): high level (tone1), high rising (tone 2), falling-rising (tone 3 - also known as low dipping), and high falling (tone 4) (Chao, 1968, Eady, 1982; Gandour, 1978; Howie, 1976).

The acoustic correlates to lexical tones in Mandarin Chinese are also used to signal word stress in English. However, as it will become apparent below, these sound cues are not implemented in the same way. Tones in Mandarin are instantiated by changes in pitch, duration and intensity (amplitude contour) (Howie, 1976; Liu & Samuel, 2004; Whalen & Xu, 1992). Of these, fundamental frequency (or more specifically, the direction of the F0 contour during the vowel) carries the greatest dimensional weighting (Gandour, 1978; Howie, 1976). And while duration and intensity come secondary in terms of their informativeness, changes in these two dimensions contribute consistently to tone perception, as evidenced by studies in performance on signal-processed stimuli in which F0 information was fully or partially unavailable. While performance was adversely affected when F0 information was neutralized, these studies have shown that Mandarin speakers can make use of durational and amplitude information to aid in their categorizations when F0 information is not available (Fu et al., 1998; Liu & Samuel, 2004; Whalen & Xu, 1992).

Mandarin Chinese learners and English prosodic acquisition

The literature has documented that adult Mandarin learners of English exhibit a range of difficulties processing English prosodic features, and one possible reason for this consistent finding could be interference from the Mandarin tonal system (Archibald, 1997; Chen et al., 2001; Hung, 1993; Juffs, 1990; Ou, 2016). Juffs (1990) found that Mandarin learners with primarily classroom-based experience showed a lack of knowledge about where stress should be located in individual words. In the speech samples where the correct stress

pattern was produced, durational information was used in a non-native manner, with participants sometimes producing syllables that were much longer than appropriate. Additionally, pitch manipulation seemed to be problematic for some Mandarin speakers who used pitch movement as opposed to changes in pitch height to mark stressed syllables. This strategy sometimes spilled into indiscriminate nonstandard stressing of even monosyllabic function words. The result was a perceptually syllable-timed effect which is highly uncharacteristic of native English productions (Adams & Munro, 1978). This finding is consistent with previous research demonstrating that when judging tone dissimilarity, most tonal language speakers (Mandarin, Taiwanese) tend to rely more heavily on pitch contour than on pitch height (Gandour, 1983), probably because in Mandarin Chinese pitch contour (the direction of pitch movement within a syllable) is more important than pitch height (Pike, 1948). Analogous findings related to excessive manipulations of F0 contours in Mandarin learners' English lexical stress processing will be a conspicuous feature in multiple other studies reviewed here.

When cue weighting strategies for categorizing lexical stress are concerned, evidence indicates that native Mandarin speakers place overwhelming reliance on pitch information when making stress judgements. When non-words were manipulated along the F0, duration, and intensity dimensions, Wang (2008) found that while native English listeners made use of all three cues in their lexical stress judgements, F0 was the dominant dimension for the native Mandarin group who showed minimal use of durational and intensity information. In this respect, Mandarin speakers displayed markedly different perceptual cue weighting strategies compared to native English speakers.

In a seminal study, Archibald (1997) examined the cross-linguistic assignment of English lexical stress by native Japanese and native Chinese speakers. Participants' perception and production abilities were tested 5 months apart and Mandarin learners showed no evidence of paying attention to the stress pattern regularities normally utilized by native English speakers in their stress processing. Instead, they seemed to rely on a purely lexical strategy to stress acquisition, treating stress as a phonemic property that is stored on an item-by-item basis. This strategy also went partially towards accounting for another finding of the study, namely the observation that for some participants stress perception scores actually

deteriorated from time 1 to time 2 of testing. This perplexing finding could reflect a continued reliance on poorly perceived stress patterns stored as part of individual lexical entries which the author argued could in fact lead to perpetuating learners' suboptimal performance. Confirmatory evidence for the application of a similar acquisition strategy by native tonal speakers was later reported in Wayland et al. (2006) in the case of adult Thai learners of English who used their knowledge of phonologically similar words to assign stress to nonwords implying a reliance on word-by-word memorization. The authors took this as an extension of the native tonal acquisition approach where tonal properties are acquired for individual words and hypothesised that this approach was making it less likely for Thai learners to use information about syllabic structure and syllable class to develop awareness about English stress regularities.

When perceiving lexical stress, native Mandarin speakers with around 2 years of immersion experience in the United States have been shown to perform similarly to English listeners (Yu & Andruski, 2010). In their study, Yu & Andruski (2010) examined lexical stress perception for three types of bisyllabic test stimuli: trochaic and iambic pairs of real words, pseudo words and hums. Mandarin listeners showed similar performance to that of the native English group when tested on real words (despite showing longer response latencies). However, the two groups' response behaviour differed, with English speakers showing higher accuracy when categorizing stimuli with trochaic stress patterns, while the Mandarin group exhibited the reverse pattern. Additionally, acoustic measurements of the test stimuli revealed that Mandarin listeners, unlike native English subjects, showed consistent bias towards pitch reliance in their judgements regardless of stimulus type or stress pattern. Also, unlike native English listeners, the Mandarin group made no use of vowel quality information in any of the stress contexts tested.

A more recent study failed to corroborate this last finding instead suggesting that Mandarin learners with short-term immersion experience (just over a year) can successfully make use of all four acoustic dimensions signalling word stress including vowel quality which emerged as the strongest cue in Mandarin listeners' identification responses on par with native English speakers' performance (Chrabaszcz et al. 2014). However, it is worth noting that the testing stimuli in this experiment included modified recordings of the iambic and trochaic

versions of one single nonword ("maba") which may have influenced Mandarin speakers' performance. One possible explanation of the study's findings proposed by the authors is that Mandarin listeners may have treated the stressed /ɑ/ and the unstressed /ə/ sounds in this isolated stimuli set as a segmental difference rather than a vowel structure change used to mark stress distinction.

Sentence stress production has also emerged as an area of difficulty for Mandarin speakers. They have been shown to produce stressed and unstressed words with higher F0, inappropriate duration and greater intensity than native speakers (Chen et al., 2001). The authors attributed this finding to L1 interference. Examining the continuous speech of Mandarin Chinese and American English speakers for instance, Eady (1982) found the two groups displayed systematic differences in the patterns of F0 movements when reading declarative sentences. Compared to the average rate of F0 change in the speech of the American group, Mandarin speakers produced their samples with greater F0 fluctuations expressed as a greater amount of dynamic movement. These findings appear to be a direct reflection of F0 movement patterns characteristic of tonal speech (Gandour, 1983), and are not restricted to the English learning context only but have been reported for Chinese learners of French, too (Shen, 1990; Sneppe & Wei, 1984).

Similar findings were reported by Zhang et al (2008). In a production experiment, native speakers of Mandarin successfully utilized F0, duration and intensity to produce minimal stress pairs (e.g. desert, subject). Moreover, consistent with native productions, they used higher F0, longer length and greater intensity to mark stressed syllables. But here too, many of the problems noted by Chen et al. (2001) were also present. Mandarin speakers used significantly higher F0 than that of native English speakers. Their samples tended to either not contain vowel reductions in unstressed positions or the spectral characteristics of the vowels did not approach native-like levels. This improper control over the acoustic correlates of stress gave rise to a noticeable accent reflected in significantly lower acceptability ratings given to Mandarin productions compared to those of the native English participants.

See Table 1 below for a summary of the key studies presented in this section.

Table 1

Summary of studies investigating lexical stress processing in Mandarin speakers

STUDY	PARTICIPANTS	TESTED ON	FINDINGS
Archibald, 1997	1 Cantonese native speaker 2 Mandarin native speakers 1 Japanese native speaker	Lexical stress perception and production	*Chinese speakers showed evidence of storing lexical stress on an item-by-item basis and did not use information about syllable structure or English lexical stress regularities; * Lexical stress perception became worse from Time 1 to Time 2 for some participants
Chen, et al., 2001	40 native Mandarin speakers (2+ years immersion experience) 40 native English speakers	Sentence stress production	*Stressed words: Mandarin speakers produced stressed words with significantly higher F0 and shorter duration than native English speakers *unstressed words: Mandarin speakers produced unstressed words with significantly higher F0 and greater intensity than native English speakers
Chrabaszcz et al. 2014	15 native English speakers (M = 1.3 years immersion experience) 15 native Russian speakers 15 native Mandarin speakers	Lexical stress identification using synthesised recordings to assess cue weighting	* All four cues to lexical stress (pitch, intensity, duration, and vowel quality) were significant predictors of lexical stress categorization behaviour for both native Mandain and native English speakers

			* Pitch and vowel quality were the most strongly weighted cues for both the Mandarin and English groups
Juffs, 1990	19 native Mandarin speakers (undergraduate students in China, no immersion experience)	Lexical stress production	* errors in stress placement * some speakers used pitch movement instead of pitch height to mark word stress * some speakers produced word stress by lengthening the syllable too much
Lin et al., 2014	17 Mandarin speakers (1 year and 6 months immersion experience) 18 Korean speakers 19 English speakers	Lexical stress perception	* The native Mandarin and native English groups had superior lexical stress processing performance compated to Korean speakers * changes in vowel quality did not aid Mandarin speakers in their word recognition performance
Wang, 2008	62 native Mandarin speakers (university students in China with no immersion experience) 38 native English speakers	Lexical stress perception	* Native Mandarin listeners weighted duration and intensity information significantly lower in categorizing English lexical stress than did native English speakers * The Mandarin group also placed significantly more weight on F0 in comparison to native English speakers

Yu & Andruski, 2010	30 native Mandarin speakers (< 2 years immersion experience) 30 native English speakers	Lexical stress perception	* Mandarin speakers used pitch to recognize stress in each stimulus type and stress pattern; * Mandarin listeners did not rely on vowel quality for making their identifications
Zhang et al., 2008	10 native English speakers 10 native Mandarin speakers (3-4 years immersion experience)	Lexical stress production	* F0 - Mandarin speakers produced stressed syllables with significantly higher F0 than the English group * F0 peak: Mandarin speakers produced the F0 peak location earlier in unstressed syllables compared to stressed syllables * vowel reduction: Mandarin speakers either did not reduce vowels in unstressed syllables, or they reduced them incorrectly

1.2.3 Individual differences in second language acquisition

A range of factors have consistently been associated with successful L2 speech attainment including age of acquisition (Flege et al., 1995; Flege et al., 1999; Munro et al., 1996; Muñoz, 2014; Saito, 2013), quantity and quality of the target language input (Flege, 1995a; Derwing & Munro, 2013), amount of foreign language input and practice (Larson-Hall, 2008; Muñoz, 2006; 2014; Zhang & Lu, 2013), length of immersion (Derwing & Munro, 2013; Saito & Brajot, 2013; Trofimovich & Baker, 2006), and the amount of interaction with native speakers of the target language (Flege & Liu, 2001), among others. However, the rate and degree of success among second language learners is subject to much individual variability (Díaz et al., 2012) with some learners even reaching near-native proficiency in their L2 (Abrahamsson & Hyltenstam, 2008; Birdsong, 2007; Bongaerts et al., 1997). When exploring factors that can account for the broad range of L2 outcomes, researchers have also proposed the role of aptitude-based learning abilities (see Doughty, 2018; Li, 2016 for an overview).

Language learning aptitude is conceptualized as a set of perceptual cognitive abilities that are predictive of the rate of language learning success in mostly instructed classroom settings (but see Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000; Granena & Long, 2013, for the role of language learning aptitude in immersive L2 acquisition). The concept of aptitude was initially operationalized as a set of four basic components (Carroll, 1981), namely, phonemic coding ability, grammatical sensitivity, rote learning and inductive learning abilities. However, the focus of research has gradually shifted from an emphasis on deliberate, analytical abilities to more implicit, cognitive-related factors including inhibitory control (Ghaffarvand & Werner, 2019; Mercier et al., 2014), attention (Guion & Pederson, 2007; Safronova & Mora, 2013), working memory (Miyake & Friedman, 1998; Révész, 2012; Service, 1992), and phonological short-term memory (Darcy et al., 2015; Lee & Révész, 2021; O'Brien et al., 2007).

Another important factor underlying individual differences among L2 learners which has received attention in recent years relates to learners' abilities to process the spectral and

temporal characteristics of sounds, hereafter referred to as domain-general auditory processing abilities. Much research has already been devoted to individual differences in auditory processing abilities and their role in first language acquisition in both children and adult populations.

Domain-general auditory processing and first language acquisition

Speech sounds are redundantly encoded by multiple and often overlapping acoustic dimensions (Winter, 2014), with some more informative of category membership than others and consequently receiving greater relative perceptual weight (Holt & Lotto, 2006). An example of cue redundancy is the case of voicing (e.g., *rapid* vs *rabid*) where the distinction between voiced and voiceless consonants is conveyed by over a dozen different acoustic cues (Lisker, 1986), of which voice onset time (VOT) and fundamental frequency are the most diagnostic. However, the presence of redundancy is not confined only to the segmental level. Prosodic features such as lexical stress (Fear et al., 1995; Mattys, 2000), sentence focus (Breen et al., 2010), and phrase boundaries (Streeter, 1978) are similarly conveyed by multiple input dimensions. An additional layer of complexity when processing speech is introduced by the fact that the speech signal is transient and rapidly changing, necessitating both precise and efficient processing of its acoustic components (Holt & Lotto, 2008). Returning to the example discussed above, a subtle temporal change on the order of just 10 ms shifts a voiced /b/ to a voiceless /p/ (Stevens, 1998). Thus, having the ability to detect subtle changes in the temporal and spectral characteristics of sounds becomes a fundamental prerequisite for successful language development (Kuhl, 2004).

However, even listeners with normal hearing exhibit individual differences in their auditory processing abilities (Kewley-Port, 2001; Kidd et al., 2007). Theoretically, variability in individuals' abilities to detect subtle differences in acoustic information could carry implications for both language development in children and speech processing in adults. Empirical evidence seems to support this showing that individual differences in perceptual sensitivity including psychoacoustic discrimination abilities and music memory explain some of the variance in children's receptive and production language outcomes (Anvari et al.,

2002; Bavin et al., 2010; Douglas & Willatts, 1994; Lamb & Gregory, 1993; Tierney et al., 2021).

Longitudinal investigations have reported similar results. An electrophysiological study capturing infants' cortical responses to complex tones has highlighted the predictive power of early auditory processing abilities in subsequent linguistic development (Choudhury & Benasich, 2011). Similarly, recent research by Kalashnikova et al. (2019) found that children with better discrimination acuity for amplitude risetime measured in infancy, showed larger vocabulary sizes when tested again at 3 years of age. The cited sources support a positive relationship between precise auditory temporal and spectral resolution and superior language performance outcomes in children.

Studies conducted with atypical populations with neurodevelopmental language disorders have also reported a variety of auditory sensory difficulties including temporal and spectral processing deficits in children with specific language impairment (SLI) (McArthur & Bishop, 2005; Wright et al., 1997) and developmental reading disabilities (Ahissar et al., 2000; Casini et al., 2018; Goswami et al., 2002; Montgomery et al., 2005; Rosen & Manganari, 2001; Wright & Conlon, 2009). However, while these studies have reported a correlational link between auditory processing abilities and language outcomes, there is no conclusive evidence establishing the existence of causality between the two.

There is, in fact, evidence suggesting that the mere existence of a perceptual deficit does not necessarily imply corresponding negative consequences for language. For instance, adults with a congenital perceptual deficit for processing pitch (amusia, Vuvan et al., 2015), report no everyday language comprehension difficulties despite displaying impaired intonational processing in behavioural testing (Liu et al., 2010). This suggests that while individuals' fine-grained auditory processing might be compromised, they need not necessarily experience dramatic effects in their daily communication. This is further supported by a cue weighting study testing amusics on a prosodic contrast (where pitch was the maximally informative cue), which found that amusic listeners relied more heavily on durational information when classifying stimuli compared to controls (Jasmin et al., 2020). The authors concluded that individuals impaired along a specific dimension could adapt

their perceptual strategies to better fit the constraints of their auditory system. But while having poor auditory processing abilities does not necessarily lead to poor language outcomes, the cited work suggests that in some cases poor language abilities seem to coincide with less precise domain-general auditory perception. Work in recent years has extended findings linking domain-general auditory processing abilities and language outcomes from the field of L1 acquisition to that of second language speech acquisition (Mueller et al., 2012).

Domain-general auditory processing and second language acquisition

While a host of variables have been linked to L2 learning outcomes including age-, input- and general aptitude-related factors (Flege & Bohn, 2021), they do not explain all the variability characteristic of L2 speech performance.

Using native language acquisition as a starting point, it is a well-established fact that native speech categories take years of linguistic input to develop and refine. Monolingual native English children tested on the perception and production of the /r-l/ phonetic contrast, for instance, have been shown to rely on the contrast's primary distinguishing cue, F3, by 4 years of age, while the incorporation of F2 information only begins to develop around the 8-9-year mark (Idemaru & Holt, 2013). Similarly, children's productions of the /d-t/ contrast in word-final positions do not exhibit adult-like characteristics even as late as 4 years old (Smith, 1979). The results of these and other studies (Morrongiello et al., 1984; Nittrouer, 2004) imply a gradual developmental trajectory of L1 speech categories requiring years of ambient language input for the adoption of adult-like cue weighting perceptual strategies. In the context of second language acquisition, this could mean that learners' abilities to effectively exploit the acoustic information in the available L2 input (which is significantly more limited than the ambient exposure to the L1) could help explain the wide variability in L2 learners' success. Promising new research into the relationship between domain-general auditory processing abilities and L2 speech acquisition offers compelling evidence in support of this.

Studies involving short-term perceptual training have shown that auditory acuity plays a role in determining how much learners benefit from training (see Hazan & Kim, 2010 for a correlation between F2 sensitivity and learning gains in the acquisition of a non-native phonetic contrast; Lengeris & Hazan, 2010 for the role of frequency discrimination abilities in short-term L2 vowel learning; and Wong & Perrachione, 2007 for pitch identification scores as predictors of the learning attainment of tones superimposed on pseudowords). When immersive contexts are concerned, individual differences in a number of auditory processing and neural encoding measures have been shown to correlate with L2 linguistic performance. Kachlicka et al. (2019) examined the speech perception and grammatical knowledge of adult native Polish speakers living in the UK. The study isolated lower auditory discrimination thresholds, better auditory-motor synchronization, and more robust neural encoding of sounds as accounting for variance in L2 perception and grammatical judgements. In another cross-sectional study, Saito et al. (2022a) collected spontaneous elicitations of Polish-English residents living in the UK and found a significant relationship between the level of L2 spoken vocabulary attainment and precise auditory processing abilities assessed in terms of auditory discrimination for pitch, duration, and amplitude risetime.

In a longitudinal context, Sun et al. (2021) examined the relationship between a range of behavioural and neural auditory processing measures and found that in the early stages of L2 immersion, music memory abilities correlated with prosody perception gains as measured 5 months apart. The same authors found that auditory discrimination sensitivity was linked to greater improvements in the fluency and accuracy of Chinese learners' English pronunciation after an 8-month study-abroad immersion (Saito, et al., 2020a).In combination, evidence from these studies suggests that learners' abilities across multiple dimensions of auditory perception can relate to both the rate and degree of language acquisition allowing learners with more precise auditory perception to make better use of L2 input and immersive exposure.

1.2.3 L2 Lexical stress teaching

Research has clearly shown that non-standard prosodic productions have a consistently adverse impact on L2 pronunciation scores independent of segmental or syllable structure errors. Poorly acquired prosodic features in the L2 frequently result in a range of communication costs associated with measures of accent, poor comprehensibility and intelligibility (Anderson-Hsieh et al., 1992; Field, 2005; Gallego, 1990; Isaacs & Trofimovich, 2012; Munro & Derwing, 1995). The accuracy of suprasegmental features can even be predictive of oral proficiency levels in the L2 (Kang & Johnson). However, failure to successfully acquire the prosodic contrasts of the target language can have subtler and often unexpected effects spilling outside the purely linguistic realm. Substantial evidence has emerged suggesting that the presence of a noticeable foreign accent can adversely impact listeners' attitudes towards L2 speakers, cause negative feelings and irritation, and even affect perceptions about the speaker's competence and trustworthiness (Anisfeld et al., 1962; Brennan & Brennan, 1981; Fayer & Krasinski, 1987; Fuertes et al., 2012).

This serves to highlight the importance of developing effective training methods for teaching L2 suprasegmentals in addition to the more traditional classroom-based instructional approaches that often treat suprasegmentals as add-ons to segmental pronunciation training. Practice has shown that classroom-based phonological instruction is mostly focused on addressing segmentals, even though the added benefit of suprasegmental instruction is associated with speech performance rated closer to that of native speakers (Moyer, 1999). However, reviewing the literature into L2 suprasegmental training exposes a lack of standardized, widely available training methods for teaching prosodic features to non-native speakers. Considering the case of English lexical stress, for instance, Dickerson (2004, 2015) designed a rule-based system for teaching L2 learners to predict the location of English stressed syllables in polysyllabic words by referring to the orthography of individual words. This approach requires that learners dissect their L2 productions against a set of spelling rules that will help them predict the location of lexical stress, then self-correct, and proceed with this practice in a strategy known as *covert rehearsal,* until they notice an improvement (Dickerson, 2000). This self-learning strategy

has been applied in classroom settings with instruction in the relevant orthographical rules and accompanying out-of-classroom practice showing significant improvements in lexical stress production abilities after extensive monthslong teaching interventions (Sardegna, 2012; Sardegna & Dickerson, 2023). In another study, following a 4-week intervention using Dickerson's orthographic-based approach (2004) combining instructional worksheets with access to extensive Youtube speech samples input, the authors found that the role of learner motivation was a key factor for determining the outcomes in autonomous learning (Sardegna & Jarosz, 2023). And while the covert rehearsal method has proven effective for helping learners process polysyllabic lexical stress, the training interventions normally require extensive periods of teacher-supported instruction and self-regulated practice spanning months.

Other instructional approaches have explored the efficacy of corrective feedback techniques such as recasts in facilitating the acquisition of L2 lexical stress (Goo & Mackey, 2013). Studies focusing on the production domain have reported mixed results. In Parlak and Ziegler (2017), L1 Arabic learners of English received corrective feedback on lexical stress errors during natural communication. In a pre-post-test paradigm, participants received feedback while engaging in either face-to-face communication or synchronous video chats. The authors hypothesized that interactional feedback would positively influence learners' L2 lexical stress pronunciation. However, an acoustic examination of key auditory dimensions signalling lexical stress revealed no statistically significant differences between learners' lexical stress productions from pre- to post-test regardless of training group. In a follow-up study with the same population, Parlak (2024) found that recasts provided during training with a role-play task had a positive effect on learners' productions of primary stress. At post-test, the trained learners demonstrated statistically significant improvements in their production of primary stress, characterized by longer duration and higher pitch for target words compared to pre-test values. However, a limitation noted by the author, was that production gains were measured for trained items only and more investigation is needed to determine whether improvements would generalize to untrained lexical stress contexts.

An overview of the literature on L2 segmental training has promoted other computer-based training approaches for teaching L2 phonology which do not require active self-monitoring

and have been found to generalize to untrained contexts demonstrating excellent results for teaching non-native speech contrasts. One of these methods is the high variability phonetic training (HVPT) paradigm presented in the section below.

HVPT overview and applications

HVPT beginnings

Early L2 phonetic training studies such as the perceptual fading technique (Jamieson & Morosan, 1986; Jamieson & Morosan, 1989) used synthetic speech stimuli as training materials with mixed results (Strange & Dittman, 1984). In this type of training, natural recordings were manipulated to first perceptually enhance relevant acoustic information and then gradually reduce the level of perceptual contrast to promote the development of appropriate sensitivity to the relevant acoustic dimensions. In later experiments, Logan et al. (1991) and Lively et al. (1993) introduced a procedure for training non-native speech contrasts with natural tokens where the target contrasts occurred in different phonetic contexts produced by multiple training voices. The research design followed a pre-test – training – post-test paradigm, with participants tested before and after training to measure their levels of improvement. The training in Logan et al. (1991) spanned fifteen sessions and used an identification task with accompanying performance feedback after each trial. Participants improved on both trained and untrained tokens, including novel items and novel voices. The authors posited that both their decision to use high stimuli variability along with the chosen training procedure created the optimal conditions for learners to zero in on the acoustic features most diagnostic of the trained contrast. This in turn facilitated the creation of robust category representations.

In their follow-up experiment, the authors compared the effectiveness of phonetic training by manipulating the type of variability in the training stimuli (Lively et al., 1993). This time they included 2 training conditions, where one group of participants were exposed to training stimuli produced by multiple talkers and the other group were trained on stimuli produced by a single talker (with phonetic context variability held constant between conditions). Participants who received the high variability talker training improved in their

pre-to-post test performance for both novel words and novel talkers. Improvements in the second, low talker variability training were confined to novel words but did not extend to novel talkers.

Since then, the proliferation in HVPT training studies has revealed that learning not only extends to novel items and speakers (Lively et al., 1993; Logan et al., 1991), but often also to the production domain (Bradlow et al., 1997; Bradlow et al., 1999). Additionally, training with HVPT seems to bring lasting learning benefits with retention ranging from two weeks to twelve months post-test assessment (Bradlow et al., 1999; Flege, 1995b; Iverson & Evans, 2009; Nishi & Kewley-Port, 2007; Rato, 2014; Thomson, 2012).

Flexibility of the HVPT training approach

Most HVPT studies have centred around providing training input that is naturally variable (stimuli produced by multiple talkers in varied lexical and phonetic contexts), embedded in forced-choice judgement tasks and accompanied by explicit feedback (Iverson et al., 2005; Lively et al., 1993; Logan et al., 1991; Pruitt et al., 2006). While these are considered the main elements characterising the paradigm, there is no rigidly prescribed way for conducting HVPT training. This has made the technique immensely popular as it can be adapted to different learning contexts and indeed, research practice has seen HVPT tailored to suit multiple experimental needs and constraints.

HVPT has been used with various native and target language pairings, different training focus (ranging from single (Logan et al., 1991; Shinohara & Iverson, 2018) to multiple phonetic contrasts (Iverson & Evans, 2009; Nishi & Kewley-Port, 2007; Thomson, 2012; 2016), and even broader stretches of discourse (Hirata, 2004) different types of training stimuli (e.g., real words and nonwords; Carlet & Cebrian, 2019; Thomson & Derwing, 2016)), and types of perceptual training tasks: identification (ID) and discrimination (DIS) tasks (Carlet & Cebrian, 2019; Flege, 1995b; Rato, 2014; Shinohara & Iverson, 2018).

HVPT effectiveness

Research utilizing the HVPT perceptual training has documented robust training effects averaging around 15 percentage points (Bradlow et al., 1999, 1997; Iverson et al., 2005; Shinohara & Iverson, 2018), with some variability depending on the target non-native contrast (5 – 29% improvements, Hirata et al., 2007; Nishi & Newly-Port, 2007; Rato, 2014). It is generally assumed that by being exposed to highly variable training stimuli, listeners learn to cope with category variability under conditions simulating real-world communicative contexts (Logan et al., 1991; Shinohara & Iverson, 2018; Thomson, 2018, but see Zhang, Cheng & Zhang, 2021). More traditional phonetic training approaches offering less variability seem to support learning, but the effects are more limited, typically confined to improved performance on the trained stimuli with poor generalization to novel instances of the target contrasts (Lively et al., 1993; Logan et al., 1991). It has been argued that talker variability in the HVPT context provides better clues to learners about which acoustic cues are relevant to category membership and which cues are less important, and consequently more variable across talker productions. In this sense, training with less variable stimuli (both relating to phonetic context, and number of speakers), affords lower opportunities for exposure to talker idiosyncrasies such as vocal tract specifics or variations in speaking rate, thus resulting in more limited improvements from the training.

Expanding the HVPT paradigm to prosodic training

Due to its versatility and the positive outcomes associated with it, HVPT training has been applied successfully to the acquisition of notoriously difficult non-native contrasts such as the English /r/ and /l/ contrast by Japanese adult learners (Bradlow et al., 1997; Lively et al., 1993; Logan et al., 1991; Shinohara & Iverson, 2018), as well as the acquisition of vowel contrasts (Fuhrmeister & Myers, 2017; Iverson & Evans, 2009; Lambacher et al., 2005; Nishi & Kewley-Port, 2007; Thomson, 2012), and other language acquisition contexts (Pruitt et al., 2006) for English and Japanese speakers acquiring Hindi stop contrasts.

But while HVPT interventions have proven to be an effective phonetic training method, research has almost exclusively applied high variability training to the problems of improving non-native vowel and consonant acquisition. An overview of the literature shows that addressing prosodic acquisition contrasts using the HVPT paradigm has received little attention. Several studies have applied the HVPT training approach to the acquisition of Mandarin tones by non-tonal language speakers (Perrachione et al., 2011; Sadakata & McQueen, 2014; Wang et al., 1999). These studies reported improved lexical tone perceptual accuracy at the post-test stage, and trained participants demonstrated long-term retention (Wang et al., 1999). However, short-term suprasegmental training using the high-variability procedure has so far been confined to the acquisition of lexical tones.
The current study in its second stage (Experiment 2) aimed to expand the application of the high variability method by training L2 learners on another instance of non-native prosodic contrasts – that of English word stress.

1.2.4 Literature summary

Adult second language acquisition is marked by varying degrees of challenges mostly associated with the distance between the native and target phonological systems (Best, 1995; Flege, 1995a; Flege & Bohn, 2021). Often sound contrasts which are not used in the native phonetic inventory (Fox et al., 1995), or which employ acoustic dimensions in a different way to how they are used in the L1 (Goto, 1971; Strange & Dittmann, 1984) are particularly difficult to master as an adult learner. And while research has documented many problematic instances of non-native contrast acquisition (Best et al., 2003; Flege et al., 1997; Flege & MacKay, 2004; Iverson et al., 2003), an equally profitable area of inquiry has sought to explain the sources behind widely observed differences in learner outcomes. Indeed, adult phonological L2 acquisition is characterized by great individual variability in ultimate attainment even after accounting for experience-related variables such as AOA, quantity and quality of input, and length of residence (Díaz et al., 2012, Abrahamsson & Hyltenstam, 2008; Birdsong, 2007; Bongaerts et al., 1997). In an attempt to capture the sources of this variability, researchers initially explored differences in language learning aptitude skills as possible predictors of L2 acquisition success (Carroll, 1981; Doughty, 2018; Li, 2016). This was followed by a surge of interest in more implicit, cognitive-based abilities

such as attentional, inhibitory control, or working memory factors (Darcy et al., 2015; Ghaffarvand & Werner, 2019; Lee & Révész, 2021; Mercier et al., 2014; Révész, 2012; Safronova & Mora, 2013; Miyake & Friedman, 1998).

More recently there has been increased focus on investigating the role of domain-general auditory processing abilities in L2 learner attainment (see Saito et al., 2022b; Saito, et al., 2024, for primers on the work carried out in this area of research). Examining the relationship between auditory processing abilities and language acquisition was initially motivated by similar investigations into first language development (Anvari et al., 2002; Bavin et al., 2010; Douglas & Willatts, 1994; Lamb & Gregory, 1993; Tierney et al., 2021). Research into establishing a link between how well individuals perceive various aspects of acoustic information and their L2 language performance has shown promising results in both cross-sectional studies (Kachlicka et al., 2019; Saito et al., 2022a), and over longitudinal projects probing both L2 production (Saito et al., 2020a), and L2 perception skills (Sun et al., 2021). Importantly, several studies have also reported that perceptual acuity along acoustic dimensions relevant for the acquisition of specific non-native contrasts, can be predictive of learning gains after short-term perceptual training (Hazan & Kim, 2010; Lengeris & Hazan, 2010; Wong & Perrachione, 2007). However, research so far has investigated relatively broad aspects of L2 language outcomes (Saito et al., 2022a; Saito et al., 2020a), or segmental acquisition (Lengeris & Hazan, 2010 for vowels; Omote et al., 2017 for consonants).

The current dissertation aimed to expand research into the role of domain-general auditory perception abilities in L2 acquisition by examining a specific suprasegmental feature – the acquisition of English lexical stress by learners of a non-intonation language (native Mandarin speakers). The decision to investigate prosodic acquisition was motivated by the fact that it has received little attention despite empirical evidence showing that poorly acquired L2 prosodic features result in a range of communicative costs including a noticeable foreign accent and low speech intelligibility (Isaacs & Trofimovich, 2012; Kang et al., 2010; Moyer, 1999; Munro & Derwing, 1995; Pennington & Richards, 1986). Mandarin speakers were tested as their native language experience with lexical tones affects how they process English suprasegmentals by relying more heavily on F0 information in perception

and manipulating prosodic features in a non-native manner in production (Archibald, 1997; Chen et al., 2001; Hung, 1993; Juffs, 1990; Ou, 2016; Jasmin et al., 2021). To this end, in Experiment 1 of this dissertation I collected a battery of domain-general auditory processing measures aimed to assess Mandarin speakers' baseline perceptual acuity and explore whether individual differences in these measures can explain variations in lexical stress performance.

Furthermore, the present dissertation aimed to scrutinize the suitability of high variability phonetic training – a well-established perceptual training paradigm used primarily in segmental training (Lively et al., 1994; Logain et al., 1999), for teaching lexical stress to tonal language speakers. To this end, I designed a high variability prosodic training using natural recordings by multiple speakers, and offering learners corrective feedback. Experiment 2 presents this training and it reports on the effectiveness of the HVPT paradigm in a suprasegmental context. I further examine the possible predictive role of domain-general auditory processing abilities in explaining lexical stress acquisition following perceptual training.

1.3 Current dissertation scope

The focus of the current dissertation is the prosodic acquisition of L2 lexical stress by native Mandarin speakers. It covers two separate but related areas of investigation from a cross-sectional (Experiment 1), and a longitudinal (Experiment 2), perspective. The first overarching goal of the two experiments centred around scrutinizing the role of individual differences in a set of domain-general auditory processing abilities and Mandarin learners' lexical stress processing (Time 1, Chapter 2) on the one hand, and examining the predictive power of these same measures in accounting for variability in learners' lexical stress learning gains after a 6-session short-term high variability prosodic training (Time 2, Chapter 3). The second main goal of the dissertation involved implementing a training intervention in the same sample of participants and exploring the effectiveness of the HVPT training paradigm in Mandarin learners' acquisition of English lexical stress (Experiment 2, Chapter 3).

Experiment 1

In Experiment 1 of the current text, I tested the link between domain-general auditory processing abilities and L2 prosodic perception and cue weighting. To this end 132 inexperienced native Mandarin learners of English were recruited from Chinese universities (testing at Time 1). Participants were tested on their English lexical stress perception, and lexical stress cue weighting strategies. Further, a range of auditory processing measures were collected to assess their abilities in the following areas: auditory discrimination, dimension selective attention, and auditory-motor integration. For Experiment 1, I investigated two main research questions:

1) Can individual differences in domain-general auditory processing abilities (including auditory discrimination, dimension-selective attention, and auditory-motor integration skills) explain variability in lexical stress performance in native Mandarin learners of English?

2) Do individual differences in participants' domain-general auditory processing abilities explain variability in their lexical stress cue weighting strategies?

Details about the theoretical motivations for the measures collected in Experiment 1 are presented below.

Lexical stress processing

Mandarin learners' lexical stress processing was measured in 2 separate ways. Data was collected on the most commonly tested mode of L2 processing: perception abilities. The study also employed another method of assessing lexical stress processing which examined the contribution of specific acoustic information to participants' lexical stress categorizations. To recap, listeners integrate information from multiple acoustic sources (dimensions) to arrive at categorizations about which speech contrast they've heard (Francis et al., 2000; Holt & Lotto, 2006; Idemaru et al., 2012), a process called dimensional weighting. However, research has found that even within native speakers of the same language there are large individual differences in how listeners weigh acoustic information which remain stable across time and are not related to their speech productions (Idemaru et

al., 2012). Cue weighting plays a fundamental role when learning a second language as learners often experience difficulties in acquiring L2 native-like patterns of cue weighting, and often persistently rely on familiar L1 strategies which are often ill-suited to the L2 context (Ingvalson et al., 2012). Native Mandarin speakers, for instance, are known to rely on pitch information significantly more compared to native English speakers both in the categorization and production of English lexical stress (Wang, 2008; Zhang et al., 2008; Yu & Andruski, 2010; but also refer to Chrabaszcz et al., 2014). This could be related to the crucial role pitch plays in Mandarin Chinese to convey meaning (Howie, 1976). Bearing this in mind, participants' cue weighting strategies were also assessed using a prosodic cue weighting task.

With respect to auditory processing, their domain-general auditory perception was conceptualized along 3 separate processing abilities: perceptual acuity, auditory attentional control, and auditory-motor integration abilities (see Saito et al., 2024 for an overview of the auditory processing construct as it has been implemented in SLA research).

Perceptual acuity

The first set of auditory processing skills that was tested was participants' domain-general auditory perceptual acuity, or the level of perceptual sensitivity they have to specific acoustic dimensions. This sensitivity is represented through individual discrimination thresholds and is commonly tested using adaptive discrimination tests (Kachlicka et al., 2019, Sun et al., 2021). These tests employ synthesized complex tones (i.e., non-verbal sounds) varying along a single dimension and they have been used for online testing with a fair-to-good test-retest reliability (Saito & Tierney, 2022). Using an ABX discrimination paradigm, participants are presented with 3 tones on each trial. One of the tones is always a baseline and participants have to decide which of the other two sounds differs from the baseline. The stimuli presentation is controlled such that the level of difficulty (i.e., the size of the perceptual difference between the baseline and comparison tones) is continually adapted to participants' performance. The resulting thresholds indicate the smallest size difference within a specific dimension that participants can perceptually detect. Lower thresholds mean that the individual can detect subtler changes in a given dimension, and therefore, they have better perceptual acuity for that dimension.

In the context of L2 acquisition, a number of studies using tests for auditory acuity have linked better perceptual sensitivity to greater improvements from short-term phonetic training (Hazan & Kim, 2010; Lengeris & Hazan, 2010; Qin et al., 2021; Wong & Perrachione, 2007), and to better L2 outcomes in immersive settings (Kachlicka et al., 2019).

In the current experiment, I collected Mandarin learners' auditory discrimination thresholds for pitch, formant, and amplitude risetime. These 3 acoustic dimensions were selected for testing as they contribute to distinguishing lexical stress contrasts (Beckman, 1986; Bolinger, 1961; Fry, 1955, 1958, 1965; Lieberman, 1960), and would be particularly relevant for investigating any possible effects of individual differences in auditory acuity and L2 lexical stress acquisition.

Auditory attentional control

The second set of auditory measures assessed auditory selective attention. As listeners integrate information from multiple overlapping acoustic dimensions to arrive at perceptual categorization (Lisker, 1986; Winter, 2014), they prioritize certain dimensions more than others (i.e., cue weighting). They do this primarily by relying on their experience about which dimensions reliably signal category membership (Holt et al., 2018; Toscano & Murray, 2010). And while perceptual weighting strategies in the L1 can show individual differences reflecting variability in perceptual acuity (Jasmin et al., 2020), native language experience can also affect perceptual weighting strategies in the L2 (Ingvalson et al., 2012; Iverson et al., 2003; Jasmin et al., 2021). The existing literature points to attentional allocation among different sources of acoustic information as playing a role in speech cue weighting, and ultimately speech perception (Francis & Nusbum, 2002; Gordon et al., 1993; Holt at al., 2018). This attentional perspective proposes that acoustic information that has been highly relevant in perceptual experience tends to capture auditory attention more and leads to the upweighting of that information in perceptual judgements. In their study, Jasmin et al. (2021) showed that native Mandarin speakers exhibited greater reliance on pitch in both speech and non-speech categorization compared to native English speakers. They also showed increased perceptual salience for pitch and an inability to selectively attend to amplitude while ignoring pitch when this was required by the task. This dimensional salience

for pitch has been shown to remain highly resistant to change even after extensive immersion in a non-tonal language environment (Petrova et al., 2023).

Given that in certain contexts, the successful acquisition of non-native speech contrasts depends on listeners' abilities to direct their attention to acoustic cues they may historically have treated as irrelevant in their L1 (Ingvalson et al., 2012), or to ignore dimensions they normally find highly useful, researchers have proposed that listeners with better abilities to selectively attend to acoustic dimensions might be more successful at acquiring L2 categories (Saito et al., 2024; Symons et al., 2021). To test this hypothesis, the current experiment employed 3 dimension selective attention tasks aimed at assessing Mandarin learners' abilities to selectively attend to a specified dimension and ignore simultaneous changes in a different dimension. The tested dimensions were those of pitch, formant, and amplitude risetime.

Auditory-motor integration

The final set of domain-general auditory processing abilities tap into participants' auditory-motor integration skills by testing their abilities to accurately reproduce rhythmic and melodic patterns. There is evidence from studies into L1 language development that the ability to remember and reproduce rhythmic and melodic sequences (tapping into temporal and spectral encoding abilities, respectively) is linked to individual differences in reading and verbal memory (Anvari et al., 2002; Douglas & Willatts, 1994; Lamb & Gregory, 1993; Tierney et al., 2017; Tierney et al., 2021). Moreover, research into L2 speech acquisition has found a link between music reproduction abilities and gains in L2 prosody after a short immersion period (Sun et al., 2021). Notably for the present study, research carried out with congenital amusics (i.e., individuals with a music deficit accompanied by a dimension-specific deficit for pitch) has indicated that when processing prosody, listeners with poor perceptual resolution for pitch upweight reliance on duration and downweight reliance on pitch even when intervals along the pitch dimension were large enough to be detected (Jasmin et al., 2020). Following this line of work, two auditory-motor integration tasks testing melodic and rhythmic reproduction abilities were included in the battery of tests assessing domain-general auditory processing as performance on them could have a bearing on participants' lexical stress processing.

Expected outcome

Based on the reviewed work, I predicted that individual differences in participants' abilities on the tested domain-general processing skills would be related to lexical stress processing. In fact, some of the collected auditory processing measures assessed different aspects of processing pitch, formant, and risetime – three dimensions important for marking lexical stress in English. Albeit there is no agreement in the literature about which of the the four acoustic cues (pitch, intensity, duration, or spectral quality) are more important for signalling lexical stress, native Mandarin speakers have a clear bias towards weighting pitch more heavily than other acoustic dimensions in a variety of contexts (Wang, 2008; Yu & Andruski, 2010; Zhang, 2012; Zhang et al., 2008; Zhang & Francis, 2010). One possibility, therefore, is that enhanced pitch processing abilities will translate into better English lexical stress processing. This is a likely result if as it has been suggested pitch is also the most robust cue to lexical stress in English (Beckman, 1986; Fry, 1958; Lieberman, 1960; Morton & Yassem, 1965). Alternatively, if pitch is less important for marking lexical stress, then variability in either risetime or formant performance could be predictive of lexical stress acquisition performance.

Additionally, as a deficit in music reproduction abilities has been linked to downweighting pitch reliance in favour of duration in prosodic perception (Jasmin et al., 2020), I also hypothesised that music reproduction skills would have an effect on lexical stress cue weighting strategies.

Experiment 2

102 of the participants in Experiment 1 elected to also participate in the second stage of the study (Experiment 2, Chapter 3). These learners were exposed to either a 6-session high variability phonetic training programme, or to the same number of vocabulary training sessions. Data on their lexical stress performance was collected in a post-test (Time 2). The effectiveness of the HVPT paradigm was examined by comparing the trained and untrained learners' performance on lexical stress processing in terms of perception, and cue weighting.

In Experiment 2 I had several research goals. Firstly, I wanted to assess the effectiveness of high variability phonetic training for teaching L2 prosody. As outlined in the previous section, participants' lexical stress processing operationalized as perception, and categorization were collected at Time 1. The HVPT-trained group's performance at post-test was benchmarked against the same measures collected at Time 1 to examine learning gains in lexical stress perception compared to those of the vocabulary-trained group. Next, I compared participants' cue weighting strategies from Time 1 to Time 2, and between the experimental and control conditions to investigate if short-term training led to shifts in perceptual cue weighting. Finally, I investigated if individual differences in the domain-general auditory processing measures recorded at Time 1 had modulated the benefits of the perceptual training.

HVPT prosodic training

I based the perceptual training used in this experiment on the empirically successful HVPT training paradigm and adapted it to the prosodic context. The training protocol followed HVPT's original paradigm including the following elements: multiple training voices, multiple phonetic contexts, a forced choice judgement task, and training feedback (Lively et al., 1993; Logan et al., 1991).

Expected outcomes

The HVPT approach has a solid track record of successfully training L2 learners on non-native segmental contrasts (Bradlow et al., 1997; Iverson & Evans, 2009; Lambacher et al., 2005; Lively et al., 1993; Logan et al., 1991; Nishi & Kewley-Port, 2007), and even for teaching lexical tones to non-tone language speakers (Sadakata & McQueen, 2014; Wang et al., 1999). However, there has not been enough research into how well the paradigm would fare in other prosodic training contexts. However, on the strength of its previous training applications, I predicted that adapting the HVPT paradigm for training lexical stress to non-native speakers would result in significant learning gains.

In terms of participants' cue weighting strategies, I also hypothesised that their reliance on pitch information would undergo changes after training. Here there are several considerations that can be made about the weight of different acoustic dimensions in contributing to lexical stress. On the one hand, Mandarin learners of English are known to place significantly more weight on pitch compared to native English speakers in prosodic categorizations (Wang, 2008). It is possible that exposure to rich, highly variable acoustic information in the training stimuli would prompt a shift towards more native-like cue weighting distributions and down-weighting of pitch reliance. Alternatively, as pitch is often regarded as the primary cue for lexical stress processing by some authors (Bolinger, 1958; Fry, 1955; Morton & Yassem, 1965; but see Beckman & Edwards, 1994), it is possible that Mandarin learners would shift their cue weighting strategies towards even greater reliance on pitch in categorising stress.

Lastly, there is some evidence to suggest that individual differences in auditory sensitivity has some predictive power for the size of learning gains after short-term perceptual training (Hazan & Kim, 2010; Lengeris & Hazan, 2010; Wong & Perrachione, 2007). Consequently, I expected that individual differences in domain-general auditory processing abilities recorded at Time 1 would explain some variability in lexical stress improvements at Time 2 for the HVPT-trained group.

1.4 Dissertation outline

In this chapter I discussed the main issues related to the acquisition of second language pronunciation, the potential underlying second language processing difficulties, the possible sources of individual differences in L2 acquisition success, as well as some of the tried- and-tested L2 training solutions.

Chapter 2 focuses on Experiment 1 – a cross-sectional investigation of English lexical stress acquisition by native Mandarin learners based in China. It outlines the motivation, methodology, and results of the experiment. Chapter 3 explores lexical stress acquisition

longitudinally and deals with perceptual training. It introduces the high variability prosodic training paradigm and discusses Time 2 findings. Finally, Chapter 4 provides a general discussion of the experimental findings from the studies presented in Chapters 2 & 3. This summary is followed by a discussion of the theoretical implications of the findings, some limitations of the current experiments, and future directions for research.

Chapter 2 Experiment 1: Domain-general auditory processing and prosodic acquisition (cross-sectional perspective)

2.1 Introduction

The focus of Experiment 1 was examining the relationship between domain-general auditory processing abilities (conceptualized in terms of auditory discrimination sensitivity, auditory-motor integration skills, and dimension selective attention), and English prosodic acquisition. Specifically, the current study explored the lexical stress processing of inexperienced native Mandarin learners of English with the aim of gaining more insights into the factors influencing the success of L2 learners' lexical stress performance. The study was designed to address two main research questions:

1) Can individual differences in domain-general auditory processing measures (encompassing psychophysical discrimination thresholds, dimension selective attention abilities, and auditory-motor integration skills) explain individual differences in lexical stress perception performance in native Mandarin learners of English?

2) Do individual differences in participants' domain-general auditory processing abilities explain variability in their prosodic cue weighting strategies?

As outlined in the literature review (Chapter 1), a number of perception and auditory processing abilities have been linked to language performance, both in the context of child first language acquisition (Boets et al., 2008, 2011; Casini et al., 2018; Goswami, 2015; Goswami et al., 2011; Gibson et al., 2006; White-Schwoch et al., 2015), and more recently, in adult second language speech acquisition (Kachlicka et al., 2019; Lengeris & Hazan, 2010; Saito et al., 2020b; Saito et al., 2022b; Sun et al., 2021). However, while these studies have added to our understanding of the role played by auditory perception abilities in second

language acquisition, for the most part research has either related auditory processing to broad measures of language outcomes (Saito et al., 2022a for vocabulary acquisition; Saito et al., 2020b for phonological and grammatical learning outcomes), or to isolated segmental contrasts (Lengeris & Hazan, 2010 for vowel contrasts; Omote et al., 2017 for consonant perception). The current investigation took an approach more focused on prosodic acquisition by collecting participants' performance data on a range of auditory processing abilities and examining if individual differences in those could explain learners' performance with a specific prosodic contrast, that of lexical stress processing.

Mandarin speakers' perceptual experience with their native tonal system has been shown to have a profound and wide-ranging impact both at the level of their auditory processing abilities and the perceptual strategies they rely on in a variety of contexts. For instance, experience with tones where pitch is a robust cue to changes in lexical meaning (Gandour, 1978; Howie, 1976, but see Liu & Samuel, 2004; see Whalen & Xu, 1992 for use of the other acoustic correlates of lexical tones), has been linked to a range of enhanced pitch-related auditory processing abilities (see Bidelman et al., 2013; Giuliano et al., 2011; Zheng & Samuel, 2018 for better pitch acuity; Deroche et al., 2000 for enhanced sensitivity to pitch contours, and Creel et al., 2018; Giuliano et al., 2011; Pfordresher & Brown, 2009 for better pitch interval discrimination, but see also Bent et al., 2006; Peretz et al., 2011; Stagray & Downs 1993 for evidence against a distinct pitch processing advantage). Native tonal experience has also been shown to have consequences in the context of foreign language acquisition. Research into prosodic acquisition in this population has documented both perception and production challenges in the context of English lexical stress (Archibald, 1997; Hung, 1993; Juffs, 1990; Yu & Andruski, 2010; Zhang et al., 2008), phrase boundaries (Zhang, 2012), and broader sentential stress (Chen et al., 2001), as well as increased perceptual salience of pitch with related inability to ignore pitch information when explicitly required by the task (Jasmin et al., 2021). For the most part these challenges can be traced back to native language experience and are manifested in the L2 as overreliance on pitch information in perceptual categorization, and overemphasis on the manipulation of pitch characteristics (both pitch height and pitch contour) in prosodic production.

The current experiment aimed to investigate if individual differences in learners' domain-general auditory processing abilities measured in several domains (discrimination acuity, dimension selective attention, and auditory-motor integration abilities) and across several dimensions important in signalling lexical stress (pitch, formant, amplitude risetime), can explain individual differences in English lexical stress processing. Based on the reviewed literature, it was hypothesized that individual differences in participants' domain-general auditory processing abilities will help explain some of the observed variance in their lexical stress perception and cue weighting strategies.

More specifically, given native Mandarin speakers' well-established pitch processing advantages, several predictions could be made. Firstly, it was hypothesised that more precise pitch processing abilities (in terms of discrimination and the ability to selectively attend to pitch), would relate to learners' English lexical stress perception and their reliance on pitch information for categorizing lexical stress. As outlined in Chapter 1, English lexical stress is cued variously by pitch, duration, intensity, and vowel quality (Beckman, 1986; Bolinger, 1958; Fry, 1955). As the primary acoustic dimension most frequently proposed among the four, is pitch (Beckman, 1986; Fry, 1958; Lieberman, 1960; Morton & Yassem, 1965), Mandarin learners' native tonal experience could confer an advantage in their English lexical stress perception and categorization performance. On the other hand, as some authors have taken the view that pitch, due to concurrent phrase- and sentential-level stress, might actually be less reliable in conveying lexical stress contrasts (Sluijter & van Heuven, 1996), it is possible that learners' auditory processing abilities along the other acoustic cues signalling stress (specifically, formant and risetime) could actually play a more important role in determining lexical stress performance. Finally, as prosodic perception has been linked to music reproduction abilities (Saito et al., 2020b; Saito et al., 2019, 2021), it is also hypothesised that participants' music memory performance could relate to their lexical stress perception. Additionally, auditory-motor integration abilities could also relate to listeners' lexical stress cue weighting strategies as suggested by evidence that individuals with congenital deficits in musical abilities such as congenital amusia, also characterized by poor pitch perception (Ayotte et al., 2002; Dalla Bella et al., 2009; Tillman et al., 2016), seem to down-weight reliance on pitch information when categorizing prosodic contrasts (Jasmin et al., 2020). In the context of the present study, it was hypothesized that individual

differences in native Mandarin speakers' music memory abilities would relate to the weights they assign to the pitch and duration dimensions when categorizing lexical stress.

In sum, the current study attempted to examine the relationship between native Mandarin learners' domain-general auditory processing abilities and their English lexical stress acquisition. The stated research questions were tested by recruiting relatively low proficiency (lower- to upper-intermediate) Mandarin learners of English living in China.

2.2 Methods

2.2.1 Participants

All participants recruited for the experiments reported in this dissertation lived in Mainland China at the time of testing. Both experiments were widely advertised on Chinese student message boards, social media, and messaging platforms. To ensure sufficient power for detecting medium-sized effects (f^2=0.15) with 11 predictors, a significance level of α=0.05, and a desired power of 1−β=0.80, a total sample size of N=100 participants was determined. This corresponds to approximately 9.1 participants per predictor. Calculations were performed using established methods for F-test power analysis (Faul et al., 2007).

One hundred and thirty-two (132) participants registered and completed Experiment 1. To maximize the quality of the reported data, a screening threshold for participants' lexical stress performance was set a priori. Only participants who achieved identification accuracy greater than 55% in the lexical stress perception task were included. This criterion resulted in the inclusion of 102 participants in the final analysis (87 females, 15 males), aged 18 – 48 (*M* = 22.46, *SD* = 4.45). For the analysis that investigated the relationship between individual differences in auditory processing abilities and variability in learners' cue weighting strategies, the data of a further 7 participants were also excluded as they did not show a significant relationship ($p < .05$) between at least one of the stimulus dimensions (pitch or

duration) and their categorization decisions in the lexical stress cue weighting task (see section 2.3).

Participants answered a questionnaire designed to collect demographic and language background information and completed a short vocabulary test (LexTALE taken from Lemhöfer & Broersma, 2012; see section 2.2.6.1) to establish their English proficiency levels. Mean scores on the LexTALE vocabulary test were *M* = 0.56 (*SD* = 0.1), placing participants in the following proficiency levels – Lower intermediate and lower (N = 68), Upper intermediate (N = 31), and Advanced and proficient (N = 3). Refer to section 2.2.6.1 for an approximate correspondence between LexTALE scores and CEFR proficiency levels.

All participants were born and raised in Mainland China, and while some of them were also familiar with other regional dialects, they were all native Mandarin speakers. Apart from short stays abroad (< 1 month), none of the participants reported having lived in a foreign country. They all learned English outside of the home in formal instructional settings, starting on average at around 8.26 years of age (*SD* = 2.57), and had spent approximately 12.39 years studying English (*SD* = 3.17). Participants self-rated their English language proficiency skills on a scale of 1-10 (*1* for no proficiency and *10* for high proficiency) for their reading (*M* = 5.61, *SD* = 2), writing (*M* = 4.18, *SD* = 1.75), comprehension (listening) (*M* = 5.03, *SD* = 1.9), and speaking abilities (*M* = 3.69, *SD* = 1.71). In terms of English use participants estimated they spent approximately 6.08 hours on average using English per week (*SD* = 6.10). Refer to Table 2 below for a breakdown of their English use into the four domains of reading, writing, listening, and speaking.

Only twenty-five (25) participants indicated they received pronunciation training as either part of their school curriculum, or by attending dedicated pronunciation training classes. Twenty (20) participants reported they received training on English lexical stress. To ensure clarity, the questionnaire included an example of a minimal lexical stress word pair ("REcord – reCORD") to illustrate the concept of lexical stress. However, as the questionnaire was administered remotely and without the presence of a researcher, it is unclear if all participants understood the concept of lexical stress correctly. Consequently, only participants' general pronunciation training experience, rather than their reported lexical

stress training, was included as a predictor variable in the statistical analyses presented in the Results section (section 2.3). When asked to rate their confidence in the perception of English lexical stress, participants reported moderate confidence in their ability to perceive stress (*M* = 4.45, *SD* = 2.12, elicited on a 10-point scale). In terms of musical training background, 63 participants reported having received musical training for an average of 3.51 years (*SD* = 3.27). For full details about the demographic characteristics of participants, see Table 2 below.

Table 2

Participant characteristics – Experiment 1

Native Mandarin speakers (N= 102)			
	M	**SD**	**range**
Age	22.46	4.45	18 - 48
AOA	8.26	2.57	3 - 17
Years of English language learning	12.38	3.17	0 - 18
Self-reported English reading skills	5.60	2.00	1 - 10
Self-reported English writing skills	4.18	1.75	1 - 9
Self-reported English listening skills	5.03	1.90	1 - 8
Self-reported English speaking skills	3.69	1.71	1 - 7
Hours English use per week	6.08	6.10	0 - 33
Hours reading in English per week	3.18	4.31	0 - 28
Hours writing in English per week	0.53	0.81	0 - 4
Hours listening in English per week	1.97	2.47	0 - 15
Hours speaking in English per week	0.38	0.82	0 - 4

Confidence in perceiving English word stress	4.45	2.12	1 - 10
Music training	3.51	3.27	0.5 - 15

Note. M = mean, *SD* = standard deviation, *AOA* = age of English language acquisition.

All testing materials and procedures were approved by the Institute of Education, UCL and informed consent was obtained from all participants. None of the tested participants reported having any impairments or disabilities, and they all confirmed normal hearing and normal or corrected-to-normal vision.

2.2.2 Design and study setup

Data collection for this experiment took place online, hosted on the Gorilla Experiment Builder platform (Anwyl-Irvine et al., 2019, www.gorilla.sc). This platform has been empirically validated as a reliable tool for conducting reaction-time-sensitive experiments in online settings. Gorilla has been tested across various technical setups (e.g., in-lab and personal computers), participant groups (children and adults), experimental contexts (laboratory and remote), and diverse web browsers, operating systems, and internet connections (Anwyl-Irvine et al., 2019; Anwyl-Irvine et al., 2020). Using a Flanker Task designed to measure attentional skills, Anwyl-Irvine et al. (2019) demonstrated that the Gorilla platform successfully captured the expected significant differences in reaction times (RTs) between congruent and incongruent trials in the task. Variability across participants and devices was within acceptable limits for experimental research. Additionally, in a comparative analysis of online experimental platforms, Gorilla, along with other leading platforms, was shown to deliver reasonably accurate timing for both stimulus presentation and response time recordings, maintaining reliable performance across diverse setups (Anwyl-Irvine et al., 2020).

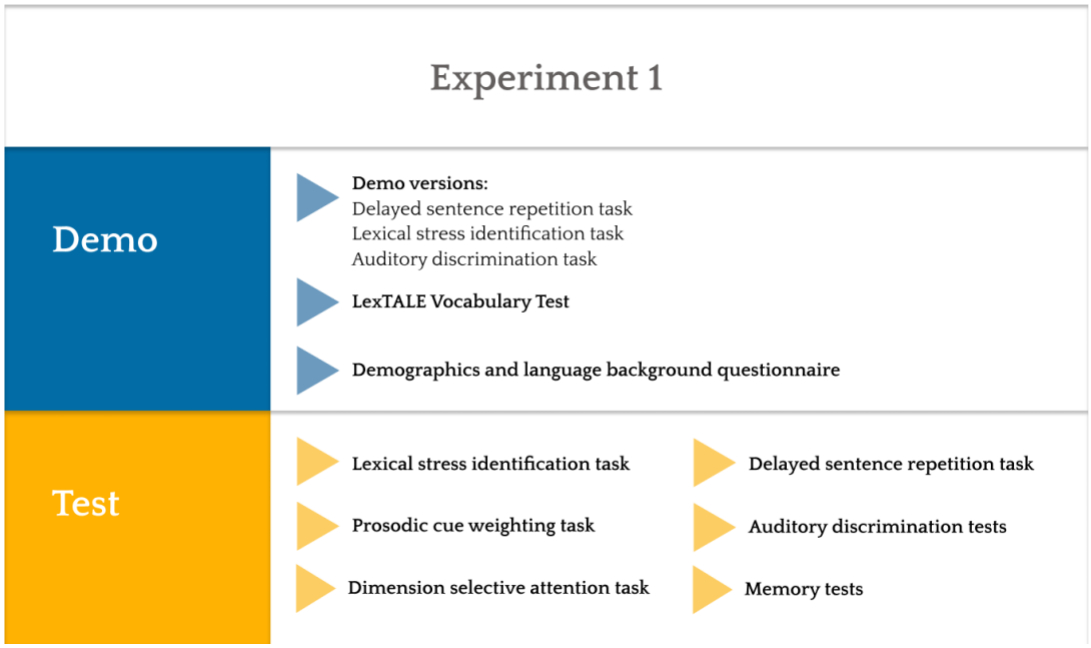
In the current study, participants received a link to access the experiment on their personal computers through a web browser. Prior to the day of testing, all participants received detailed instructions to ensure the successful running of the experiment. Through automated procedures for device and browser compatibility, access to the study was

allowed only for computer and laptop users running the Chrome browser. Restricting the testing environment to a specific device and browser type made it possible to maximize the consistency of stimulus presentation across participant sessions thus improving the quality of data collection.

All initial instructions and the full experiment were translated into Simplified Chinese to ensure full understanding of the procedures and tasks involved. After registering for the study, participants were sent a troubleshooting guide in case of encountering any problems with the testing platform or technical setup. A collaborator based in China was communicating directly with participants, sending out all experiment-related documentation (instructions and troubleshooting guide) and individual login details for the experiment platform. Participants first completed a short demo session designed to maximize the quality of the data collection. This session allowed participants to get familiar with the Gorilla interface and it gave them the opportunity to test out their equipment prior to the full experiment. Figure 2 below shows the full research design for the current experiment.

Figure 2

Schematic representation of the experimental design illustrating the number and order of tasks presented to participants in Experiment 1



2.2.3 Overview of materials and task presentation

Participants first completed a short 15-minute taster session with demo versions of the lexical stress perception and auditory discrimination tasks. In the same session demographic information was collected, and participants were tested on their vocabulary knowledge to establish general proficiency levels.

The actual experimental session took place on a different day and comprised seven tasks, which were presented in the following order: (i) a lexical stress identification task; (ii) a lexical stress cue weighting task; (iii) a dimension selective attention (DSA) task; (iv) a delayed sentence repetition (DSR) task[1]; (v) auditory discrimination tests (formant, pitch,

[1] Data collected for this task is out of the scope of this dissertation. While production data was collected using a Delayed sentence repetition task, subsequent inspection of the recorded samples revealed severe quality problems. As testing was conducted fully online, there were wide differences in the recording equipment used by participants. This resulted in multiple instances of data loss, and varying degrees of static, background noise, and poor sound quality. Consequently, analysis of the production data was not pursued further.

and amplitude rise time discrimination presented in a randomized order); (vi) a working memory digit span test; and (vii) two auditory-motor integration tasks (testing melodic and rhythmic memory).

See Table 3 below for a list of the abilities tested in the current experiment. To recap, participants' English lexical stress proficiency was measured using 2 tasks tapping into lexical stress perception and lexical stress cue weighting. This two-pronged approach aimed to examine Mandarin learners' lexical stress processing holistically by testing not only perception abilities, but also perceptual reliance on acoustic dimensions when categorizing L2 lexical stress.

As highlighted in the literature review, the acoustic dimensions that covary with speech categories are not perceptually equivalent (Francis et al., 2000; Holt & Lotto, 2006; Idemaru et al., 2012) with some of the integrated sources of information weighted more heavily than others in perceptual categorization. The existence of language-specific cue weighting strategies, or the tendency for language communities to utilize specific sets of acoustic cues to signal speech categories, can be a source of difficulty for L2 learners' trying to acquire their target language cue weighting strategies (Ingvalson et al., 2012). In the case of native Mandarin speakers, research has shown that they weigh F0 information more heavily than do native English speakers when categorizing and producing L2 English stress (Wang, 2008; Zhang et al., 2008; Yu & Andruski, 2010; but also refer to Chrabaszcz et al., 2014). In the current study, I collected participants' lexical stress cue weighting functions in an attempt to tease apart the role of individual differences in learners' domain-general auditory abilities and the degree of reliance on acoustic information when processing English lexical stress.

A second area of testing focused on domain-general auditory processing abilities conceptualized in terms of auditory perceptual acuity, auditory attentional control, and auditory-motor integration abilities, which are hypothesised to relate to lexical stress learning outcomes (see section 1.3 of the literature review for a more comprehensive discussion of the motivation for including these 3 sets of auditory processing abilities).

Finally, participants' general language proficiency and linguistic background were collected through a demographics questionnaire (see APPENDIX D) and a LexTALE vocabulary test. Furthermore, performance on a working memory task was measured with a view to controlling for possible working memory effects on one of the training tasks in Experiment 2 (see section 3.2.3.3 for more information), and as such data from this task will not be explored in this chapter.

Table 3

List of the speech and auditory-processing abilities tested in Experiment 1

Domain of testing	**Task**	**Measures**
L2 prosodic performance	Lexical stress perception task	Lexical stress perception
	Prosodic cue weighting task	Lexical stress cue reliance
Domain- general auditory processing performance	Auditory discrimination tasks – 3 tasks (measuring auditory discrimination for pitch, formant, and risetime)	Auditory perceptual acuity
	Dimension selective attention tasks – 3 tasks (measuring dimension selective attention for pitch, formant, risetime)	Auditory attentional control
	Rhythm and melody memory tasks (2 tasks – for rhythm memory and for melody memory)	Auditory-motor integration abilities
Cognitive and language background	LexTALE vocabulary test	General language proficiency
	Digit span task	Working memory

	Background questionnaire	Demographics and language experience

The average running time for the experiment was approximately 60 minutes. The sections that follow detail the stimuli generation, test procedure, and data analysis performed for each of the seven tasks.

2.2.3.1 Lexical stress test stimuli

The speech test stimuli for Experiment 1 were recorded by 2 native speakers of British English (1 female, 1 male). Due to the Covid-19 crisis, recordings were obtained remotely. In each case, the voiceover actors had access to a home recording studio. All readings were taken at a sampling rate of 44.1 kHz with 16-bit quantization.

The target tokens were real English minimal stress pairs produced in question-response sentences (Table 4 below shows typical sentence pairs used for the elicitation of the recordings). Sentence length was controlled for by matching the number of syllables in each response sentence to ensure they fell within the 6-8 syllable range. This constraint was specifically introduced to ensure that participants' working memory would not be disproportionally taxed for some trials in comparison to others when collecting their elicitations in the sentence production task which has not been analyzed for this dissertation (see section 2.2.3 for more details). In the context of the lexical stress perception task, however, participants were presented only with the extracted minimal pair target words, and they did not hear the full sentences.

A total of 58 minimal pairs were recorded and materials for 3 separate tasks were derived from these natural recordings. The materials were processed using the following procedure. Firstly, the target words forming the minimal stress pairs were excised and normalized for amplitude in Adobe Audition 2.0. Only 47 pairs were judged to be unambiguous by a native speaker of Southern British English and those were used for testing in the lexical stress perception task (see section 2.2.4.1). This yielded a total of 94 test words used in the

experiment. For a separate task, one minimal stress pair produced by the male speaker (the lexical stress pair "PROtest" – "proTEST") was used to create synthesized speech stimuli for a lexical stress cue weighting categorization task (presented in section 2.2.4.2).

All word perception stimuli and sentence pairs used in Experiment 1 are listed in APPENDIX A. The supplementary materials also report the word frequencies of the selected pairs, as extracted from the Compleat Lexical Tutor (Cobb, 2021). All words fell into the first 4000-word (K4) families, with the exception of *perfume* (K5), and *intern* and *refund* (K6).

Table 4

Example sentence pairs used for the stimuli recordings.

Prompt sentence	**Response sentence** (target words in bold)
Are they investigating the company?	Yes, for unethical **conduct.**
Is this the new teacher?	Yes, he will **conduct** the training.
Did you pay full price?	No, they gave me a **discount**.
Is this possible?	Yes, I won't **discount** it.
Can I park here?	Yes, but you need a **permit**.
Can I copy the files?	The law does not **permit** it.

2.2.4 Prosodic processing measures

2.2.4.1 Lexical stress perception task

Participants' lexical stress perception abilities were tested in a forced choice identification task (Cooper et al., 2002; Ou, 2010). The test stimuli were 47 minimal stress pairs produced by two speakers (1 male, 1 female).

On each trial, participants first heard a word and then they saw 2 response buttons appear on the screen. The same word appeared on both buttons, but the stress was indicated with capital letters and showed a trochaic stress pattern (always the left button), and an iambic stress position (the right-hand side button).

The task started with two familiarization trials where participants were given feedback (a green check mark appeared on the screen after a correct response, and a red "x" mark appeared if the response was incorrect). The feedback was provided to ensure participants understood the premise of the task and what they were required to do before proceeding to the test trials. Participants heard 94 identification trials (47 minimal pairs) grouped in 4 test blocks. No response accuracy feedback was provided during the testing stage. However, participants could monitor their progress with a progress bar indicating how far along they had moved through the block trials (Figure 3).

Figure 3

Screenshot of a trial taken from the lexical stress perception task. Participants responded by selecting the button representing the stress pattern they thought they heard. The instructions in Simplified Chinese read “Which word did you hear?”

Data analysis

The data collected in the lexical stress perception tasks were analyzed as outlined below. Portion correct values were calculated from each participant’s identification responses by dividing the number of correct responses by the total number of trials.

2.2.4.2 Prosodic cue weighting task

A prosodic cue weighting task was used to measure the relative importance of acoustic cues contributing to lexical stress judgements for native Mandarin listeners (Chrabaszcz et al., 2014; Wang, 2008 for similar paradigms investigating lexical stress cue weighting in Mandarin listeners; and Jasmin et al., 2021; Petrova et al., 2023 probing dimensional weightings in the perception of phrase boundaries). Participants categorized stimuli drawn from a two-dimensional acoustic space which varied in the extent to which pitch, and duration information cued stress on either the first syllable, or the second syllable of the word.

The stimuli set for this task consisted of modified natural recordings of the disyllabic word "protest", produced with stress on the first syllable, as in /PRO-test/, and with stress falling on the second syllable, as in /pro-TEST/, reflecting a shift from a noun to a verb form. The stimuli varied orthogonally along two manipulated dimensions, with different combinations of pitch and durational information extracted from the original recordings. The 2 dimensions either jointly cued a trochaic or an iambic stress pattern on some of the trials, while on other trials they suggested conflicting stress placements. The words in the minimal stress pair were cropped from the following context sentences: "No, he left in *protest*" (noun frame), and "Yes, but we plan to *protest*" (verb frame). For more information on the acquisition of the recordings, refer to section 2.2.3.1 above.

Stimuli creation

All relevant acoustic manipulations were conducted using the STRAIGHT speech analysis and re-synthesis software application (Kawahara & Irino, 2005), which allows for flexible control over the acoustic dimensions of the source signal. The application follows a multi-step standard procedure involving the decomposition of the input signal into its component acoustic characteristics before morphs can be resynthesized. For this task, the noun and verb recordings of "protest" were morphed onto one another by taking the steps outlined below.

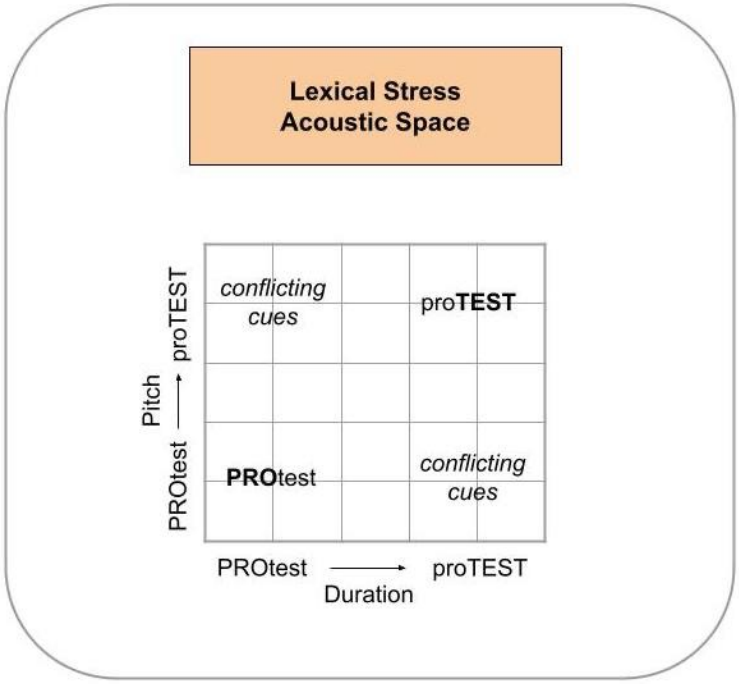
In the first instance, the two recordings were subjected to a source and filter analysis where the speech signal was deconstructed into three components. First, the F0 information contained in the voiced segments of the audio files, was extracted. At this stage, the program performed an automatic recognition of voiced and voiceless segments that was also manually examined and corrected to maximize the quality of the extracted information. Next, the aperiodicity-related aspects of the signal were extracted. As a final step, the software computed the filter parameters of the signal based on the speech spectrogram and the F0 structure. The same procedure was followed for both recordings, and the resulting morphing substrates were saved as separate files.

These files were then used as a basis for manually time-aligning corresponding sections in the frequency and temporal domains of the two recordings. To do this, a distance matrix of the saved substrates was visually inspected for corresponding salient phonemes and syllables, and anchor points were assigned in those places.

The morphed stimuli were then generated in MATLAB (The MathWorks, Inc., Natick, Massachusetts) using a custom script. Manipulations of the fundamental frequency (x 5 morphing levels) and duration (x 5 morphing levels) dimensions were applied, resulting in a two-dimensional acoustic space. The fully crossed combinations of the cues yielded 25 unique tokens representing a morphed continua between the two original recordings. The different morphing levels determined the degree of relative contribution that each recording had in the morphed stimuli, expressed in percentage terms. Contributions of 0% and 100% indicate the manipulated acoustic characteristics were entirely drawn from one of the two endpoints of the continuum, with 0% carrying F0 and duration information from the "PROtest" recording, and the 100% rate deriving its F0 and duration values from the "proTEST" recording. The 25% morphing rate indicated greater acoustic contribution from the trochaic stress recording, while the iambic stress recording contributed to a greater extent in the stimuli morphed at the 75% rate. Lastly, the 50% morphing level included acoustic information drawn to an equal extent from both recordings. Figure 4 can be consulted for a schematic representation of the morphed acoustic space. It is worth noting that these percentage rates relate only to the contributions of the F0 and duration dimensions. All other acoustic dimensions making up the speech signal were held constant at intermediate values, ensuring equal contributions from both stress pattern audio files. Essentially, judgments about the stress location of stimuli could be made only by utilizing information across the two manipulated dimensions (pitch and duration). In some instances, F0 and duration signalled the same interpretation, in others they conflicted, and in certain sections of the acoustic space they were perceptually ambiguous with no correct answers (towards the centre of the acoustic space in Figure 4).

Figure 4

A schematic diagram depicting the "protest" stimulus space created through morphing the minimal lexical stress pair "PROtest"-"proTEST". The manipulated stimulus space was defined along 5 levels of F0, and 5 levels of duration information derived from the original recordings. In some cases, the F0 and duration dimensions cued the same interpretation (cells in the upper right and bottom left corners of the acoustic space), while for other stimuli (cells occupying the upper left and bottom right corners), the two dimensions conflicted resulting in the perception of lexical stress ambiguity.



Procedure

The stimuli were presented in a two-alternative, forced-choice categorization task in which participants had to categorize the words they heard as having stress on the first, or on the second syllable. The task commenced with an explanation of the difference in stress placement between the two versions of the minimal stress pair. Participants then responded to two practice trials where they categorized acoustically unambiguous stimuli (from the endpoints of the stimuli distribution) and feedback was provided on each trial. Participants progressed to the actual test blocks only after logging correct responses in both practice trials. In the testing phase, a word token was played in each trial, and participants had to select one of two buttons, labelled "PROtest" or "proTEST" to indicate if they

perceived stress on the first, or on the second syllable (the stressed syllable was marked orthographically in capitals).

Participants completed 10 blocks (x 25 stimuli) for a total of 250 categorization responses. The trials in each block were randomly presented without any accuracy feedback.

Data analysis

To calculate participants' cue weighting functions, logistic regressions were conducted on individual participants' categorization responses. Response choices were entered as a binary outcome variable, while pitch (x 5 levels) and duration (x 5 levels), were submitted as continuous predictor variables. Next, the logistic regression coefficients were normalized to sum to one as per the following equation (Jasmin et al., 2021):

|Pitch cue weight coefficient|/|pitch cue weight coefficient|+ |duration cue weight coefficient|

The resulting normalized perceptual weight of each dimension ranged between the values of 0 and 1 (Berg, 1989; Christensen & Humes, 1997; Doherty & Turner, 1996; Lutfi, 1992), giving an indication of the relative extent to which participants relied on each acoustic cue when making judgements about the assignment of English lexical stress. For instance, a large, normalized pitch cue weighting reflected greater reliance on the pitch dimension relative to duration.

2.2.5 Auditory processing measures

Three broad auditory processing abilities were tested with the following behavioural measures: 1) auditory acuity to discriminating sounds differing in specified characteristics by collecting auditory discrimination thresholds for three specific acoustic dimensions (pitch, formant, and amplitude risetime), 2) scores for listeners' abilities to selectively attend to specific dimensions (pitch, second formant, and amplitude risetime), and 3) music memory

by assessing participants' abilities to remember and reproduce melodic and rhythmic sequences.

2.2.5.1 Auditory Discrimination Tasks

Participants' psychophysical thresholds were recorded to assess their discrimination sensitivity to three of the key correlates of English lexical stress – pitch, formant, and amplitude rise time (the stimuli were taken from Kachlicka, et al. 2019).

Stimuli creation

The stimuli were generated using MATLAB (The MathWorks, Inc., Natick, Massachusetts), and consisted of 500-ms four-harmonic complex tones with equal amplitude across harmonics, a fundamental frequency of 330 Hz and a 15-ms on-off linear ramp (for the pitch and rise time discrimination tests) and varied along a single dimension. One hundred target stimuli and one base stimulus were constructed for each test dimension. The stimuli for the pitch discrimination test consisted of a baseline tone set at 330 Hz F0, while the target sounds varied along a continuum between 330.3 to 360 Hz in 0.3 Hz steps.

The stimulus set for the amplitude rise time discrimination test had a duration of 500 ms, and F0 of 330 Hz, with a baseline sound of 15 ms, while the targets ranged from 17.85 to 300 ms, in steps of 2.85 ms.

Lastly, the stimuli for testing formant sensitivity were complex tones with 500 ms duration and the following parameters. The F0 of the tones was fixed at 100 Hz, the first formant (F1) was set at 500 Hz, and the third formant (F3) was 2,500 Hz, with harmonics of up to 3,000 Hz. The second formant frequency (F2) was set at 1,500 Hz for the baseline stimulus, and it ranged between 1,502-1,700 Hz in 2-Hz steps for the target stimuli.

Procedure

On each trial, participants heard three complex tones presented with a 500 ms ISI that varied along a single acoustic dimension (the dimension of interest for each specific test). One of the sounds differed from the other two, and the different sound was always either the first or the third sound, with the middle sound being the baseline. Participants indicated their responses by selecting either one of two response buttons numbered “1” and “3” appearing on the screen.

Stimulus presentation was controlled through a three-down, one-up adaptive staircase procedure whereby participants’ performance on a preceding trial determined the stimulus level on the next trial (Levitt, 1971). The full stimulus set contained 100 tones. The default stimulus level for the start of the task was set at step 51 for each participant, and the initial step size by which stimulus level changed after each trial was set at 10 steps. Working this into the adaptive staircase procedure, the task became easier by 10 steps (i.e., reaching stimulus level 61), after an incorrect response on a trial, and the accumulation of 2 correct responses on 2 consecutive trials led to the task becoming more difficult by 10 steps (lowering the stimulus level to 41). After the first reversal, the stimulus step size was reduced to 5 steps. After the second reversal, the stimulus step size was further decreased down to 2 steps, and then to 1 step after the third reversal. The latter step size was kept until the end of the task. The task lasted until 75 stimuli were presented or the program went through seven reversals.

Data analysis

Data pre-processing for this task was performed in MATLAB. A performance score was computed for each acoustic dimension separately, and it was calculated as the mean stimulus level across all reversals from the second reversal until the end of the test. The final score (ranging between 0-100 score points) is, effectively, the discrimination threshold a participant has reached for each specific dimension. The threshold score gives an indication of the minimal difference between the stimuli that participants can discriminate. The scores

can be interpreted as follows: lower scores indicate better sensitivity, or the ability to hear smaller differences in a given dimension, while higher scores suggest poorer sensitivity to differences in that dimension.

2.2.5.2 Dimension selective attention (DSA) task

In this task, participants' abilities to track changes in one dimension and ignore simultaneous changes in another, were tested. Participants were presented with non-verbal stimuli – sequences of complex tones varying in 2 acoustic dimensions at a fixed rate - and they were instructed to listen for repetitions in a specified dimension. The stimuli were modified from Symons, et al. (2021).

Stimuli generation

The auditory stimuli for this task were synthesized in MATLAB and consisted of sequences of complex tones (650 ms long) with 40 harmonics and a 5-ms linear on-off ramp to avoid transients. The tones were created in a three-dimensional acoustic space defined by pitch (F0), formant (single formant at F2), and amplitude rise time. The task itself was set up so that in any given condition, only 2 of the 3 dimensions would vary simultaneously, while the third dimension remained fixed. This resulted in 6 different dimension pairings, presented under 3 test conditions – an attend to pitch condition (where in half of the trials the ignored dimension was formant, and in the other half it was risetime), an attend to formant condition (with pitch and risetime being the ignored dimensions), and finally, an attend to rise time condition (where pitch and formant were the ignored dimensions).

The tones for the sequences were created using the following criteria. The two levels for pitch (F0) as attended (and ignored) dimension were 100 Hz and 118.9207 Hz (separated by 3 semitones), respectively. The values at the F2 frequency were 2040 Hz and 2884.9957 Hz (separated by 6 semitones), and the 2 levels for rise time were 0.005 s and 0.095 s. Finally, pitch and formant when these were stable (the dimension which didn't change) were set as half the distance in semitones between the 2 levels, resulting in 109.0508 Hz for pitch, and

2425.9825 Hz for the formant dimension. The stable value for rise time was computed as the mean of the 2 levels, or 0.05 s, respectively (refer to Table 5 for the dimension values for each level of the three dimensions). The fully crossed combinations of all cues resulted in 12 unique tones (3 attended dimensions x 2 levels x 2 ignored dimensions; note the 3rd dimension was at the stable mid-level value).

Table 5

Stimuli dimensions used for the generation of the tones. Every dimension, pitch, formant (F2), and amplitude rise time could be either an attended dimension (Levels 1 & 2), an ignored dimension (Levels 1 & 2), or a stable dimension (Mid- level).

Stimuli parameters			
Dimension	Level 1	Level 2	Mid-level
Pitch	100 Hz	118.9207 Hz	109.0508 Hz
Formant (F2)	2040 Hz	2884.9957 Hz	2425.9825 Hz
Rise time	0.005 s	0.095 s	0.05 s

Sequence structure

The synthesized tones were concatenated into 2 Hz sequences where the attended and the ignored dimensions varied at different rates (either every three tones, at 0.667 Hz, or every two tones, at 1 Hz). The third, stable dimension was kept constant (held at the mid-level value). Repetitions were inserted into some of the concatenated sequences, in a way that the rate of dimension change was doubled (instead of the tones changing every 2 or 3 notes, they changed every 4, or 6 tones, respectively). The repetitions appeared randomly in the sequences, but 2 repetitions never appeared consecutively. The combinations between attended and ignored dimension, and the presence and absence of repetitions, resulted in 4 trial types: repetition only in the attended dimension, repetition only in the ignored dimension, repetition in both dimensions, and no repetition in any of the dimensions (see Table 6 below for the trial types in the attend to pitch condition). The sequences were blocked into an attend pitch condition, attend formant condition, and attend risetime

condition. Participants heard 16 trials per condition (2 ignored dimensions x 2 rates of tone change x 4 trial types, see below), or a total of 48 trials for the full task.

Table 6

Example trial types for the attend pitch condition. In half of the trials, the pitch rate was changing every 2 tones, and for the other half, the change rate was every 3 tones

Trial types					
		Repetition			
Attended dimension	Pitch	Yes	No	Yes	No
Ignored dimension	Formant	No	Yes	Yes	No
Attended dimension	Pitch	Yes	No	Yes	No
Ignored dimension	Rise time	No	Yes	Yes	No

Counterbalancing was achieved by creating 2 versions of the same task containing identical test stimuli but where these stimuli were assigned to the opposite attention condition – i.e., the focus of attention differed in the two versions. For instance, the same sequence varying in pitch and formant, would have pitch as the attended dimension in 1 of the versions, and formant as the attended dimension in the other version of the task. Both versions presented the same number of trials and conditions.

Training and testing procedure

At the beginning of the task participants were given detailed instructions guiding them through the type and nature of changes in the three dimensions. The dimensions were defined in terms of the quality characteristics of the sound. Pitch was described as how "low" or "high" the sound was. The formant dimension was described as "brightness", and risetime was described in terms of "how fast the sound reaches its maximum peak". Participants heard a two-tone sequence for each dimension (pitch, formant, and risetime) where only the target dimension was changing and the other two dimensions remained constant, exemplifying the different levels of pitch, formant and risetime that they would encounter in the task. They also heard concatenated single dimension sequences where

only the target dimension varied at different rates. Participants also had the opportunity to hear stimuli with and without repetitions.

Before proceeding to the test trials, participants completed 3 blocks of training for each dimension, consisting of single dimension sequences. For each block instructions specified the attended dimension and its rate of change that would apply for that block. The instructions remained on the screen for the duration of the trials, and participants received feedback on their response accuracy. Eight (8) trials were presented per attended condition (2 rates of tone change x 2 ignored dimensions x 2 trial types (presence and absence of repetition)). The training blocks were randomized across participants.

The main task consisted of 6 blocks presented in a random order (3 attention conditions each presented with 2 rates of tone change). At the start of each block, participants received instructions about the attended dimension for that block and the rate of change. The attended dimension and the rate of change remained visible on the screen for the duration of the block as a reminder of which dimension had to be tracked. Each trial began with a 500 ms silence followed by the presentation of the stimulus. Participants were asked to listen to the sequence and respond whether they had detected a repetition in the attended dimension or not. Responses were made by selecting either a “Yes”, or a “No” button presented on the screen. Feedback accuracy was provided after each response in the form of a green check mark for correct responses, and a red “x” mark for incorrect responses. Participants could take a short break between blocks.

Data analysis

Performance on this task was measured by calculating response accuracy for each attended dimension separately. Portion correct was computed as the number of correct responses per condition, divided by the total number of trials in that condition.

2.2.5.3 Music memory tasks

Participants' abilities to reproduce musical sequences were tested through the use of two music memory tasks (Tierney, et al. 2017, Sun et al. 2021). Both tasks required participants to listen to and remember a series of melodic and rhythmic patterns and then use either their computer keyboard or screen buttons to enter the patterns they remembered. The melodic memory task assessed memory for melodies, and by extension participants' spectral integration abilities, while the rhythmic memory task assessed memory for rhythms, or participants' temporal integration abilities.

Melodic memory task

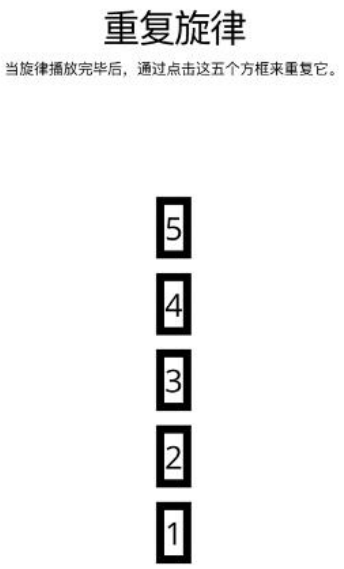
The stimuli for this task were seven-tone melodic patterns formed from the first five notes of the A major scale (specifically, 300 ms notes carrying the following frequencies: 220, 246.9, 277.2, 311.1, and 329.6 Hz). Each note contained 6 harmonics with equal amplitude across harmonics and had a 50-ms on-off cosine ramp applied to avoid transients. Test melodies were generated using the procedure outlined below. The first note in a stimulus track always contained the third tone (277.2 Hz). The next note was randomly selected to be either lower on the pitch scale, or higher on the pitch scale. Five more iterations of the same procedure followed until a 7-note sequence was constructed. If either one of the scale limits was reached (a note of 220 or 329.6 Hz), the next note was selected to either be closer to the centre of the scale or identical to the previous note.

Participants completed a total of 10 melody reproduction trials. On each trial, they heard one melody repeated three times with 1000 ms interval between repetitions. Following the three melody repetitions, a vertical array of 5 boxes (numbered 5-1) each associated with a tone, was displayed on the screen (shown in Figure 5). By clicking on the boxes, which changed colour when selected, participants could hear the tones play. The middle box (Box 3) corresponded to the middle pitch value (i.e., 277.2 Hz) which was also the first note of each melody. The boxes with higher number (4 and 5) corresponded to notes higher on the scale, and boxes 1 and 2 played the lower notes.

Participants were instructed to reproduce the melody by clicking on one box at a time so that they matched the notes to the melody they heard. Before proceeding to the test trials, participants practiced clicking through the boxes and they did a practice trial.

Figure 5

Screenshot of the group of 5 note boxes presented in a vertical layout. The boxes corresponded to the following frequencies in ascending order 220, 246.9, 277.2, 311.1, and 329.6 Hz. Participants were informed that each melody started with the 3rd note. They had to reproduce the melody from memory by clicking one box at a time. The instructions in Simplified Chinese read "Repeat the melody. When the melody is finished playing, repeat it by clicking on these five boxes."



Data analysis

The data collected in this task was pre-processed in MATLAB. Response accuracy was computed by processing participants' responses on a trial-by-trial basis by making a 1-to-1 comparison between the notes logged by the participant, and the notes in each target trial. Responses where the logged response vector exceeded the length of the stimulus vector

were excluded from the analysis. A final percentage score was calculated for each participant reflecting the mean response accuracy ratio of the 10 trials.

Rhythmic memory task

The stimulus set for the rhythmic memory task consisted of 10 rhythmic patterns, with 5 taken from the weakly metrical list and 5 from the strongly metrical list to ensure a range of difficulty across the rhythmic patterns (Povel & Essens, 1985). The rhythmic patterns were made up of 16 segments (each 200 ms long) in a combination of drum hits and segments of rest. Nine segments contained a drum hit, while the remaining were rest segments. Drum hits consisted of a 150-ms conga drum sound obtained from freesound.org. On each trial, the rhythm was repeated three times with a 600 ms interval between repetitions. Participants had to listen to the three repetitions and then use their keyboards to recreate the rhythmic sequence they had heard by pressing the space bar. Key presses corresponded to drum hits while segments of rest were represented by longer or shorter rests between spacebar presses.

Data analysis

The response time of each spacebar press was recorded, and it was later compared to the pattern of drum hit segments in each sequence track. The inter-response intervals (IRIs) were converted to the nearest multiple of 200 ms (e.g., 200, 400, 600, 800 ms). Response accuracy was calculated for each participant by making segment-by-segment comparisons between the pattern of drum hits and rest segments logged by that participant on each trial and the corresponding pattern in the target track. A final percentage score was computed for each participant.

2.2.6 Cognitive and language proficiency measures

2.2.6.1 Language proficiency test - LexTALE

A short standardized yes/no vocabulary test was used as a measure of English language proficiency. The Lexical Test for Advanced Learners of English (LexTALE, www.lextale.com) is a lexical decision task validated as a reliable predictor of vocabulary competence and shown to have high correlation with other measures of general language proficiency (Lemhöfer & Broersma, 2012). However, some limitations of the LexTALE test have been identified in more recent research. For instance, Puig-Mayenco et al. (2023) found weaker correlations between LexTALE scores and standardized measures of L2 proficiency, such as the Quick Placement Test (QPT, UCLES, 2001). Specifically, LexTALE appears to be more reliable for assessing proficiency in advanced learners, whereas for lower proficiency learners, correlations between LexTALE and QPT scores tend to be low or non-significant. Despite these limitations, LexTALE was chosen for the current experiment due to its practicality and short administration time (approximately 5 minutes), making it a more feasible option than longer tests such as the QPT, which typically require up to 30 minutes to complete.
The test contained 63 vocabulary items, of which 40 were low-frequency real words, while the remaining 20 items were pseudowords conforming to English phonological and phonotactic rules. Participants first completed 3 dummy trials (2 real words, 1 nonword). All test items are listed in APPENDIX B. On each trial participants saw a word on the screen, and they had to indicate whether they recognized it as an existing English word or not. Participants completed 63 trials in total. The final analysis excluded the responses from the first 3 dummy trials, resulting in a final total of 60 trials.

Accuracy scoring

Two types of items were tested, words and nonwords. Accuracy scores, therefore, were calculated by averaging the portion correct responses for each item type, correcting for the unequal number of words and nonwords using the following equation:

Portion correct = ((number of correct words/40) + (number of correct nonwords/20)) / 2

The portion correct scores were interpreted in correspondence to the Common European Framework (CEFR) levels (as suggested by Lemhöfer & Broersma, 2012; see Table 7).

Table 7

LexTALE score ranges and their rough correspondence to CEFR language proficiency levels. Taken from Lemhöfer and Broersma, 2012.

CEFR Level	CEFR Description	LexTALE score
C1 & C2	Upper & lower advanced/proficient	80 - 100 %
B2	Upper intermediate	60 - 80 %
B1 & lower	Lower intermediate and lower	< 59%

2.2.6.2 Digit span task

Participants' working memory abilities were tested using a visual backward digit span task (Olsthoorn et al., 2014). The task was selected as previous research has shown that numerical items can reliably measure verbal working memory capacity (Daneman & Merikle, 1996. Additionally, the use of the visual modality was deemed more appropriate for measuring working memory in L2 learners, as visually presented stimuli are more likely to be processed in the learners' dominant language. This approach may provide a more accurate reflection of their working memory capacity by minimizing potential interference from L2 language proficiency (Olsthoorn et al., 2014). Participants were presented with sequences of numerical digits and were instructed to repeat them in reverse order (backwards) using a numeric pad provided on the screen. On each trial, the digit sequences were presented visually item by item with each digit displayed for 500 ms, followed by a 250-ms fixation cross. Participants completed a total of 8 trials with 2 trials per digit length. At the start, they were presented with 2 trials of 2 digits, and the length of the sequences increased by 1

digit every 2 trials. The number of digits entered on each trial was controlled by showing empty data entry fields on the screen that matched the number of digits in each trial. The highest number of digits the task reached was 5. The scores for this task were calculated as the total number of trials correct. The data from this task was collected for the purposes of controlling for potential working memory effects on performance in one of the lexical stress training tasks and as such will not be discussed in this chapter. Refer to section 3.2.3.3 for more details.

2.3 Results

Unless specified otherwise, all data pre-processing and subsequent statistical analyses reported in this dissertation were performed using R (R Core Team, 2023). All data visualization graphics were created with the ggplot2 package in R (Wickham, 2016).

According to Kolmogorov-Smirnov tests of normality, participants' lexical stress identification scores did not show significant deviation from the normal distribution (D = 0.10, p = 0.258). Their formant and risetime discrimination thresholds were comparable to a normal distribution (D = .046, .054, p = .980, .920), as were their rhythm and melody memory scores (D = .064, .136, p = .783, .045), and their age of acquisition (D = 132, p = .056) However, participants' weekly English use, their pitch discrimination thresholds, as well as their dimension selective attention scores differed significantly from the normal distribution (D = .168, p < .01 for weekly English use, D = .201, p < .001 for pitch discrimination, and D = .159, .144, .181, p = .011, .028, .002, for formant, pitch, and risetime dimension selective attention, respectively). These variables were submitted to transformations with dimension selective attention scores for formant and pitch undergoing arcsine transformations (D = .103, p = .22, for dimension selective attention to formant, and D = .089, p = .39 for dimension selective attention to pitch) and dimension selective attention for risetime and pitch discrimination thresholds, as well as weekly English use were transformed with a log10 function (D = .14, p = .03 for dimension selective attention to risetime, D = .04, p = 96 for pitch discrimination, and D = .085, p = .45).

Auditory processing abilities and lexical stress perception

To investigate the extent to which domain general auditory processing abilities can explain variability in lexical stress perception, participants' auditory discrimination thresholds (for pitch, formant, and risetime), their dimension selective attention scores (for pitch, formant, and amplitude risetime), their music memory scores (for rhythmic memory, and melodic memory), and demographic measures (age of acquisition, pronunciation training, and weekly English use) were submitted as potential predictors in a multiple linear regression analysis. Pronunciation training was entered as a categorical variable with two levels, 1 for participants who reported receiving pronunciation training, and 0 for those who had no pronunciation training. Participants' lexical stress perception scores were entered as the outcome variable. This model was significant, $F(11, 90) = 5.94$, $p < .001$, $R^2 = .35$ (see Table 8 below). Variance inflation factors for all 11 predictor variables were within acceptable ranges (1.08 - 1.78) and did not yield evidence for multicollinearity. In this model, pitch discrimination abilities emerged as a significant predictor ($t = -2.62$, $p = .01$, Figure 6) indicating that learners' with lower discrimination thresholds (or higher discrimination sensitivity), showed better English lexical stress perception. Additionally, melody memory ($t = 1.75$, $p = 0.08$), and pronunciation training ($t = 1.83$, $p = 0.07$) were also marginally significant predictors. The remaining discrimination (formant, risetime), dimension selective attention (formant, pitch, risetime), rhythm memory, and demographic measures were not predictive of lexical stress performance.

Table 8

Multiple regression model predicting lexical stress perception from a number of potential predictors (auditory discrimination (formant, pitch, risetime), dimension selective attention (formant, pitch, risetime), auditory-motor integration (rhythm memory, melody memory), and demographics (age of acquisition, pronunciation training, English use)). Pitch

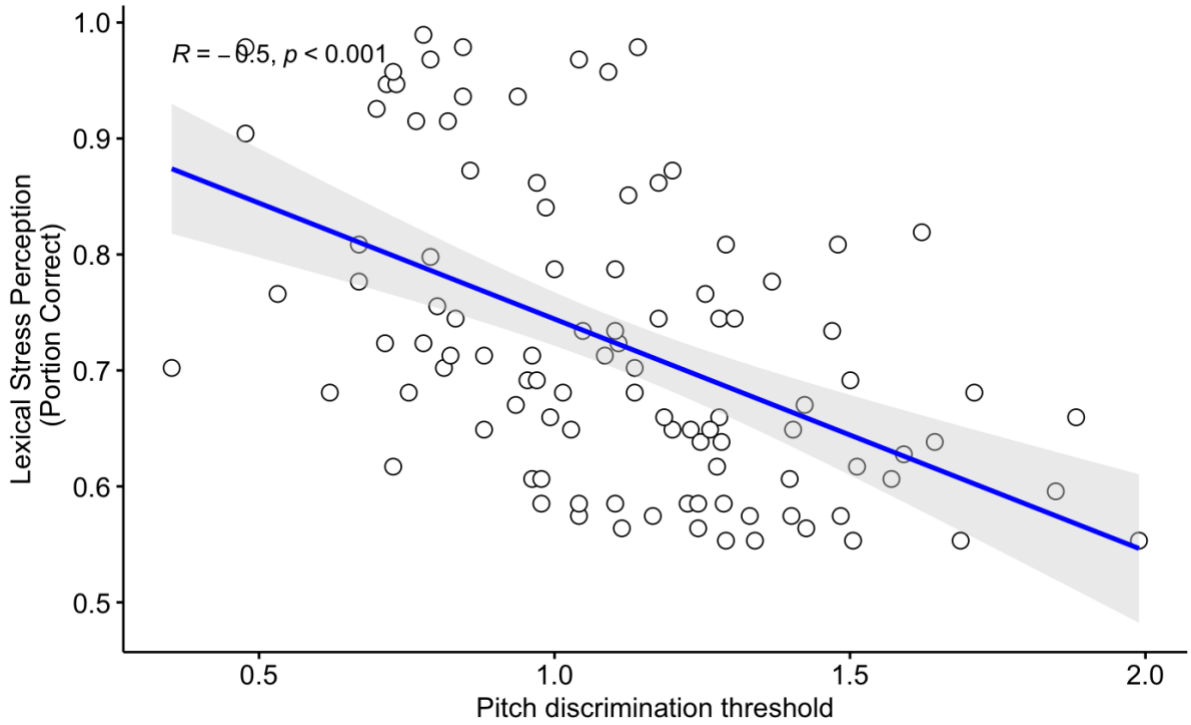
discrimination emerged as a significant predictor of lexical stress perception for Mandarin learners.

Outcome variable	Predictor variables	Estimate	*SE*	95% CI		*t*	*p*
				LL	*UL*		
Lexical stress perception	Formant auditory discrimination thresholds	0.00	0.00	0.00	0.00	0.51	0.61
	Pitch auditory discrimination thresholds	-0.11	0.04	-0.19	-0.03	-2.62	**0.01**
	Risetime auditory discrimination thresholds	0.00	0.00	0.00	0.00	-1.45	0.15
	Formant dimension selective attention	0.02	0.05	-0.08	0.13	0.48	0.63
	Pitch dimension selective attention	0.07	0.07	-0.03	0.17	1.30	0.20
	Risetime dimension selective attention	-0.01	0.10	-0.22	0.20	-0.11	0.91
	Rhythm memory	0.00	0.00	0.00	0.00	1.23	0.22
	Melody memory	0.11	0.06	-0.01	0.23	1.75	0.08
	Age of acquisition	0.01	0.00	0.00	0.01	1.64	0.10
	Pronunciation training	0.05	0.03	0.00	0.10	1.83	0.07
	Weekly English use	0.01	0.03	-0.05	0.06	0.23	0.82

Note. SE - standard error, CI = confidence interval, *LL* - lower limit, *UL* - upper limit, *t* - t value. Statistically significant *p* values are bolded.

Figure 6

Negative linear relationship between pitch discrimination abilities and lexical stress perception scores. Participants with better discrimination acuity for pitch showed better English lexical stress perception.



Upon closer examination of the data and in the interests of confirming the robustness of the model, the raw data of participants identified as outliers in the auditory discrimination tests (defined as performance that was more than 2 standard deviations away from the sample mean) were examined manually and those who showed anomalous response behaviour (i.e., consistent evidence for misinterpreting the task instructions), were excluded (N = 1). Additionally, participants who failed to perform above chance level (< .55) on any of the three auditory selective attention tasks were also excluded (N = 5). As such, 96 participants were included in this more stringent analysis, which produced a very similar pattern of results, with auditory pitch discrimination abilities identified as a significant predictor of lexical stress perception. The complete model output is provided in APPENDIX C.

Auditory processing abilities and lexical stress cue weighting strategies

To explore the possible role of domain-general auditory processing in shaping listeners' dimensional weighting strategies, a second multiple regression analysis was performed with participants' dimension selective attention scores (for pitch, formant, and amplitude risetime), auditory discrimination thresholds (for pitch, formant, and amplitude risetime), and music memory scores (rhythmic and melodic memory) entered as predictors, and their normalized pitch cue weights entered as the dependent variable. Aiming to improve the quality of the data, this analysis excluded spurious data points (participants who did not show a significant relationship ($p < .05$) between at least one of the stimulus dimensions (pitch or duration) and their categorization decisions in the lexical stress cue weighting task). This resulted in a total of 95 participants' data included at this stage of the analysis.

The entered model was significant, $F(8, 86) = 3.21$, $p < .001$, $R^2 = .158$. The variance inflation factors for all 8 predictor variables were within acceptable ranges (1.20 - 1.68) and did not yield evidence for multicollinearity. Rhythm memory scores, pitch discrimination thresholds, and dimension selective attention for risetime emerged as significant predictors of participants' reliance on pitch in their lexical stress dimensional strategies. All three predictors inversely correlated with reliance on pitch (Table 9).

Table 9

Multiple regression model predicting pitch cue weights in lexical stress perception from a number of potential predictors (auditory discrimination (formant, pitch, risetime), dimension selective attention (formant, pitch, risetime), and audio-motor integration (rhythm memory, melody memory). Three variables emerged as significant predictors for participants' reliance on pitch information in English lexical stress processing: rhythm memory abilities, pitch auditory discrimination thresholds, and dimension selective attention for risetime.

	Predictor variables	*SE*	95% CI	*t*	*p*

Outcome variable		Estimate		*LL*	*UL*		
Pitch cue weights	Formant auditory discrimination thresholds	0.00	0.00	0.00	0.00	-0.69	0.49
	Pitch auditory discrimination thresholds	-0.20	0.07	-0.33	-0.06	-2.90	**0.01**
	Risetime auditory discrimination thresholds	0.00	0.00	0.00	0.00	-1.17	0.25
	Formant dimension selective attention	-0.12	0.09	-0.29	0.06	-1.29	0.20
	Pitch dimension selective attention	0.08	0.08	-0.09	0.25	0.96	0.34
	Risetime dimension selective attention	-0.35	0.18	-0.70	0.00	-2.01	**0.047**
	Rhythm memory	-0.01	0.00	-0.01	0.00	-3.55	**<0.001**
	Melody memory	0.03	0.10	-0.17	0.23	0.28	0.78

Note. SE - standard error, CI = confidence interval, *LL* - lower limit, *UL* - upper limit, *t* - t value. Statistically significant *p* values are bolded.

Pitch discrimination abilities negatively correlated with reliance on pitch information in lexical stress categorizations (t = -2.90, p < .01), indicating that learners' with lower discrimination thresholds (or higher discrimination sensitivity) to pitch, relied more heavily on pitch information when categorizing English lexical stress contrasts. A significant negative relationship was also observed between memory for rhythmic patterns and reliance on pitch in categorization responses (t = -3.55, p < .001), suggesting that participants with excellent rhythmic memory relied more heavily on duration information in their lexical stress cue weighting strategies. Lastly, selective attention to the amplitude risetime dimension was also a significant predictor of the utilization of pitch information in lexical

stress cue weighting (t = -2.01, p = .047). This relationship suggests that participants who are better able to attend to amplitude rely more on the duration dimension in lexical stress categorizations. The remaining discrimination (formant, risetime), dimension selective attention (formant, pitch), and melodic memory scores were not predictive of dimensional weighting strategies.

2.4 Discussion

In Experiment 1 I set out to investigate if individual differences in performance on a battery of domain-general auditory measures, could explain variability in adult L2 learners' prosodic acquisition. I tested 102 native Mandarin learners of English on their lexical stress acquisition and asked if individual differences in domain-general auditory processing abilities, conceptualized as auditory discrimination acuity, dimension selective attention and auditory-motor integration, could account for some of the variability in their lexical stress processing.

Relationship between auditory processing abilities and L2 lexical stress perception

In a regression model including measures of domain-general auditory processing, pitch discrimination ability emerged as a significant predictor of Mandarin speakers' lexical stress perception. Learners with lower pitch discrimination thresholds, or better discrimination acuity also had better lexical stress perception. This relationship is in line with studies that have stressed the importance of pitch in signalling lexical stress, arguing that pitch is in fact the primary acoustic cue correlating with lexical stress in English (Beckman, 1986; Bolinger, 1958; Fry, 1955). Pitch, or fundamental frequency variations, have consistently been shown to affect the perception of stressed versus unstressed syllables to a greater extent than changes along other dimensions such as duration or amplitude (Fry, 1958; Lieberman, 1960; Morton & Jassem, 1965).

However, there is a lack of consensus in the literature about the degree of informativeness that the different acoustic cues bring to lexical stress categorization. Some authors have even proposed that pitch is the least useful dimension for signalling lexical stress (Beckman & Edwards, 1994; Sluijter & van Heuven, 1996). This claim is raised primarily on account of the complex interplay between segmentals and suprasegmentals (Chen, 1970; House, 1961; Peterson & Lehiste, 1960) on the one hand, and word-level versus phrase- and sentential-level prosody (Ladefoged, 2014; Ou, 2016; Sluijter & van Heuven, 1996) on the other, which makes it challenging to draw firm conclusions about the prominence of acoustic cues in lexical stress processing. And while this is a valid consideration, the findings from the current study offer strong evidence in support of the importance of auditory sensitivity to pitch for the superior processing of lexical stress.

The results of Experiment 1, also hinted at a possible relationship between melody memory, a subtest of the auditory-motor integration battery, and lexical stress perception. While melody memory scores did not reach statistical significance, there was a tendency for participants who were more accurate in their reproductions of melodic sequences to identify lexical stress better. The melodic reproduction task tested participants' ability to remember and reproduce patterns of complex tones varying in pitch (Saito et al., 2021). Whereas the relationship was marginally significant (at p = .08), it lends further support to the importance of pitch processing abilities in explaining variability in Mandarin learners' lexical stress performance.

Finally, pronunciation training missed statistical significance (p = .07) but there was an indication in the data that receiving pronunciation training was associated with better scores in learners' lexical stress identifications. Some of the participants who received pronunciation training also reported receiving lexical stress-specific instruction. This trend in the data is in line with previous studies reporting on the benefits of more holistic classroom-based pronunciation instruction going beyond the segmental level to include suprasegmental phonology (Derwing et al., 1998; Derwing & Rossiter, 2003; Moyer, 1999; Zhang & Yuan, 2020).

Relationship between auditory processing abilities and lexical stress cue weighting strategies

In a prosodic cue weighting task Mandarin listeners had to categorize a disyllabic word as either having stress on the first, or on the second syllable. Two acoustic dimensions, pitch and duration, were manipulated orthogonally so that both dimensions cued lexical stress to varying degrees. This gave a measure of the extent to which listeners' lexical stress judgements were influenced by pitch and duration.

Based on the analysis presented in the results section, three auditory processing abilities were predictive of lexical stress categorization behaviour. I found that participants with lower pitch discrimination thresholds, indicating better sensitivity for pitch, also tended to rely more on the pitch dimension (relative to duration), when perceiving lexical stress. This is consistent with evidence suggesting that individual differences in listeners' abilities to perceive acoustic dimensions are related to the cue weightings assigned to those dimensions (Jasmin et al., 2020). In their study, Jasmin et al. (2020) found that acoustic information which is better perceived was also upweighted in perceptual judgements.

Two auditory measures related to temporal processing were also significant predictors of lexical stress cue weighting, namely rhythm memory scores and dimension selective attention for risetime. In the present context, the rhythm memory test assessed individuals' abilities to integrate acoustic information across time (Tierney et al., 2017). Participants with better memory for rhythm relied more heavily on duration for their lexical stress judgements. The third auditory processing ability which showed a significant negative relationship with pitch weighting in stress categorization behaviour, was dimension selective attention for risetime. Listeners who were better able to direct their attention to risetime amplitude and away from simultaneous changes in other dimensions, tended to rely more heavily on duration to classify lexical stress. Referring back to the attentional models of speech processing (Francis & Nusbaum, 2002; Gordon et al., 1993; Holt et al., 2018) reviewed earlier, the results from this experiment point to the existence of a link between the perceptual salience of a dimension (here operationalized in terms of the ability to direct

attention to risetime and ignore irrelevant changes in other dimensions) and reliance on that dimension. The data here suggests that better dimension-selective attention ability for a temporal cue (risetime) was associated with increased weighting of duration in prosodic categorization. In combination, these findings speak to an overall tendency for individuals with better temporal processing abilities to also weight durational information more heavily. This lends further support to the theories that selective attentional processes might underlie the integration of acoustic information from different dimensions (Francis & Nusbaum, 2002; Gordon et al., 1993; Holt et al., 2018), whereby the ability to selectively focus attention on a specific dimension might affect the perceptual weighting assigned to that dimension so that more perceptually salient information is also weighted more heavily.

Chapter 3 Experiment 2: High-variability prosodic training for teaching English lexical stress to native Mandarin speakers (Longitudinal investigation)

3.1 Introduction

Experiment 1 presented in Chapter 2 was a cross-sectional study of the relationship between individual differences in Mandarin speakers' domain-general auditory processing abilities and variability in their English lexical stress performance (in terms of perception and cue weighting strategies). The next phase of this dissertation project took a longitudinal approach to the investigation of lexical stress acquisition in the same population. The broad purpose of Experiment 2 was twofold: to design and test the effectiveness of a novel lexical stress perceptual training paradigm, and to explore the possible effects of domain-general auditory processing abilities on learners' lexical stress processing outcomes after receiving short-term online perceptual training.

The current study aimed to fill a gap in the phonetic training literature by exploring the efficacy of the high variability phonetic training (HVPT) method in a new prosodic learning context – that of the acquisition of L2 lexical stress. The HVPT training approach involves intensive exposure to target contrasts produced by multiple talkers and appearing in multiple lexical and phonetic contexts (Lively et al., 1993; Logan et al., 1991) and has successfully been used in the training of both segmentals (Bradlow et al., 1997; Iverson & Evans, 2009; Lambacher et al., 2005; Lively et al., 1993; Logan et al., 1991; Nishi & Kewley-Port, 2007) and some suprasegmentals (Perrachione et al., 2011; Sadakata & McQueen, 2014; Wang et al., 1999) . Experiment 2 of this dissertation, therefore, trained native Mandarin speakers on English lexical stress perception. As outlined in the literature review (Chapter 1), Mandarin Chinese and English have typologically different prosodic systems. While English is an intonation language using word-level stress to convey lexical distinctions,

Mandarin Chinese is a tone language encoding lexical differences by changing the tone of a given syllable. Consequently, the processing of English lexical stress and controlling the acoustic cues that covary with stress in a native-like manner frequently prove problematic for native Mandarin learners (Archibald, 1997; Chen et al., 2001; Hung, 1993; Juffs, 1990; Zhang et al., 2008).

The high variability training approach was chosen as the training method for this study as it has proven highly effective in improving non-native segmental acquisition (see Barriusu & Hayes-Harb, 2018; and Thomson, 2018 for comprehensive reviews of the HVPT technique). Over two decades of research have shown that the HVPT training paradigm is an effective, flexible and relatively simple perceptual intervention that can be implemented in a variety of contexts. The highly variable training stimuli (incorporating both phonetic and lexical variability on the one hand, and talker variability on the other) are reflective of real-world listening conditions which is a crucial component for promoting learning by exposing listeners to naturally occurring variability which is lacking in other types of training where low variability stimuli are used (Logan et al., 1991; Lively et al., 1993). And while studies over the years have introduced numerous variations adding greatly to the original HVPT paradigm (Logan et al., 1991; Lively et al., 1993), the core tenets of the methodology including high variability training input, forced-choice perceptual training tasks, and immediate corrective feedback have been consistently shown to enhance learning for a number of target language sounds. In general, reported improvements extend beyond the trained context (to novel items and novel talkers), can lead to positive transfer to production abilities (Bradlow et al., 1997; 1999; Iverson et al., 2012), and learning gains from training are frequently retained over time (Lively et al., 1994; Bradlow et al., 1999; Flege, 1995b; Iverson & Evans, 2009; Thomson, 2012). For the most part, however, empirical interest in the method has almost singularly centred around teaching consonant and vowel perception (Bradlow et al, 1999; Bradlow et al, 1997; Iverson et al, 2003; Iverson & Evans, 2009; Logan et al., 1999; Lively et al, 1993; Nishi & Kewley-Port, 2007; Rato & Rauber, 2015; Shinohara & Iverson, 2018). The one notable exception in the use of HVPT training beyond the segmental context has been the acquisition of Mandarin tones by non-tonal, or other tonal language speakers. This research has provided support for the efficacy of HVPT in a suprasegmental context, finding it offers similar benefits to learners as those documented in

segmental studies (Sadakata & McQueen, 2014; Wang et al, 1999; Zhang et al, 2018). And while a recent study has also explored the possible HVPT training effects on L2 learners' lexical stress cue weighting strategies (Tremblay et al., 2023), and found that the training was effective for shifting perceptual weightings, there have been no attempts to investigate the efficacy of HVPT for improving L2 lexical stress perception.

However, this is an important area of research and deserves more attention since L2 suprasegmental acquisition plays a prominent role in L2 English comprehensibility and intelligibility outcomes (Gallego, 1990; Isaacs & Trofimovich, 2012; Kang, 2010; Kang et al., 2010; Munro & Derwing, 1995). A case in point is that prosodic performance has actually been shown to predict oral proficiency scores at internationally recognized proficiency exams (Kang & Johnson, 2018). These findings bring into sharp focus the need to develop effective targeted phonetic training methods that will improve L2 learners' suprasegmental abilities. To this end, and building on the well-attested strengths of this paradigm, the training study presented here aimed to enhance the prosodic performance of native Mandarin learners of English by adapting the HVPT training technique to the acquisition of a previously untested prosodic context - that of the perception of lexical stress. In the present study I explored the trajectory and scope of improvements in English lexical stress processing for a group of Mandarin speakers who received high-variability prosodic training, compared to a control group of Mandarin speakers who were trained on English vocabulary. I also investigated the possible sources of individual differences in learner outcomes by probing the relationship between participants' performance on a battery of auditory processing tests administered pre-training and their learning outcomes post-training. Chapter 3 of this dissertation, therefore, combines data presented in Chapter 2 (Experiment 1) where baseline measures of participants' lexical stress perception, cue weighting and auditory processing abilities were collected, with data from Experiment 2, namely the high variability phonetic training intervention, and post-test testing.

The prosodic training developed for the purposes of this study adopted the well-established key elements of successful HVPT training designs (Logan et al., 1991; Lively et al., 1993). Participants were exposed to natural recordings of non-word stimuli (Ortega et al., 2021; Thomson & Derwing, 2016) featuring multiple phonetic environments and voices,

embedded in forced choice judgement tasks and presented with corrective feedback. In the control training condition, participants received visual-only vocabulary training for the full duration of the training.

The training study explored the following research questions:

1) Will high variability phonetic training (relative to the control vocabulary-based training) be effective in improving Mandarin learners' lexical stress perception measured at post-test and compared to their performance at pre-test (Experiment 1)?

2) Will short-term HVPT training lead to a measurable shift in participants' underlying cue weighting strategies used for categorizing lexical stress (i.e., their relative reliance on pitch versus duration information in lexical stress perceptual judgements)?

3) Can individual differences in the auditory processing measures collected at Time 1 (pre-test) help predict the observed learning gains in lexical stress perception at Time 2 (post-test)?

Based on the literature demonstrating the effectiveness of the HVPT paradigm, it was predicted that the high variability prosodic training used in this experiment would facilitate L2 lexical stress acquisition and native Mandarin learners were expected to show learning gains in lexical stress perception. More specifically, I predicted that participants trained in the experimental group would show greater improvement on their lexical stress perception scores from pre-test to post-test compared to participants trained in the control group. The generalization of learning was assessed through the use of real word pairs produced by 2 novel talkers not encountered during training.

Next, I set out to investigate if Mandarin participants in the experimental training group would undergo changes in their cue weighting strategies. In the present study participants' cue weighting strategies were assessed based on the weights they assign to pitch relative to

durational information when categorizing a lexical stress contrast (using the morphed versions of a lexical stress minimal pair in which the pitch and duration information were manipulated to vary in the extent to which they cued lexical stress on the first or on the second syllable). The second research question, therefore, seeks to investigate if short-term phonetic training would alter participants' cue weights. However, since there is a lack of agreement in the existing literature with respect to the hierarchy of importance and relative weighting of the four acoustic correlates of English lexical stress, this research question has an exploratory nature with respect to the direction of a potential change in perceptual strategies.

While some authors have proposed that pitch is, in fact, the primary acoustic cue for categorizing lexical stress (Beckman, 1986; Bolinger, 1958; Fry, 1958; Lieberman, 1960; Morton & Jassem, 1965), others have highlighted the fact that lexical, phrase and sentential stress are often confounded (Beckman & Edwards, 1994; Sluijter & van Heuven, 1996), making pitch a less reliable cue. There is also no strong consensus about the ranking of duration in signalling lexical stress, which while considered an important cue (Adams & Munro, 1978; Taylor, 1981), is frequently affected by vowel type and surrounding consonants (House, 1961; Peterson & Lehiste, 1960). Given the contradictory evidence outlined so far, I had no strong predictions about the direction of change in Mandarin speakers' reliance on pitch and duration.

On the one hand, prior research has shown that native Mandarin speakers attach greater weight to F0 in English lexical stress perception compared to native English speakers (Wang, 2008). It is, therefore, possible that Mandarin learners would down-weight reliance on F0 when categorizing stress, effectively acquiring more native-like perceptual weighting strategies (Tremblay et al., 2023). However, it is also possible that Mandarin speakers through being exposed to more naturalistically variable stimuli, would instead upweight their reliance on pitch-related information, thus leaning more into their enhanced pitch processing abilities (Bidelman et al., 2013; Giuliano et al., 2011; Deroche et al., 2000; but see Bent et al., 2006; Peretz et al., 2011) to try and maximize their perceptual performance. Similar adaptations have been observed, when for instance due to the idiosyncrasies of the

perceptual system, individuals learn to rely more on cues they have historically been able to perceive better (Jasmin et al., 2019).

Finally, participants' auditory processing abilities recorded before the training were hypothesized to have predictive power for their stress processing outcomes at the post-test. A number of studies, for instance, have demonstrated that individual differences in auditory acuity are related to learning improvements from perceptual training (Hazan & Kim, 2010; Lengeris & Hazan, 2010; Wong & Perrachione, 2007), and to L2 speech acquisition in immersive settings (Sun et al., 2021). For the context of this study, I hypothesised that individual differences in Mandarin learners' auditory processing abilities collected at Time 1 could predict their learning gains at Time 2.

An additional measure that was collected at Time 1 was the digit span task which assessed participants' working memory abilities. These scores were used in Experiment 2 to control for possible working memory effects on training performance in the category discrimination task. The testing procedure for this task involved listening to 3 different non-words all produced by different talkers and deciding which of the 3 non-words had a stress placement different to that of the other 2 tokens. Identifying the odd stress pattern, however, requires holding each token in memory while the next is being played, which may signify an increased memory load for participants (Strange & Shafer, 2008). To investigate the possibility of working memory capacity impacting training performance for this task, I collected participants' digit span scores at Time 1 (Olsthoorn et al., 2014).

These research questions were explored in a short intensive high variability prosodic training study aimed at improving native Mandarin learners' acquisition of English lexical stress. Participants' learning gains were measured through a perception test. The perceptual intervention was distributed over 2 weeks and consisted of 6 training sessions amounting to 3 hours of speech training in total (approximately 30 minutes of training per session). Participants were tested on 2 separate occasions (before and after training) on their lexical stress perception and cue weighting abilities. Their auditory processing and working memory abilities were collected only before the training (in Experiment 1).

3.2 Methods

3.2.1 Participants

As reported in Chapter 2, 132 native Mandarin learners of English living in China completed Experiment 1. Of these, 102 participants signed up for the training and post-test components of Experiment 2. For this second stage of the experiment the following exclusion criteria were set a-priori. First, to maximize the quality of the data, participants who scored below chance level (< 55 %) in their performance on the lexical stress perception task, at both Time 1 (pre-test) and Time 2 (post-test), were excluded from the analysis. Additionally, as this longitudinal stage involved a training intervention, participants with lexical stress perception accuracy of more than 90% at pretest were also excluded to allow reasonable room for improvement from the perceptual training (Iverson et al., 2005; Saito et al., 2022c). On the basis of this exclusion criteria, 63 participants were included in the final data analysis (56 female, 7 male), ranging in age from 18 - 48 (*M* = 22.12, *SD* = 4.51). This sample size is in line with previous research as reported by Thomson (2018) who reviewed 32 HVPT studies and found that sample sizes tended to fall in the range of 4 to 36 participants (*M* = 15.3). Note that in the interest of thoroughness, I reran the analysis with the full dataset (N = 102) and found that all of the effects reported as significant in this chapter remained significant when the full dataset was analyzed. The full dataset analysis appears in APPENDIX E for comparison purposes.

Refer to Table 10 below for participant details. The 63 participants were assigned to one of two training conditions - experimental (for lexical stress perception training), and control (for vocabulary training) using stratified randomization to control for possible effects of language proficiency at the start of the study. Participants were first stratified into subgroups based on their LexTALE scores, to include Lower intermediate and lower, Upper intermediate, and Advanced and proficient groups, and then participants from each subgroup were randomly assigned to either the experimental or control training condition. The LexTALE scores for the two trainings were as follows: *M* = 0.54 (SD = 0.09) for the experimental group, and *M* = 0.57 (*SD* = 0.10) for the control group corresponding to lower

intermediate and lower proficiency levels. Refer to section 2.2.6.1 for more details on the LexTALE scoring and interpretation).

Table 10

Participant characteristics

Training group	Trained on	Number of participants (female, male)	Age	Lexical stress perception accuracy at T1	Language proficiency level (mean)
Experimental	Lexical stress perception	33 (28, 5)	18 - 48 (*M* = *22.83, SD* = 5.66)	*M* = 0.678 (*SD* = 0.094)	0.54 (*SD* = 0.09)
Control	Vocabulary	30 (28, 2)	18 -28 (*M* = 21.33, *SD* = 2.62)	*M* = 0.679 (*SD* = 0.085)	0.57 (*SD* = 0.10)

Note. M - mean, *SD* - standard deviation.

3.2.2 Design and study setup

The study design for Experiment 2 was split into three stages (pre-test (i.e., the testing session presented in Chapter 2), training, and post-test). The complete research design is illustrated in Figure 7. The pre-test session, training component, and the final post-test were all administered on the Gorilla Experiment Builder (Anwyl-Irvine et al., 2019, http://www.gorilla.sc). After completing Experiment 1, participants who wished to receive training were registered for Experiment 2, and proceeded to the training component of the study. The same research collaborator who was communicating with participants in Experiment 1 remained their first point of contact. For each stage of the study, participants received emails with their individual login details and for the duration of the training, they also received training reminders.

Participants were tested twice – in Experiment 1 (pre-test session), and after completing the training (post-test session). Training for both the experimental and control groups consisted of 6 training sessions completed within 2 weeks with access allowed for no more than one training session per day.

3.2.2.1 Task presentation

Refer to Chapter 2, section 2.2 for a full description of the tasks and stimuli administered in the pretest. One hundred and two (102) participants proceeded to the training stage of the study and they were assigned to one of two training conditions – an experimental and a control training. Using stratified random sampling, group assignment was counterbalanced for language proficiency, assessed through a lexical test administered as part of the demo session in Experiment 1 (see section 2.2.6.1 for more details on this task).

Figure 7

Schematic representation of the full experimental design illustrating the number and order of tasks presented to participants in the pre-test, training, and post-test phases

In the training phase of the experiment, participants in the experimental condition were trained using two perception tasks: (i) a lexical stress identification task; and (ii) a lexical stress category discrimination task. The estimated time commitment for each training session was 30 minutes. Learners were expected to train for 6 days, and automated procedures in Gorilla ensured they completed no more than 1 training session per day.

The length (6 training sessions) and duration of training (30 minutes x session) were matched for in the control group. Participants there received vocabulary training by way of (i) an L1-to-L2 word matching task; and (ii) a word-picture matching task.

All participants were tested a second time, after completing the training. The post-test session included (i) a lexical stress identification task, (ii) a lexical stress cue weighting task, (iii) a delayed sentence repetition task[2], and (iv) a brief post-test survey. The latter was designed to capture participants' impressions after taking part in the study. Testing took approximately 20 minutes.

3.2.3 Experimental training

3.2.3.1 Exposure stimuli

The stress patterns of interest in this study were disyllabic word pairings with trochaic (stress on the first syllable) and iambic stress (stress on the second syllable). This stimuli selection was motivated by the fact that most real word minimal stress pairs in English have two syllables. The training stimuli were natural recordings of disyllabic nonwords produced by native speakers of British English. A total of 50 nonwords (25 word-stress minimal pairs) were recorded by 4 speakers (2 female, and 2 male).

The use of nonwords for the training materials in this study was motivated by previous research into the benefits of using nonwords when training segments (Ortega et al., 2021; Thomson & Derwing, 2016), and by practical considerations. On the one hand, very few true minimal pairs contrasting only in lexical stress placement exist in English. This constraint severely limits the pool of available options for stimuli creation. On the other hand, there is evidence to suggest that participants who receive training on nonwords demonstrate larger learning benefits than those trained on real words (Thomson & Derwing, 2016). The authors

[2] As noted in Chapter 2, data collected for this task is out of the scope of this dissertation. While production data was collected using a Delayed sentence repetition task, subsequent inspection of the recorded samples revealed severe quality problems. As testing was conducted fully online, there were wide differences in the recording equipment used by participants. This resulted in multiple instances of data loss, and varying degrees of static, background noise, and poor sound quality. Consequently, analysis of the production data was not pursued further.

of that paper have proposed that unlike in the case of using real words where lexical information can prove to be distracting, exposing learners to nonwords makes it easier to direct their attention to the important phonetic characteristics of the target contrasts, essentially directing attention to form as opposed to meaning (Schmidt, 2001). Therefore, pairs of nonwords differing in their stress location were used as stimuli for the training stage of this experiment.

A pseudoword generator was used for the creation of the training stimulus set. The Wuggy software package (Keuleers & Brysbaert, 2010) applies a flexible algorithm to generate nonwords based on a list of real words supplied to the program. As a first step, the algorithm segments each word into its constituent syllables. Next, information about the letters contained in each syllable, the syllable position in the word, the number of syllables making up the word, as well as the neighbouring syllabic elements, is extracted. This forms the basis for producing all possible pseudowords matching the input list which are phonotactically permissible in English.

To generate the stimuli for this study, a list of English lexical stress minimal pairs was fed into the program (see APPENDIX A), thus ensuring the generated exemplars would match the consonantal and vowel structure of existing stress pairs. Twenty-five disyllabic non-words were selected for use in this study. Each word was recorded with stress on the first syllable (trochaic stress pattern), and with stress on the second syllable (iambic stress pattern), thereby forming 25 minimal stress pairs, or a total of 50 words. For a full list of training materials, refer to APPENDIX F.

The training stimuli were recorded by four voiceover artists who were all native speakers of British English (2 female, 2 male). Each nonword was read in a noun sentence frame “I need a...”, and a verb sentence frame “I need to...”. In this way speakers were able to easily apply the necessary shift in stress pattern reflecting the grammatical category required by the carrier sentence. It also ensured a constant prosodic environment across recordings from different speakers. Refer to section 2.2.3.1 for more details on the stimuli acquisition.

The target words were excised and normalized for amplitude in Adobe Audition 2.0. The number of tokens used in the training was 200 (25 nonwords x 2 stress patterns (iambic & trochaic) x 4 speakers), and the same tokens were used across the two HVPT training tasks (lexical stress identification task & lexical stress category discrimination task, see below).

3.2.3.2 Lexical stress identification task

The main task for training participants' stress perception abilities was a perception task involving overt judgements of lexical stress location. This was a forced choice identification task used in most canonical HVPT training paradigms (Iverson et al., 2005; Lively et al., 1993; Logan et al., 1991; Shinohara & Iverson, 2018). In this task, participants heard a linguistic stimulus and responded by selecting one of two response choices presented on the screen. In the present context, participants first heard an exemplar nonword, and were then presented with 2 response buttons on the screen. They showed the same non-word with a different stress pattern, with capital letters designating the stressed syllable on each. For instance, as shown in Figure 8, if participants heard the word "imbert" with stress on the first syllable, they had to select the button appearing on the left-hand side of the screen with the first syllable shown capitalized.

Figure 8

Screenshot of an example trial from the lexical stress perception training task. Participants responded by selecting the button representing the stress pattern they thought they heard. Overt feedback was provided in the form of a green check mark for correct responses, and a red "x" mark for incorrect responses. Additionally, percentage correct responses and response latency could be seen after each response. A progress bar at the top of the screen gave a rough indication of the remaining trials in the current block. (The instructions in Simplified Chinese read "Which word did you hear?").

Trial-by-trial performance feedback was provided (a green check mark appeared after correct responses, and incorrect responses were followed by a red “x” mark). Learners were encouraged to respond as quickly and as accurately as possible. All the while they were able to monitor their response reaction times. The screen also showed a tally of the correct responses out of the total number of submitted responses for the current training block. Lastly, a progress bar was displayed on the screen showing participants’ progress throughout the training blocks. To boost the levels of engagement, at the end of the 4th block of the task, a summary screen displayed participants’ accuracy rate and response speed for that day.

Participants completed 4 blocks of 50 trials per session (200 trials in total). Talker variability was incorporated into every training session (Ortega et al., 2021; Thomson, 2012; Thomson, 2016; Thomson & Derwing, 2016), with participants exposed to all 4 talkers (and the full stimulus set) in every training session. The order of stimuli presentation was randomized within each block.

Data Analysis

Practice trials were excluded from the final data analysis. Two types of measures were computed for this task.

Overt measures of learning

Perceptual learning gains across the 6 days of training were measured by calculating participants' identification response accuracies for each of the training sessions. Portion correct values were computed for each participant individually by dividing the number of correct responses per session by the total number of trials in the same session (i.e., 200 trials).

Covert measures of learning (reaction times (RTs))

Reaction time data were also collected for participants' identification responses as a covert indicator of learning effects. Average RT scores were computed individually for each participant and for each of the six training sessions. Mean RTs for each training session were computed from the median reaction times for all correct trials in each block.

3.2.3.3 Category Discrimination Task

The category discrimination task was a variant of the oddity task in which participants are asked to identify which stimulus from a number of stimuli presented in an array differs based on a specified characteristic (Strange & Shafer, 2008). In the context of the present study, participants heard three stimuli exemplars presented one after the other. The tokens were three different words spoken by three different talkers (Gottfried, 1984; Gottfried et al., 1985; Rato & Rauber, 2015). Two of the nonwords were stressed either on the first or on the second syllable, while one of the tokens had a stress pattern different to the other two.

Participants had to choose which of the three tokens had a different stress pattern, and respond by selecting one of 3 buttons (1, 2, or 3) corresponding to the word with the odd stress pattern (see Figure 9). Trial-by-trial performance feedback was provided as detailed in the previous section. In this task, participants had to focus their attention on the relevant prosodic feature (the location of the lexical stress), to the exclusion of any accompanying variability related to talker voice, and segmental environment. This type of category discrimination task has successfully been implemented in HVPT studies alongside the more traditional forced-choice identification task (Flege, 2003; Iverson et al., 2012; Shinohara & Iverson, 2018). Participants categorized a total of 50 trials per training session, for a total of 300 trials over the 6 training sessions.

Figure 9

Example screenshot of the category discrimination task. Participants responded by clicking on the button of the word they thought had a different stress pattern from the other two. Overt feedback was provided in the form of a green check mark for correct responses, and a red "x" mark for incorrect responses. Additionally, participants could see their percentage correct responses as well as their RTs following each response. A progress bar at the top of the screen gave a rough indication of the remaining trials in each block. (The instructions in Simplified Chinese read "Which word had a different word stress?").

The same measures were collected and analyzed as outlined in section 3.2.3.2 (the previous task).

3.2.4 Control training

Participants in the control group received visual-only vocabulary training aimed to enhance their L2 vocabulary knowledge. The choice of training was motivated by several considerations. Firstly, it was important to offer training to avoid the potential demotivating effects of a no-treatment control group which could have a negative impact on the quality of the collected data at post-test.

Secondly, it was crucial to design a training paradigm unrelated to perceptual training to mitigate the potential for any cross-training transfer effects. Insofar as pronunciation training normally presents target sounds embedded in larger phonetic structures (syllables, or words), there is a theoretical risk that exposure to sounds and structures that are not the targeted focus of training, can still result in perceptual improvement for those specific sounds. In fact, there is some evidence supporting this possibility as shown in a study by Rato and Rauber (2015) who used identical CVC tokens to train vowel production (experimental training), and consonant production (control training) in native Portuguese learners of English. The experimental training group improved on all trained vowel categories, and while the paper does not discuss the control group's gains for consonants, it notes that the control group showed improvements on one of the non-trained vowel categories. In view of this cross-transfer possibility, visual-only vocabulary training was chosen as the control intervention in the current experiment to ensure learners would not inadvertently be primed for any aspects of English phonology by being exposed to auditory-visual vocabulary training.

Lastly, it was crucial to ensure control participants remained engaged and continually motivated for the duration of the full study. To promote retention and maximize the quality of the collected data, both training conditions were designed to maintain comparable interest levels during training. To achieve this, I employed the same engagement and feedback procedures for both training conditions, including corrective feedback, a running tally of correct responses, trial-by-trial response times, as well as end-of-session summary of performance for the day. By ensuring the same level of engagement across both groups, we minimized the potential confounding influence of motivational differences on outcomes.

Similarly, the training schedule for the control group was modelled after the training schedule of the experimental group and consequently, the same time commitments applied. Participants received 6 training sessions spaced over 2 weeks and could access only one training session per day. The vocabulary materials and tasks used in the training are presented in the sections that follow.

3.2.4.1 Vocabulary training sets

The vocabulary training items were selected from Schmitt and Schmitt's Vocabulary Levels Test (2014). The test is a language assessment tool based on the frequency distribution of words in English. Unlike more traditional approaches to vocabulary size testing (Nation, P., 2006) which included low-frequency, high-frequency, academic and technical vocabulary, the Vocabulary Levels test uses five vocabulary categories: 2000-, 3000-, 5000-, and 10000-word frequency families, and academic vocabulary. The word family bands are updated to cover not only high-frequency (the first 2000, and 3000-word families) word families necessary for developing successful everyday communication skills (Adolphs & Schmitt, 2003), low frequency words (word families going beyond the 9000+ level, according to latest estimates (Nation, 2006)), but also vocabulary contained in the in-between word frequency bands (3000 – 9000-word families). These were labelled mid-frequency bands by the authors and argued to confer "authentic purpose" benefits to English learners allowing them to engage more deeply with their interests (watching movies, reading books, etc), as well as the ability to use English for a wider range of topics (Schmitt & Schmitt, 2014). Finally, academic vocabulary is also included to facilitate studies in academic contexts.

These word sets allow for a more effective assessment of learners' vocabulary knowledge as they are more representative of the lexical resources required for real-world English communication in a variety of contexts. For the duration of the training, participants in the control group were exposed to 72 vocabulary items from each category for a total of 360 words (5 vocabulary categories x 72 items). The full vocabulary list used in the training can be found in APPENDIX G. Participants received training on two vocabulary learning tasks: an L1-to-L2 word matching task, followed by a word-picture matching task.

3.2.4.2 L1-to-L2 word-matching task

In each training session, participants were first trained on a Mandarin-to-English word matching task. On each trial, they saw a Chinese word written on the screen and they had to match it with its English counterpart. Three possible choices were presented. If participants responded correctly on their first attempt, they automatically proceeded to the next trial. In order to facilitate the acquisition of novel vocabulary items, in the cases when participants logged an incorrect response on their first try, the selected option was removed until the correct response was chosen, at which point they advanced to the next trial. This meant that in practice participants could log a maximum of 3 attempts (i.e., responses) per trial given that 3 options were presented on each trial (Figure 10).

A green check mark appeared after participants provided a correct answer, and a red "x" mark followed incorrect answers. Similar to the feedback for the experimental training tasks, here too, participants saw their percentage correct responses, reaction times and a progress bar at the top of the screen. At the end of each block, they were also given a summary of their performance for that block.

Figure 10

Screenshot of an example trial from the L1-to-L2 word matching task. Participants matched the Chinese word to its English counterpart. If their first choice was correct, they were moved on to the next trial. If their choice was incorrect, however, that choice was removed until

they selected the correct translation. Immediate feedback was available in the form of a green check mark (for correct responses) and a red "x" mark (for incorrect responses). Participants could also see their percentage correct responses and response latency following each response. A progress bar at the top of the screen gave a rough indication of the remaining trials in a block.

Participants completed 4 blocks of 45 trials per block in each training session (for a total of 180 trials per session). Below is a breakdown of the number of vocabulary items used in the training by word frequency bands (Table 11). Participants were exposed to different sets of words for each training session.

Table 11

Vocabulary training materials per word frequency levels. In each training session participants experienced the following:

Word frequency band	**Number of items**	**Number of presentations of each item**

2,000	12	3
3,000	12	3
5,000	12	3
10,000	12	3
Academic	12	3

Data analysis

Unlike the experimental training condition where participants were exposed to the same stimuli on every session of the training and gains could be measured in an incremental way from session to session, in the vocabulary training condition, the trained items were different in every training session. Consequently, vocabulary learning gains were measured across the training blocks within a session and were ultimately collapsed across block types. Portion correct scores were computed for each participant individually by dividing the number of correct responses per block type by the total number of trials in the same block (i.e., 270 trials (45 trials x 6 training sessions for each block)). Only responses entered on the first attempt were included in the calculations.

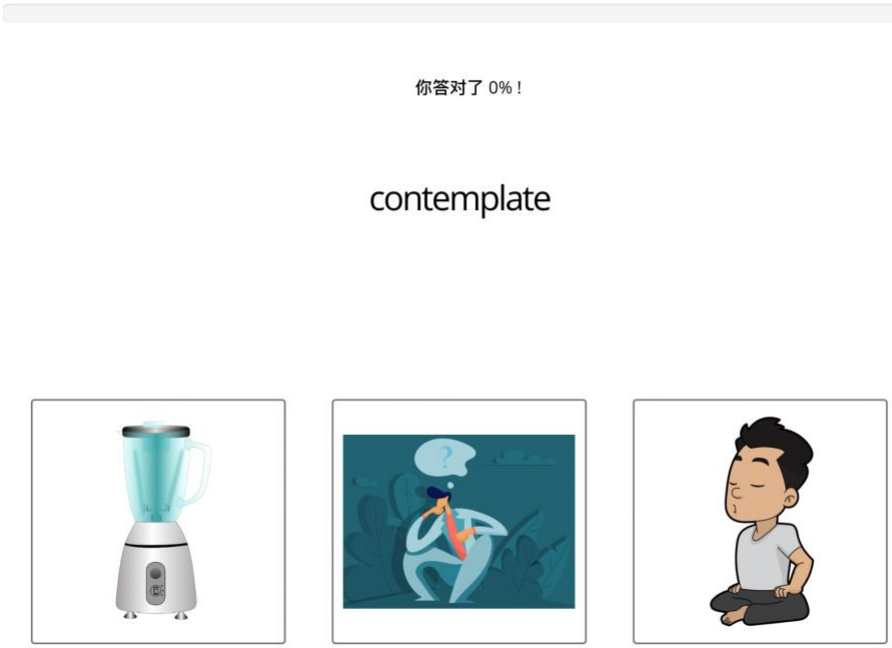
Reaction time data were also collected for participants' vocabulary responses which were used to investigate covert learning effects. Average RT scores were computed individually for each participant and for each of the 4 training blocks. Mean RTs for each training block (1-4) were computed from the median reaction times for all correct trials in each training session block type, calculated per specific block. For instance, the mean RTs for training block 1 were calculated on the basis of the median RTs for all block 1 training blocks (for the 6 training sessions).

3.2.4.3 Word-picture matching task

In the second vocabulary training task, participants matched words to their respective pictures (all public-domain images). On each trial, they saw an English word on the screen with 3 image response options. Their job was to match the word to its picture. If they responded correctly on the first attempt, the task proceeded to the next trial. If they selected the wrong picture, the selected option was removed until the correct picture was selected (Figure 11). Participants were presented with the same feedback as outlined in the previous task (section 3.2.4.2).

Figure 11

Screenshot of an example trial from the word-picture matching task. Participants matched the English word to its best image representation. If their first choice was correct they were moved to the next trial. If their choice was incorrect, that choice was removed until they selected the correct image. Immediate feedback was available in the form of a green check mark (for correct responses) and a red "x" mark (for incorrect responses). Participants could also see their percentage correct responses and response latency following each response. A progress bar at the top of the screen gave a rough indication of the remaining trials in a block.



Twenty-five (25) trials were presented per block, for a total of 75 trials per session. Table 12 lists the number of words presented at each training session by word frequency levels. Participants were exposed to different sets of words for each training session.

Table 12

Breakdown of the vocabulary training materials for the word-picture matching task by word frequency bands

Word frequency bands	**Number of items**	**Presentations of each item**
2,000	5	3
3,000	5	3
5,000	5	3
10,000	5	3
Academic	5	3

Data analysis

Refer to section 3.2.4.2 above for details on the data analysis.

3.2.5 Post-test

Stress processing was assessed at separate time points in this longitudinal experiment. At Time 1 (data presented in Chapter 2), participants' baseline lexical stress perception and lexical stress cue weighting strategies were recorded before the training took place. At Time 2, a post-test experiment was conducted after the completion of all training sessions, to

examine participants' performance on the same two measures collected at Time 1. Participants in both the experimental and control groups were tested at both time points.

Twelve tasks were completed at Time 1 including auditory processing measures and lexical stress processing measures, and of those only the lexical stress processing measures were administered at the post-test. These were (1) the lexical stress perception task; (2) the lexical stress product task (see section 3.2.2.1) and (3) the prosodic cue weighting task. Finally, participants filled out a post-test survey (see APPENDIX H for the contents of the survey). The perception and cue weighting tasks were identical to those administered at Time 1. The methods section of Chapter 2 can be consulted for full details about the stimuli generation, test procedure, and data analysis for each task. Refer to Figure 7 for the full test protocol.

3.3 Results

All data processing and statistical analysis for this Experiment were conducted in R (R Core Team, 2023). All visualization graphics were created using the ggplot2 package in R (Wickham, 2016).

3.3.1 Overall improvement in lexical stress perception from pre-post test

To investigate improvements in lexical stress perception from Time 1 (= pretest) to Time 2 (= posttest), participants' portion correct scores and their mean response latencies were submitted to separate repeated-measure ANOVAs.

The pre- and post-test lexical stress perception scores for participants were submitted to a two-way analysis of variance (ANOVA) with Time (Time 1 (= pretest), Time 2 (= posttest)) as a repeated variable, and Group (Experimental (high variability prosodic training), and Control (vocabulary training)) as a between-subjects variable. For effect size, I also

calculated partial eta squared and interpreted it in line with Cohen's (1988) definition for small (ηp2 = 0.01), medium (ηp2= 0.06), and large (ηp2 = 0.14) effects.

The analysis showed a significant main effect of group, F(1, 61) = 6.063, $p < 0.05$, η2p = 0.09, and a main effect of time, F(1, 61) = 35.007, $p < 0.001$, η2p = 0.37. There was also an interaction effect between Group and Time, F(1, 61) = 28.088, $p < 0.001$, η2p = 0.32. Based on the results of the main effects, it followed that controlling for Time, lexical stress perception scores were different between the experimental and control groups. Further, controlling for Group, lexical stress perception scores were different for at least one of the times. Then, to investigate the effect of the between-subjects factor, Group, on lexical stress perception scores at every time point (Time 1, Time 2), two separate t-tests were conducted on the mean perception scores of the 2 groups at Time 1 and Time 2. At Time 1, the means of the experimental and control groups were not significantly different, t(61) = 0.05, p = 0.96, d = 0.01 (small effect size), means for the experimental group, M = 0.68, SD = 0.09; and for the control group, M = 0.68, SD = 0.09. At Time 2, however, the means of the 2 groups differed significantly, t(61) = -4.09, $p < 0.001$, with Cohen's d = 1.03, indicating a large effect size (means for the experimental group, M = 0.80, SD = 0.12; and for the control group, M = 0.69, SD = 0.10, see Table 13). These results indicate that the simple main effect of Group was significant at Time 2, but not at Time 1. The lexical stress perception accuracy of the trained and control groups was, therefore, comparable at the beginning of the training, but their performance was different following the training intervention.

Paired samples t-tests were also performed to evaluate if there was a difference between participants' mean perception scores from Time 1 to Time 2 within each group. For the experimental group, the mean perception scores at Time 2, M = 0.80 (SD = 0.12) were significantly higher than their mean perception scores at Time 1, M = 0.68 (SD = 0.09), t(64) = -4.63, $p < 0.001$ (Cohen's d = 1.38). For the control group, there was no statistically significant difference between their mean performance at Time 1, M = 0.68 (SD = 0.09), and Time 2, M = 0.69 (SD = 0.10), t(58) = -0.28, p = 0.78 (Cohen's d = 0.08). The results confirm that participants in the Experimental group significantly improved their lexical stress perception scores from Time 1 to Time 2 by 12.3% from 67.8% at pre-test to 80.1% at post-test. The control group did not show any significant changes in their lexical stress perception

scores between Time 1 and Time 2 (starting out at 67.9% accuracy score at pretest, and scoring at 68.6% at post-test (Table 13), see Figure 12 below illustrating individual participants' learning trajectory from Time 1 to Time 2).

Figure 12

Lexical stress identification accuracy for the two training conditions (left - control group, right - experimental group), at Time 1 (pre-test), and Time 2 (post-test) for individual participants' data points (lines)

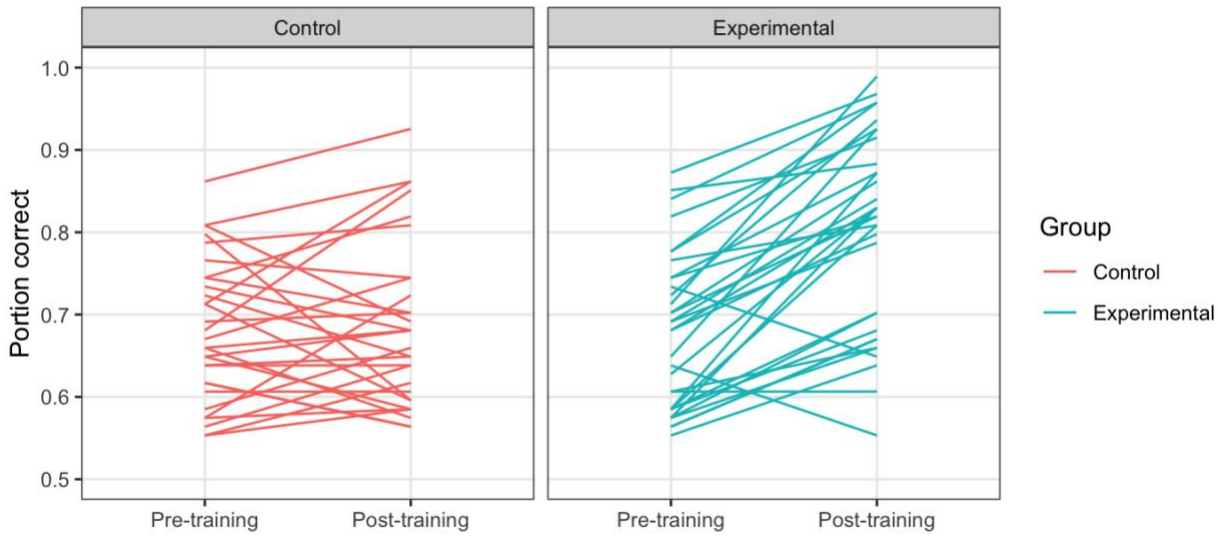


Table 13

Summary statistics of participants' lexical stress perception performance at Time 1 and Time 2 for the experimental and control groups

Group	Time	Variable	n	M	SD
Experimental	Time1	Portion correct	33	0.68	0.09
Experimental	Time2	Portion correct	33	0.80	0.12

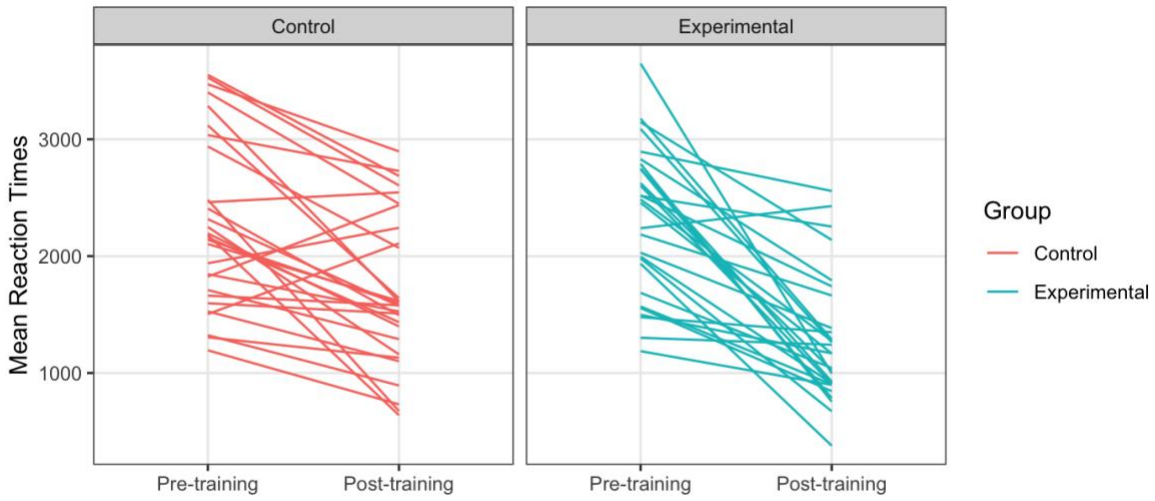
Control	Time1	Portion correct	30	0.68	0.09
Control	Time2	Portion correct	30	0.69	0.10

Note. n - sample size, M - mean, SD - standard deviation

Next, participants' mean response latencies were calculated for correct responses only. Mean response times which were more than 2 standard deviations away from the mean were excluded from the analysis. The computed mean response times were submitted to a two-way ANOVA with Group (Experimental, Control) as a between-subjects factor, and Time (Time 1, Time 2) as a within-subjects factor. The analysis revealed a main effect of Time, $F(1, 55) = 86.137$, $p < 0.001$, $\eta 2p = 0.61$, and an interaction effect between Group and Time, $F(1, 55) = 6.62$, $p < 0.05$, $\eta 2p = 0.11$, but not a main effect of Group, $F(1, 55) = 1.859$, $p = 0.18$, $\eta 2p = 0.03$. Mean response times decreased from 2317.943 ms (*SD* = 637.273 ms) to 1435.469 ms (*SD* = 725.554) for the Experimental group, and from 2326.052 ms (*SD* = 735.227 ms) to 1702.152 ms (*SD* = 652.874 ms) for the Control group from Time 1 to Time 2, respectively (Figure 13). Bonferroni-corrected pairwise tests showed no significant difference between the mean response times of the experimental and control groups at Time 1, $F(1,56) = 0.002$, $p = .96$, Cohen's d = .01; or their mean response times at Time 2, $F(1,59) = 2.259$, $p = .14$, Cohen's d = .4. However, mean response times differed significantly between Time 1 and Time 2 for both the experimental group, $F(1,58) = 24.7$, $p < .001$, Cohen's $d = 1.29$, and the control group, $F(1,57) = 11.84$, $p < .01$, Cohen's $d = .9$.

Figure 13

Response latencies in the lexical stress perception task for the two training conditions (left - control group, right - experimental group), at Time 1 (pre-test), and Time 2 (post-test). The plot shows mean response times for correct responses only presented in milliseconds for individual participants.



Additional analysis was conducted with the data from all 102 participants (no exclusion criteria applied), and the effects reported as significant here remained unchanged. As such, only the results from the more stringent analysis are reported in this section. Results from the ANOVA run with the full dataset are available in APPENDIX E.

3.3.2 Lexical stress perception performance during training (experimental group)

In the previous section, I found that the HVPT training led to improvements in the experimental groups' lexical stress perception accuracy scores from Time 1 to Time 2. Next, I examined the learning trajectories of participants in the experimental group across the 6 training sessions. They were trained on two different lexical stress tasks - a forced-choice identification task, and a category discrimination task. The following two sections present the results of this group's training performance on each of the training tasks.

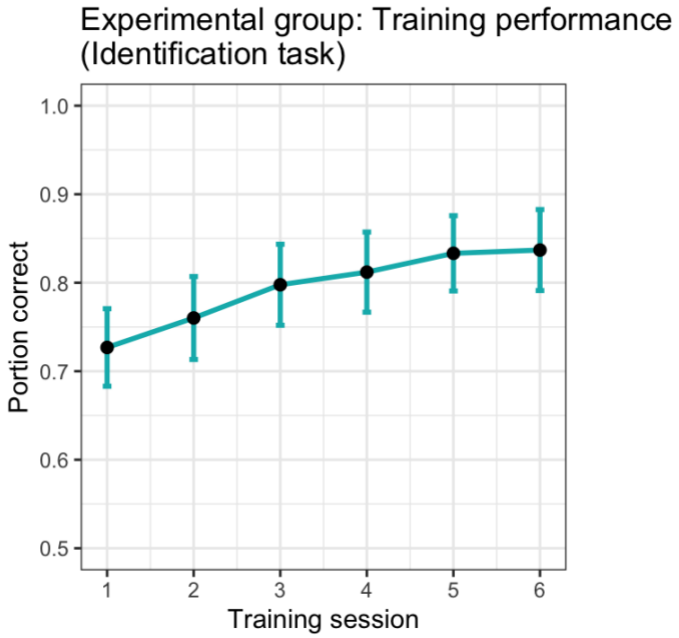
Lexical stress identification training task

For the lexical stress identification training task, participants' performance showed steady improvement across the training sessions (Figure 14).

Participants' identification accuracy scores were submitted to a one-way repeated analysis of variance with session (x6) as a within-subjects factor. Performance across the training sessions was significantly different $F(5, 160) = 22.28$, $p < 0.001$, $\eta 2p = 0.41$. To identify which training sessions were different, follow-up post-hoc Bonferroni-adjusted t-tests were conducted. The results support gradual improvement of lexical stress perception over time, with the mean accuracy scores from training session 1 ($M = 0.73$, $SD = 0.12$) differing significantly from those of training session 5 ($M = 0.83$, $SD = 0.12$), $p = .012$; and training session 6 ($M = 0.84$, $SD = 0.13$), $p < .01$.

Figure 14

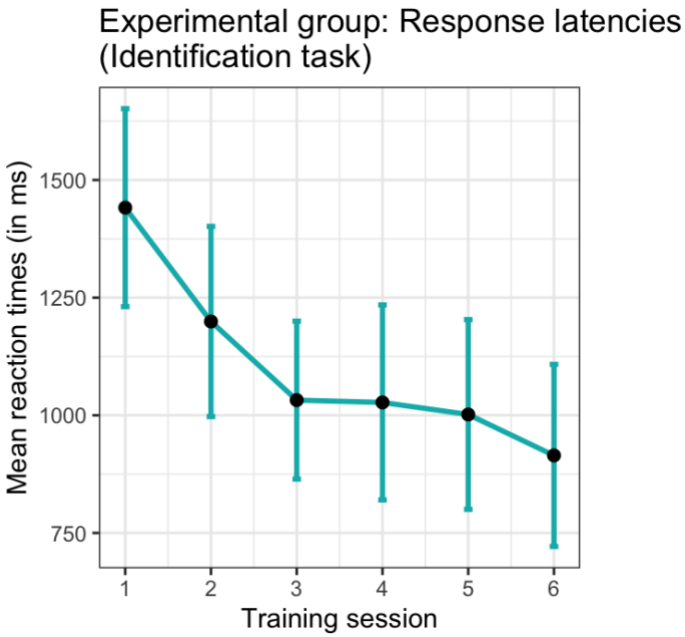
Participants' lexical stress identification accuracy throughout the training. Participants showed gradual improvement in identification accuracy across the 6 training sessions. The plot presents mean response accuracy scores with error bars indicating 95% confidence intervals.



Response latencies during training were also collected as a covert measure of learning. Note that this analysis includes response times for correct responses only. Additionally, mean response times which fell more than 2 standard deviations away from the mean were excluded from the analysis. A one-way analysis of variance (ANOVA) was performed on the mean response times with training session as a within-subjects variable to investigate how response times changed from session to session and if there was a significant difference between the mean scores. The ANOVA was significant, $F(5, 148) = 14.91$, $p < 0.001$, $\eta 2p = 0.34$. A post-hoc Bonferroni corrected analysis indicated that the mean response times for training sessions 4 (M = 1027.35 ms, SD = 564.79 ms), $p = .04$, training session 5 (M = 1001.76 ms, SD = 559.17 ms), $p = .02$, and session 6 (M = 914.75 ms, SD = 536.41 ms), $p < .01$ were significantly faster compared to mean response times recorded at the first training session (M = 1441.24 ms, SD = 573.27 ms). Refer to Figure 15 below for participants' trajectory in response time changes in the lexical stress identification training task. Mean response times were gradually becoming faster across the training intervention, such that the changes in lexical stress identification performance as seen in Figure 14 were also mirrored by changes in response latencies. Notably, as participants were becoming more accurate in their lexical stress identifications, they were also responding faster.

Figure 15

Changes in response latencies during training. The plot shows mean response times for correct responses presented in milliseconds with error bars indicating 95% confidence intervals.

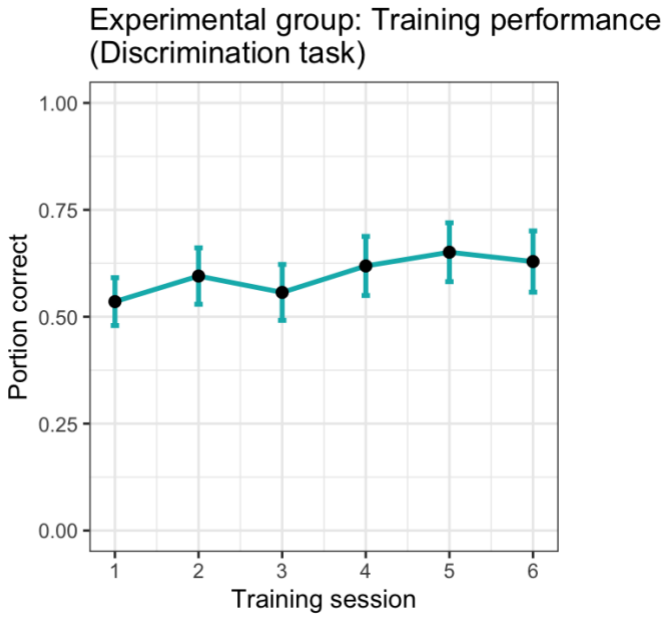


Lexical stress category discrimination task

Participants in the experimental group were also trained on a category discrimination task where they had to identify which of three tokens that they heard had a stress pattern different to the other two. A one-way ANOVA conducted on participants' categorization scores across the 6 training sessions (the within-subjects variable), revealed a significant difference between training performance across different days of training, $F(5, 160) = 7.557$, $p < .001$, $\eta 2p = .19$. Figure 16 displays participants' categorization discrimination accuracy with small incremental improvements for most of the sessions. Post hoc analysis with a Bonferroni adjustment revealed that lexical stress discrimination accuracy improved significantly from training session 1 ($M = 0.54$, $SD = 0.16$) to training session 4 ($M = 0.62$, $SD = 0.19$), $p < .05$, training session 5 ($M = 0.65$, $SD = 0.19$), $p < .01$, and training session 6 ($M = 0.63$, $SD = 0.20$), $p < .05$. Additionally, there was a statistically significant difference between performance means from training session 3 ($M = 0.56$, $SD = 0.18$) to training session 5 ($M = 0.65$, $SD = 0.19$), $p < .01$.

Figure 16

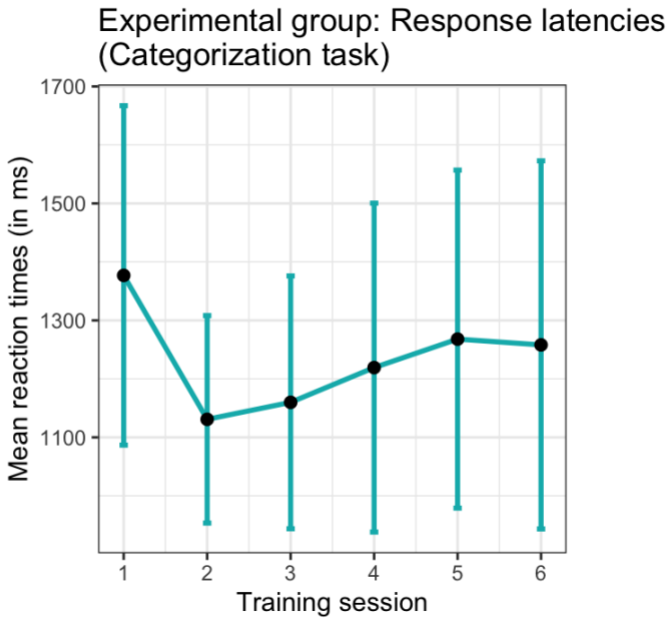
Changes in participants' lexical stress discrimination accuracy. Participants showed relatively small improvements in accuracy across the 6 training sessions. Mean response accuracy scores with error bars indicating 95% confidence intervals.



An examination of the mean response latencies across the 6 training sessions showed no statistically significant difference between the mean reaction times, F(5, 148) = 0.81, *p* = 0.55, η2p = .03. In fact, as can be seen in Figure 17 below, after a slight decrease in mean reaction times from training session 1 (*M* = 1376.87, *SD* = 804.83) to training session 2 (*M* = 1130.78, *SD* = 483.65), participants' reaction times underwent almost no changes for the remainder of the training.

Figure 17

Participants' response latencies during training in the category discrimination task. After a slight decrease in mean reaction times from training session 1 to training session 2, response times remained relatively unchanged across the remaining training sessions. Mean response times for correct responses are presented in milliseconds with error bars indicating 95% confidence intervals.



Lastly, to investigate the possible influence of participants' working memory on their training performance in this task, a series of non-parametric correlations (Spearman rank correlation coefficient) were performed on participants' digit span scores collected at Time 1 (see section 2.2.6.2), and their training performance on the lexical stress category discrimination task. Analysis was run in terms of relative gains from session to session, as well as overall relative gains (from the first to the last training session), on the one hand, and relative changes in response accuracy from training session to training session and overall relative changes in mean response times (from the first to the last training session), on the other. The results from all the correlation tests performed are reported in Table 14 below. For overall relative changes in response latencies (Day 1 relative to Day 6), there was a significant negative correlation between participants' working memory scores and the relative change in mean response times (Spearman rho = −0.47, p <.01, moderate as per Cohen (1988)). In other words, participants with higher working memory scores at Time 1 had a greater decrease in response times from training session 1 to training session 6 in the category discrimination training task. A marginally significant positive correlation between working memory scores and overall relative gains in lexical stress discrimination between session 1–session 6 was also noted (Spearman rho = 0.32, p = .07). No other significant correlations were found.

Table 14

Results of Spearman correlation tests relating working memory and session-to-session relative gains, relative changes in response latency, and overall training relative gains and response latency changes

Lexical stress discrimination	**Relative gains M (SD)**	**Working memory**
Session 1 - Session 2	0.14 (0.34)	0.07
Session 2 - Session 3	-0.04 (0.23)	0.25
Session 3 - Session 4	0.14 (0.28)	-0.09
Session 4 - Session 5	0.07 (0.22)	-0.02
Session 5 - Session 6	-0.01 (0.27)	0.27
Session 1 - Session 6	0.22 (0.43)	0.32•
Response latencies	**Relative changes in RTs M (SD)**	**Working memory**
Session 1 - Session 2	-0.08 (0.22)	-0.22
Session 2 - Session 3	0.08 (0.35)	0.03
Session 3 - Session 4	0.06 (0.37)	-0.32
Session 4 - Session 5	0.09 (0.49)	0.11
Session 5 - Session 6	0.00 (0.32)	-0.17
Session 1 - Session 6	-0.03 (0.36)	**-0.47****

•$p < .1$ **$p < .01$

Note. *M* = mean, *SD* = standard deviation. Statistically significant correlations are bolded.

3.3.3 Vocabulary performance during training (control group)

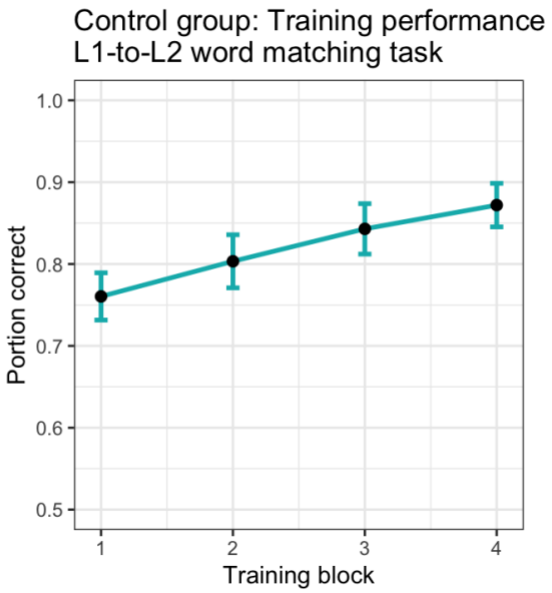
Participants in the control group completed 6 vocabulary training sessions. The following two sections report on their performance in the two training tasks – L1-to-L2 word-matching, and word-picture matching tasks.

L1-to-L2 word-matching task

As participants in this training experienced novel items in every training session, to track training improvements in this task, portion correct were collapsed across blocks, for a 4-level within- subjects factor (4 training blocks per session). A one-way repeated measures ANOVA was conducted with participants' portion correct scores as the outcome variable, and block as the within-subjects factor. The difference in the mean scores was statistically significant, $F(3, 87) = 77.53$, $p < .001$, $\eta 2p = 0.73$. Follow-up post-hoc (Bonferroni corrected) t-tests showed that participants' response accuracy differed significantly from training block 1 ($M = 0.76$, $SD = 0.08$) to training blocks 3 ($M = 0.84$, $SD = 0.08$), $p < 0.001$, and 4 ($M = 0.87$, $SD = 0.07$), $p < 0.001$, as well as from training block 2 ($M = 0.80$, $SD = 0.09$) to training block 4 ($M = 0.87$, $SD = 0.07$), $p < 0.01$ (see Figure 18).

Figure 18

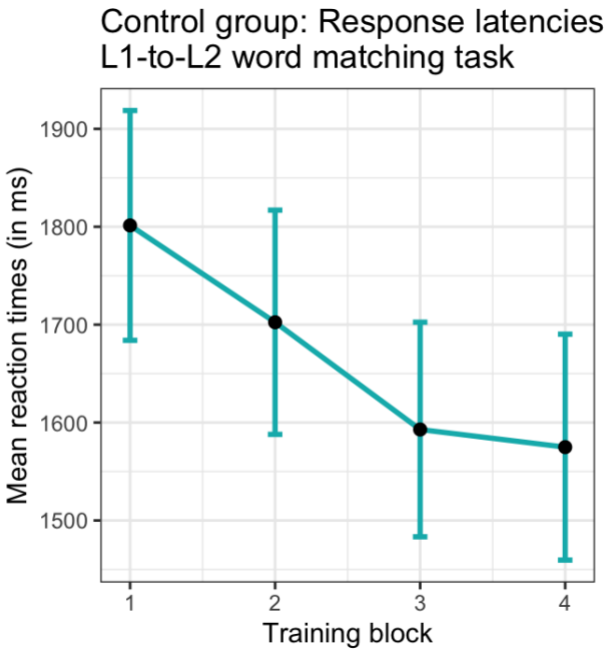
Changes in participants' vocabulary response accuracy. Participants showed consistent improvements in accuracy across the 4 training blocks within training sessions, with significant differences in group means across the blocks. Mean response accuracy scores with error bars indicating 95% confidence intervals.



For response times, a one-way repeated measures analysis of variance was also conducted, with participants' mean response times collapsed across blocks as the dependent variable, and blocks (x4) as the within-subjects factor. The effect of block was significant at the $p <$ .001 level, $F(3, 82) = 66.07$, $\eta 2p = 0.71$ (see Figure 19). Post-hoc pairwise comparisons (Bonferroni corrected) showed response time means differed between the first block (M = 1801.37, SD = 296.67) and the fourth training blocks (M = 1574.92, SD = 308.97), $p < 0.05$. The remaining pairwise comparisons were non-significant.

Figure 19

Participants' response latencies during training on the L1-to-L2 vocabulary word matching task. Response times gradually decreased from training block 1 through training block 4. Mean response times for correct responses are presented in milliseconds with error bars indicating 95% confidence intervals.

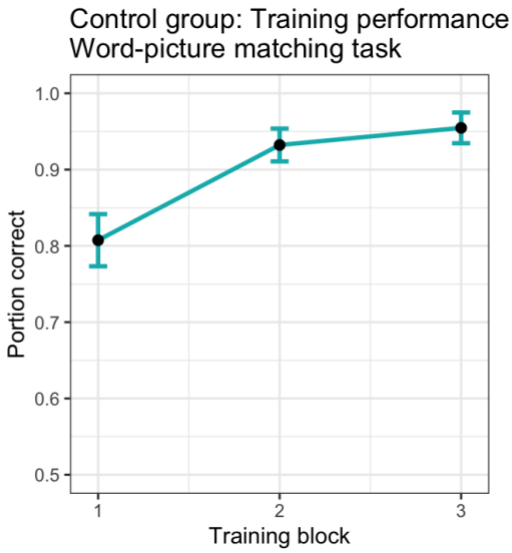


Word-picture matching task

In the word picture matching task, participants trained through 3 blocks per training session. Portion correct scores were collapsed across blocks. A one-way repeated measures ANOVA with block (3 levels) as the within-subjects factor, and portion correct scores as the dependent variable was significant, $F(2, 58) = 125.4$, $p < .001$, $\eta 2p = 0.81$. Mean accuracy scores differed between training block 1 ($M = 0.81$, $SD = 0.09$) and training block 2 ($M = 0.93$, $SD = 0.06$), $p < 0.001$, and 3 ($M = 0.95$, $SD = 0.05$), $p < 0.001$. Accuracy scores did not differ between training blocks 2 and 3, $p = 0.65$ (see Figure 20).

Figure 20

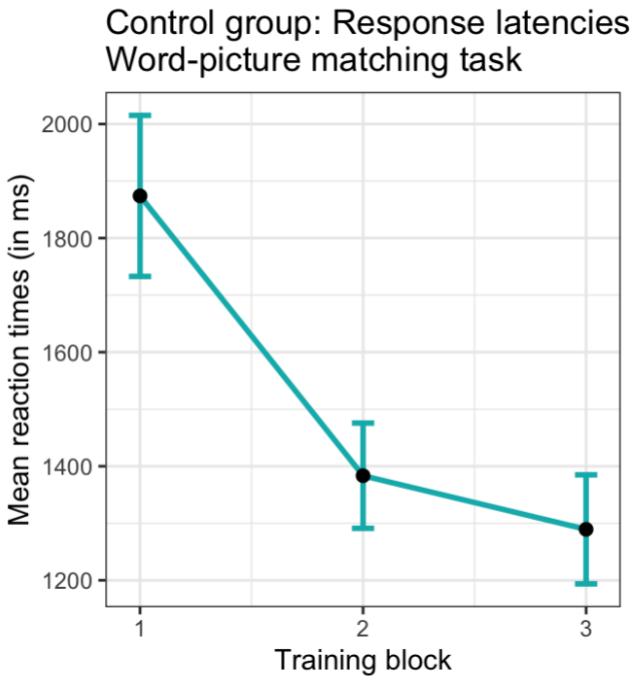
Changes in participants' vocabulary response accuracy in the word-picture matching task. Participants showed improvement in accuracy in within-session training, with significant differences in group means between blocks 1–2, and blocks 1–3. The mean response accuracy scores are shown with error bars indicating 95% confidence intervals.



Similarly, participants' response latencies decreased significantly within training sessions. A one-way ANOVA conducted on mean response times with block (x3) as a within-subjects factor, was significant $F(2, 54) = 146.5$, $p < .001$, $\eta 2p = 0.84$. Response times significantly decreased within sessions between training block 1 ($M = 1873.98$, $SD = 349.60$) and training block 2 ($M = 1383.35$, $SD = 247.15$), $p < .001$, as well as between training block 1 and training block 3 ($M = 1289.31$, $SD = 255.99$), $p < .001$ (as seen in Figure 21).

Figure 21

Participants' response latencies during training sessions on the word-picture matching task. Response times decreased from training blocks 1 to training blocks 2 and 3 within sessions. Mean response times for correct responses are presented in milliseconds with error bars indicating 95% confidence intervals.



3.3.4 Relationship between individual differences in auditory processing abilities at pre-test and relative learning gains at post-test

To investigate the possible relationship between participants' auditory processing abilities collected at Time 1 and their relative learning gains in lexical stress perception at Time 2, I first conducted an exploratory factor analysis in order to uncover any latent factors underlying the auditory processing measures and thus reduce the number of predictor variables used in the regression model. The decision to perform exploratory factor analysis was necessitated by the sample size at Time 2 (N=63), which was insufficient to detect medium-sized effects (f^2=0.15) with 8 predictors, given a significance level of α=0.05 and a power of 1−β=0.80. For multiple regression analysis with medium-sized effects, approximately 12.5 participants per predictor are recommended, corresponding to a total sample size of N=100. Consequently, I submitted a total of eight auditory processing measures collected at Time 1 to the factor analysis: pitch discrimination, formant discrimination, risetime discrimination, dimension selective attention to pitch, dimension selective attention to formant, dimension selective attention to risetime, rhythm memory, and melody memory. These variables were factor analysed across both the experimental and control groups (N = 63) with a varimax orthogonal rotation.

Preliminary testing verified the factorability of the dataset with a significant Bartlett's test of sphericity ($\chi 2 = 105.233$, $p < .001$), and a middling Kaiser-Meyer-Olkin measure of sampling adequacy, KMO = .76 (values interpreted as per Kaiser & Rice, 1974). According to the eigenvalue criterion ≥ 1, only 2 factors were extracted accounting for a cumulative variance of 42% (factor loadings are summarized in Table 15).

Table 15

Results from the factor analysis of 8 auditory processing measures using an oblique (Varimax) rotation.

	Factor 1: Spectral Processing (3.09)	**Factor 2: Temporal processing** (1.09)
Cumulative variance %	25%	42%
Pitch discrimination	**-.56**	.14
Formant discrimination	**-.42**	.19
Risetime discrimination	-.10	**.99**
Dimension selective attention for pitch	**.64**	-.10
Dimension selective attention for formant	**.60**	-.18
Dimension selective attention for risetime	.24	**-.36**
Rhythm memory	**.41**	**-.32**
Melody memory	**.71**	-.20

Note. Factor loadings above .30 are bolded. The original eigenvalues for each factor are presented in parenthesis.

A review of the first factor shows high factor loadings mostly for pitch- and formant-related processing abilities (formant and pitch discrimination, dimension selective attention to pitch and dimension selective attention to formant), spectral memory and reproduction abilities

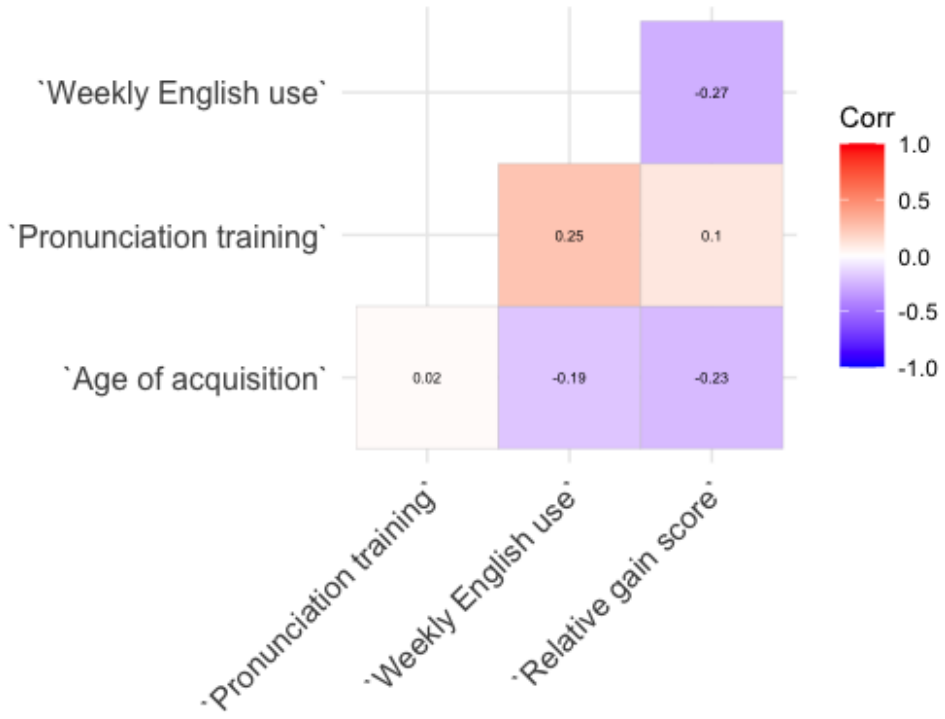
(melody memory), and rhythm memory. Thus, the first factor identified in the analysis can be interpreted as indexing mostly spectral processing abilities. The second factor is strongly dominated by risetime-related measures (risetime discrimination and selective attention to risetime), as well as temporal memory and reproduction skills (rhythm memory) and can thus be interpreted as describing temporal processing abilities. Having interpreted the factor constructs from the analysis, I then extracted individual factor scores using the regression method in the factanal function from the stats package in R (DiStefano et al., 2009).

Next, I conducted a multiple regression analysis using the computed scores for Factor 1 (spectral processing), and Factor 2 (temporal processing) as predictor variables, and participants' relative gain scores in lexical stress perception as the dependent variable. The analysis did not yield a statistically significant relationship between either of the two predictors (t = .18, p = .86, for Factor 1, and t = –.99, p =.33, for Factor 2, respectively) and participants' relative learning gains, $F(2,30) = .49$, $p = .62$, $R^2 = -.03$.

To examine the potential relationship between participants' relative gain scores (RGS) in lexical stress perception between Time 1 and Time 2 and some of their language experience measures, I also ran a series of correlations between RGS and the following measures: age of acquisition, pronunciation training, and weekly English use. There was no statistically significant correlation between any of the language experience measures and participants' RGS scores (r (31) = -.23, p = 0.2 for AOA, r (31) = .10, p = 0.58 for pronunciation training, and r (31) = -.27, p = 0.13 for weekly English use), see Figure 22.

Figure 22

Correlation matrix for participants' language experience measures and their relative gain scores from in lexical stress perception from Time 1 to Time 2.



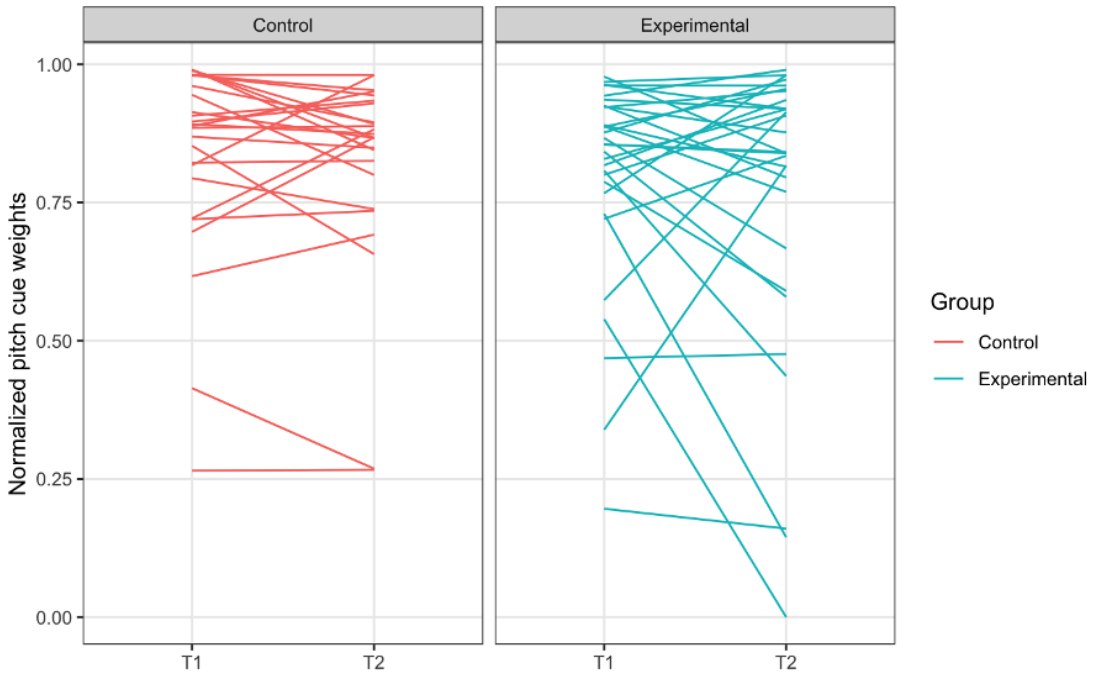
3.3.5 Changes in pre/post-test lexical stress cue weighting strategies

To explore if short-term intensive HVPT training can lead to shifts in participants' cue weighting strategies when categorizing lexical stress contrasts, the normalized pitch cue weights collected at Time 1 and Time 2 for the two groups (Experimental, Control) were submitted to a two-way ANOVA with Group as a between-subjects factor and Time as a within-subjects factor. The data of 55 participants (who had a significant relationship between at least one of the stimulus dimensions (pitch, or duration) and their lexical stress categorization responses at both Time 1 and Time 2) were analysed. No statistically significant effects were found for Group, $F(1, 53) = .777$, $p = .38$, $\eta 2p = .01$, Time, $F(1, 53) = 1.201$, $p = .28$, $\eta 2p = .02$, nor for the interaction of Group x Time, $F(1, 53) = .168$, $p = .68$, $\eta 2p = .003$. Figure 23 below presents the means and confidence intervals of the two groups' pitch reliance at Time 1 and Time 2 expressed through their normalized pitch cue weights.

For the same analysis run on the two groups' normalized duration cue weights, see APPENDIX I.

Figure 23

Participants' normalized pitch cue weights at Times 1 and 2. There were no statistically significant changes in cue reliance between the Time 1 and Time 2 testing for either of the groups. Mean normalized pitch cue weights are shown with error bars indicating 95% confidence intervals.



3.4 Discussion

In Experiment 2 of this dissertation, I set out to investigate if short-term suprasegmental perceptual training would improve Mandarin learners' L2 English lexical stress processing. More specifically, I employed a modified version of the high variability phonetic training technique (HVPT) adapted to the prosodic domain and asked whether online short-term identification and discrimination training of natural tokens produced by multiple talkers in different phonetic environments would enhance L2 lexical stress perception. Additionally, I

collected learners' lexical stress cue weightings, or the extent to which they relied on specific acoustic information to determine where stress falls within a word, before and after training to investigate if the relative reliance on acoustic dimensions would undergo significant changes at post-test compared to pre-test. Finally, I examined the possible role of individual differences in learners' domain-general auditory processing abilities at pre-test in modulating the effectiveness of the high-variability prosodic training.

Lexical stress perception

The principal finding from this study indicates that high variability prosodic training can improve L2 lexical stress perception. Mandarin learners who received perceptual training showed significantly better lexical stress perception accuracy from Time 1 to Time 2, compared to the performance of learners in the control group. Furthermore, training on nonwords generalized to the perception of real words and untrained voices. Indeed, identification accuracy improved from 67.8% at Time 1 to 80.1% at Time 2 for the experimental group; an overall improvement of 12.3% which is in line with the degrees of improvement reported in other HVPT studies (ranging from approximately 5 - 29% depending on the trained contrasts, Hirata et al., 2007; Nishi & Newly-Port, 2007; Rato, 2014; with most studies reporting gains averaging around 15 percentage points, Bradlow et al., 1999, 1997; Iverson et al., 2005; Shinohara & Iverson, 2018).

The presented findings provide preliminary evidence that the high variability phonetic training approach predominantly used to promote segmental learning (Bradlow et al., 1997; Lively et al., 1993; Lambacher et al., 2005; Logan et al., 1991; Shinohara & Iverson, 2018; Thomson, 2012) can also be used to enhance L2 prosodic perceptual learning with a reasonable degree of success. Prior work using HVPT training to aid prosodic acquisition did so in the context of teaching Mandarin tones to non-tonal language speakers (Perrachione et al., 2011; Sadakata & McQueen, 2014; Wang et al., 1999). The current data expands the available research and provides additional empirical support for the efficacy of high variability perceptual training beyond the segmental context.

Lexical stress cue weighting strategies

Results relating to the second research question, showed no statistically significant changes in dimensional cue weighting from pre- to post-training for the experimentally trained group. There could be several explanations for this finding. Firstly, the training was not designed to target cue weighting strategies specifically. Auditory training studies that target learners' perceptual cue weighting strategies normally employ synthetic stimuli where relevant acoustic dimensions are manipulated to either enhance or diminish the usefulness of specific acoustic information with the explicit aim of producing a shift in dimensional weightings. Prior work in this area has shown varying degrees of success in changing cue reliance for L2 learners which can largely be attributed to the degree of interference from the L1 phonetic representations (Iverson et al., 2005 for Japanese learners' F3 weighting in the perceptions of the English /r-l/ contrast; Ylinen et al., 2010 for cue weighting in L2 vowel perception). It should also be noted that L2 cue weightings in some instances do not undergo significant changes even after years of immersion experience in the target language environment (Cebrian, 2006; Ingvalson et al., 2011; Morrison, 2002).

Secondly, while shifts towards more native-like cue weighting strategies in suprasegmental categorization have been reported for native Mandarin learners of English after longer immersion experience (Petrova et al., 2023), it may be the case that these were prompted by communicative necessities. In the case of phrase boundary categorization examined in Petrova et al. (2023), native Mandarin speakers with extensive immersion experience (UK residence > 3 years) weighted duration information more highly compared to inexperienced learners with shorter LOR, thus exhibiting more native-like cue weighting strategies. However, in the context of phrase boundary perception, there is evidence to suggest that duration information influences categorization judgements more reliably than pitch (Streeter, 1978). In this regard, overreliance on the pitch dimension that starkly contrasts with native speaker strategies for processing prosodic boundaries could have implied poorer communication outcomes for Mandarin speakers. Practical necessity, therefore, could have triggered Mandarin speakers' changes in cue weighting strategies. Following this line of thought, if we examine English lexical stress perception where, by multiple accounts, pitch is

the primary acoustic correlate for stress location (Beckman, 1986; Bolinger, 1958; Fry, 1958; Lieberman, 1960; Morton & Jassem, 1965), then it is possible that Mandarin learners may not need to shift their reliance away from pitch to achieve optimal categorization. Some evidence of support for this possibility comes from Tremblay et al. (2023) where Korean learners' of English received HVPT lexical stress training with the aim of altering learners' cue weighting strategies. The study reported changes in cue reliance from pre-to-post test with a tendency for learners to rely more heavily on vowel quality information when the manipulated dimensions were those of vowel quality and pitch, and vowel quality and duration. However, when the test stimuli varied by pitch and duration only, the authors observed increased reliance on pitch from pre-to-post test. This finding seems to suggest that in the context of pitch versus durational perceptual weighting, native non-tonal learners of English, tend to shift their cue weighting strategies towards greater reliance on pitch after intense exposure to natural lexical stress recordings. However, it is well-documented that native Mandarin speakers already upweight pitch information in lexical stress processing (Wang, 2008; Zhang et al., 2008; Yu & Andruski, 2010; but also refer to Chrabaszcz et al., 2014), and therefore, it is possible that no changes in cue weighting strategies were necessary for this population to achieve successful lexical stress categorization.

Importantly, the latter possibility concerns the perceptual categorization of lexical stress in the tested instances constrained by combinations of changes along the pitch and duration dimensions. Given that lexical stress is cued by four different acoustic dimensions: pitch, duration, intensity, and the spectral characteristics of the vowel (Beckman, 1986; Bolinger, 1961; Fry, 1955, 1958, 1965; Lieberman, 1960), it should be noted that the current results show no statistically significant changes in Mandarin learners' cue reliance for pitch and duration information. No claims can be made about the perceptual weightings of amplitude and vowel quality, as these dimensions were not tested in the current experiment.

Training performance and methodological considerations

Participants in the experimental group of this study received training on lexical stress perception using nonwords as training stimuli created to conform with the English language

phonotactic constraints, and produced by 4 different British English speakers. The use of nonwords was partially motivated by prior work in segmental (Ortega et al., 2021; Thomson & Derwing, 2016) and tonal acquisition (Perrachione et al., 2011; Sadakata & McQueen, 2014), and the limited number of true lexical stress minimal pairs in English. Results from the current training offer strong evidence that intensive exposure to nonwords containing the target segmental or suprasegmental feature is an effective way to train L2 contrasts. Authors have suggested that pseudowords may offer an advantage over the use of real words in phonetic training, as the latter also triggers lexical representations which may interfere with perceptual acquisition (Thomas & Derwing, 2016).

Training lexical stress with nonwords yielded improved performance during training. Additionally, pre-to-post test results revealed significant improvements in the perception of real lexical stress minimal pairs for the experimentally trained group (but not for the control group). These findings highlight that training with nonwords is also effective for promoting significant learning gains for lexical stress perception which generalize to real words.

There is, however, another aspect of the training paradigm employed in the current study which warrants a special mention. Traditional HVPT procedures implement perceptual training through the use of a forced choice identification task (Lively et al., 1993; Logan et al., 1991). The authors of these papers have argued that the choice of training task is significant for promoting phonetic category formation and generalization to novel talkers and phonetic contexts. Indeed, the HVPT paradigm was developed to address some of the limitations associated with earlier discrimination-based training (Jamieson & Morosan, 1986; Strange & Dittman, 1984) which emphasized perceptual sensitivity to within-category differences and, therefore, had limited success in supporting category learning. Nevertheless, subsequent work investigating the benefit of using either identification or discrimination training tasks has reported similar learning outcomes for both training methods and no additional benefit of using both methods together (Shinohara & Iverson, 2018). In the context of the present study, Mandarin participants were trained on both identification and discrimination tasks and while isolating the impact of these two task types on perceptual learning falls outside of the scope of this study, learners showed consistent training improvements in the identification task, and more limited improvements across

training sessions in the discrimination task. Arguably, these findings do not constitute direct evidence for the effectiveness of either training task, but a correlational link between working memory capacity and changes in response latencies should be noted for the discrimination task. This link suggests that task demands may be memory-taxing which could also explain the small changes in accuracy scores and the relatively unchanged reaction times observed in the discrimination training task. Future research should attempt to disentangle the contribution of each training method (identification vs discrimination task type) to the acquisition of lexical stress to guide the design of training programs that would balance learning benefits and cognitive demands for more effective and engaging training.

Individual differences in auditory processing abilities and learning gains from short-term perceptual training

Results from the multiple regression analysis revealed no statistically significant relationship between variability in domain-general auditory processing abilities collected at Time 1 and learning gains from the HVPT lexical stress training. While auditory processing acuity was linked to lexical stress perception pre-training (at Time 1), contrary to my prediction, individual differences in Mandarin learners' auditory processing abilities were not predictive of improvements in their English lexical stress perception. The data showed that neither spectral nor temporal processing abilities at Time 1 correlated with learning gains at Time 2, suggesting that HVPT training, at least in the context of teaching English lexical stress, was effective for all participants regardless of individual differences in their auditory processing profiles. There are several possible explanations for this finding.

One possible reason might be that individual differences in auditory processing aptitude may play a role in explaining some of the variability in L2 speech acquisition depending on the learning context. In naturalistic settings, for instance, L2 gains have been linked with more precise auditory processing abilities for a number of speech outcomes (Saito et al, 2022a; Saito et al, 2020a; Sun et al., 2021). In the early stages of immersion, auditory processing has shown predictive power for prosody perception gains (Sun et al., 2021), and

improved segmental and lexical stress accuracy (Saito et al., 2020a) in the initial 5-8 months of immersion experience.

In terms of instructional contexts, however, studies have yielded mixed evidence about the role of auditory perception abilities and gains from perceptual training. Lengeris and Hazan (2010) reported that discrimination acuity for the F2 formant was predictive of vowel perception and production success after short-term HVPT training. Similarly, at the prosodic level, pitch identification accuracy has been linked with non-native lexical tone acquisition (Wong & Perrachione, 2007). However, in a study comparing the effectiveness of high- and low-variability phonetic training for the acquisition of the English /r-l/ contrast by native Japanese learners, researchers found no evidence that auditory processing performance at pre-test (conceptualised in terms of psychoacoustic discrimination and auditory-motor integration abilities) was predictive of learning success at post-test for either type of variability training (Brekelmans et al., 2022). Interestingly, in a similar training study employing three types of talker variability (low, medium, and high) to teach non-native tone perception, Sadakata and McQueen (2014) discovered a more nuanced interaction showing that perceptual aptitude can play differential effects on learner success depending on the training variability they receive. That is, increased variability in training stimuli had a hindering effect on improvements for learners with low perceptual aptitude, and facilitated learning for high-aptitude participants.

These findings, in combination with the results of the current experiment, suggest that depending on the targets of training and the characteristics of the training input, individual variability in domain-general auditory processing abilities may either play a role in facilitating or hindering targeted phonetic training, or in certain cases it may bear no relationship to the learning outcome. While the precise role of domain-general auditory processing in determining L2 speech acquisition success remains unclear, the findings from the current study move us closer to understanding its role in the targeted teaching of prosodic contrasts. In the context of short-term intensive training, individual differences in Mandarin speakers' pre-test auditory processing performance were not associated with relative gains in English lexical stress perception at post-test.

This finding carries important implications for both empirical and theoretical approaches to L2 phonetic teaching. For the narrow context of lexical stress acquisition, the current study has clearly demonstrated that high variability phonetic input experienced in a short intensive training can lead to significant improvements in English lexical stress perception for learners of all auditory processing abilities.

Chapter 4 General discussion

The current dissertation explored the acquisition of English lexical stress by native tonal language speakers. The experiments reported here aimed to advance our understanding of L2 prosodic acquisition by examining the potential variables that could explain individual differences in learner outcomes on the one hand, and design and test a perceptual training intervention for the teaching of L2 lexical stress. To accomplish this, I investigated the lexical stress acquisition of inexperienced native Mandarin learners of English in two experiments employing a cross-sectional, and a longitudinal experimental design. Basing my approach on the emerging literature on individual differences in auditory processing abilities and their role in determining L2 speech acquisition success (**Chapter 1**), in **Chapter 2** I asked if variability in learners' domain-general auditory processing abilities can explain individual differences in lexical stress processing (operationalized in terms of lexical stress perception and lexical stress cue weighting strategies) at a specific point in time (**Chapter 2**). In **Chapter 3**, I asked if individual differences in auditory processing were related to between-learner variability in perceptual gains after short-term training. Additionally, **Chapter 3** presented novel high variability prosodic training employing non-words produced by multiple talkers and featuring multiple phonetic and lexical contexts to train Mandarin learners' lexical stress perception. I sought to test the effectiveness of this instructional approach for improving perceptual performance and assess its potential effect on participants' dimensional weighting strategies. This final chapter summarizes the main findings of both studies and attempts to consolidate the theoretical and practical implications of this work to inform the direction of future research and teaching practices.

Summary of key findings

The literature in the field of second language acquisition (SLA) has seen a growing interest in the relationship between domain-general auditory processing abilities and L2 speech learning outcomes (Saito et al., 2022b; Saito et al., 2020a), offering emerging evidence that a learner's abilities to accurately process acoustic information are linked to L2 speech

acquisition success (Saito et al, 2022a; Saito et al, 2020a; Sun et al., 2021). Building on this research, in Experiment 1 (**Chapter 2**), I recruited native Mandarin speakers and collected a number of auditory processing measures tapping into learner's' abilities to discriminate, selectively attend to, and remember and reproduce information from different acoustic dimensions. I also assessed their lexical stress perception and cue weighting strategies. Experiment 1 revealed that perceptual acuity for the pitch dimension (measured in terms of discrimination thresholds) was a significant predictor of L2 English lexical stress processing in two separate ways. Firstly, Mandarin learners' with better pitch discrimination were able to perceive lexical stress more accurately. Secondly, sensitivity to pitch was also predictive of reliance on the pitch dimension for making lexical stress categorizations. Thus, participants with lower pitch thresholds (indicating better discrimination abilities) tended to give more perceptual weight to pitch when deciding if a word is stressed on the first or on the second syllable. Variability in cue weighting strategies was also related to individual differences in learners' rhythm memory and dimension selective attention for risetime. Both variables relating to individuals' temporal processing abilities were found to negatively correlate with reliance on pitch. In effect, participants who were better able to reproduce rhythmic sequences, and were better able to direct their attention to amplitude risetime and ignore irrelevant changes in other dimensions utilized durational information to a greater extent when categorizing lexical stress.

Next, **Chapter 3** presented a longitudinal investigation of lexical stress processing where I set out to design an effective, and easy to implement, training intervention for teaching lexical stress to L2 learners of English. Some of the native Mandarin speaking participants whose data was reported in Experiment 1 also participated in this short-term training study. Their Experiment 1 data served as a baseline for their domain-general auditory and English lexical stress processing abilities before the training. To create the experimental training of Experiment 2, I adapted the high variability phonetic training approach (a proven perceptual training technique characterised by lexical, phonetic, and talker variability and used successfully in L2 speech acquisition research; Lively et al., 1993; Logan et al., 1991) to the prosodic domain. After a 6-session training regimen, participants' lexical stress processing performance was retested. Experiment 2 showed that using the high variability phonetic training method to teach L2 prosody was effective. Mandarin learners in the experimental

condition improved their English lexical stress perception by 12.3% from pre- to post-test. This was a significant improvement in lexical stress performance from Time 1 to Time 2 compared to the control condition where performance remained statistically unchanged from pre- to post-test. Perceptual training, however, did not have an effect on participants' lexical stress cue weighting strategies with no statistically significant difference in reliance on pitch and duration information from Time 1 to Time 2.

Finally, in Experiment 2 I returned to the question of individual differences in learning outcomes by examining the relationship of IDs in domain-general auditory processing abilities collected at Time 1 and variability in participants' relative learning gains as computed at Time 2. Results showed that neither temporal nor spectral processing abilities were predictive of learning gains from the training. Importantly, the data provides strong evidence that the HVPT training for teaching lexical stress was effective for all participants and that individual variability in auditory processing abilities before the training did not have a statistically significant effect on learner outcomes.

Role of individual differences in domain-general auditory processing abilities in L2 speech acquisition

The current findings have important theoretical implications for existing auditory processing frameworks in L2 language acquisition, such as the auditory precision hypothesis (Kachlicka et al., 2019; Saito et al., 2022b). This framework proposes that learners' abilities to accurately perceive and encode acoustic information plays a crucial role in determining L2 acquisition success. The results presented in this dissertation support this hypothesis by showing that individual differences in domain-general auditory processing abilities are linked to L2 lexical stress performance. However, in the case of native Mandarin learners of English acquiring lexical stress, this relationship appears to be significant only at the initial stages of acquisition. Together, the findings from both studies in this dissertation indicate that individual differences in domain-general auditory processing abilities may play a role in determining L2 acquisition success depending on the different stages and contexts of

learning. In Experiment 1 I found that individual differences in Mandarin learners' lexical stress perception were linked to participants' more precise pitch discrimination abilities. The participants tested for this dissertation were inexperienced native Mandarin learners of English residing in China, who had no experience living abroad in immersion contexts, and had received classroom-based training in English. From a cross-sectional perspective, individual differences in their auditory processing abilities were shown to be associated with their L2 prosodic performance. However, Experiment 2 showed that in a longitudinal context after receiving short perceptual training, individual differences in participants' baseline auditory processing profiles before the training bore no relationship to variability in learning gains at post-test.

From the findings presented so far, the following can be gleaned: learners' auditory processing abilities were linked to variability in lexical stress perception at the initial stages of acquisition, but did not play a role in modulating the effects of high variability phonetic training. This contrasts with findings from several studies reporting a link between individual differences in auditory sensitivity and improvements shown post perceptual training in the cases of both segmental and suprasegmental acquisition (Lengeris & Hazan, 2010 for the role of F2 discrimination acuity in L2 vowel acquisition; Wong & Perrachione, 2007 for the role of pitch sensitivity in lexical tone acquisition). However, as the literature has shown, the evidence for the role of auditory processing aptitude in facilitating or hindering learning from short-term training is ambiguous. The effects of domain-general auditory processing abilities are not always observed when investigating the extent to which participants benefit from training. This was the case in Brekelmans et al. (2022) where the authors compared the effectiveness of high variability and low variability phonetic training on Japanese learners' acquisition of the English /r-l/ contrast. Importantly, they also collected a number of auditory processing and more general cognitive measures pre-training (tapping into dimension discrimination, auditory-motor integration, cognitive control and attention) to explore the effects of individual differences in aptitude on determining learning outcomes from the training. The study found no effect of individual differences in auditory processing on learners' improvements for either type of variability training (high or low).

These findings from the literature taken together with the results from Experiment 2 presented in this dissertation, suggest as a possible interpretation that when L2 short term training is concerned, individual differences in learners' auditory processing abilities may have an effect on learner success depending on the trained contrast. And while this a topic of ongoing research, not enough evidence is available yet to draw firm conclusions about the specific perceptual training contexts in which auditory processing abilities can have an impact on learning gains. More research is needed to explore the relationship of pre-training individual differences in auditory processing and variability in training outcomes for the acquisition of multiple non-native segmental and suprasegmental contrasts.

To further contextualize the role of individual differences in L2 speech acquisition, however, it's important to consider the full range of the L2 learning experience. The studies presented here have shown that auditory processing acuity for specific dimensions was associated with lexical stress identification accuracy at the early stages of acquisition but did not play a role in modulating learning outcomes from short-term training. The available literature, however, has shown that auditory processing abilities can also play a role in non-instructional settings where learning occurs in immersion or study-abroad contexts. More precise auditory processing abilities have been associated with greater improvements in both perception and production outcomes in the early stages of English-immersion experience (Saito et al, 2020a; Sun et al., 2021). An overview of the literature also seems to suggest that the existence of a relationship between individual differences in auditory processing and L2 speech outcomes may further depend on the target trained contrasts. Future work will need to establish when in the learning journey auditory processing abilities are most drawn upon to process L2 input and the contexts of acquisition where the precision of an individual's auditory processing can make the most difference to either facilitate or hinder learning. This research will have great potential to inform educational language practices by identifying the key stages of the acquisition journey where learners can benefit from receiving auditory training that will enhance their auditory processing abilities. This, in turn, can have positive effects on their L2 speech acquisition.

Role of individual differences in domain-general auditory processing abilities in shaping L2 cue weighting strategies

In Chapter 2 of this dissertation, I also investigated the relationship between individual differences in domain-general auditory processing abilities and learners' cue weighting strategies for lexical stress categorization. Previous studies have shown that native Mandarin learners of English experience prosodic-related challenges both in terms of perception and production, with their tonal linguistic experience resulting in a bias for overreliance on pitch information when processing prosody (Archibald, 1997; Hung, 1993; Juffs, 1990; Yu & Andruski, 2010; Zhang et al., 2008). But while this tendency to weight pitch more heavily when categorizing and producing English lexical stress is well-documented (Wang, 2008; Zhang et al., 2008; Yu & Andruski, 2010), the native-likeness of participants' cue weighting strategies was not the focus of the present work. Instead, I probed the potential link between individual differences in learners' auditory processing profiles and their cue reliance. Learners' dimensional weightings were collected through a categorization task whereby they had to indicate if a word was stressed on the first or the second syllable basing their judgements on different combinations of pitch and duration information either jointly cuing the same lexical stress location, or giving conflicting information. I found a clear link between performance on several auditory processing tests and reliance on pitch information relative to duration in lexical stress categorization.

Mandarin speakers who had better pitch discrimination acuity also tended to rely more heavily on pitch in their categorizations. This finding is important as it suggests that individuals better able to perceive subtle differences along a specific dimension, are also more likely to weight that dimension more heavily. This finding is in line with previous research which has investigated how individuals with congenital difficulties for processing pitch integrate acoustic information, showing they tend to rely more heavily on dimensions they are better able to perceive (Jasmin et al., 2020). Similar processing strategies have been observed when combining information from different sensory systems (Ernst & Banks, 2002), whereby a specific source of information (visual vs haptic) is weighted more heavily if the variance associated with processing that channel is lower. The authors proposed that

the human nervous system integrates input from different sources using computations similar to the maximum-likelihood estimation to minimise the variance in the outcome estimate by increasing reliance on less variable channels.

Two additional auditory processing abilities were also predictive of pitch cue weighting in the opposite direction. Firstly, better rhythm memory performance was associated with greater reliance on duration for category judgements. Participants who were better able to memorize and reproduce rhythmic patterns across time (Tierney et al., 2017) utilized duration information to a greater extent when making lexical stress categorizations. In this context too, listeners with better auditory processing abilities along a specific dimension (temporal), also tended to rely more heavily on that dimension.

Lastly, dimension selective attention for risetime also emerged as a significant predictor of participants' cue weightings. More specifically, better abilities to selectively attend to risetime (to the exclusion of simultaneous variability in other dimensions) were associated with greater duration weightings in lexical stress categorization. This finding is consistent with attentional theories of speech perception which propose that more perceptually salient dimensions tend to be given more perceptual weight in speech categorization (Francis & Nusbaum, 2002; Gordon et al., 1993; Holt et al., 2018).

The results in Chapter 2, therefore, provide evidence of a link between individual differences in domain-general auditory processing abilities and how listeners integrate information from multiple sources. At least in the context of lexical stress processing, Mandarin learners of English tended to rely more on the dimensions they could better perceive or selectively attend to. Given the redundant nature of the speech signal (Winter, 2014), these findings are compatible with the hypothesis that perceptual cue weighting strategies are shaped by several factors. One factor of note is extensive linguistic experience (such as exposure to a native language) which has been shown to lead to the development of perceptual strategies that prioritize dimensions which listeners have found to be highly informative (Francis & Nusbaum, 2002; Gordon et al., 1993; Holt et al., 2018; Jasmin et al., 2021). Similarly, dimensions that listeners are better able to focus their attention on (and suppress the interference of other simultaneously changing dimensions) also tend to up-

weight that dimension (Jasmin et al., 2021). Put another way, these findings suggest that one of the factors associated with variability in cue weighting strategies is the unique combination of strengths and weaknesses of an individuals' auditory processing system.

HVPT training for teaching L2 prosodic contrasts

One of the key components of this dissertation was the design and testing of a new phonetic training paradigm to target the acquisition of L2 lexical stress perception. Attaining suprasegmental proficiency is crucial for successful L2 acquisition and has been shown to significantly affect learners' comprehensibility and intelligibility (Isaacs & Trofimovich, 2012; Kang, 2010; Kang et al., 2010; Munro & Derwing, 1995), and predict oral proficiency success (Kang & Johnson, 2018).

In Chapter 3 I presented a new prosodic training paradigm adapted from the high variability training technique successfully used in the teaching of non-native consonants and vowels (Bradlow et al., 1997; Iverson & Evans, 2009; Lively et al., 1993; Logan et al., 1991; Nishi & Kewley-Port, 2007), and the teaching of lexical tones to non-tonal speakers (Perrachione et al., 2011; Sadakata & McQueen, 2014; Wang et al., 1999). The HVPT approach aims to expose learners to useful variability by training them on natural recordings produced by multiple talkers and featuring the target contrast in multiple lexical and phonetic contexts (Lively et al., 1993; Logan et al., 1991); elements which are all reflective of real-world listening conditions. Similarly, the training paradigm designed in the present study used disyllabic non-words produced by multiple talkers and in multiple phonetic contexts to train native Mandarin learners' lexical stress perception. The training was shown to be effective at significantly improving learners' lexical stress performance. Learners in the experimental condition were, on average, 12.3% more accurate at perceiving English lexical stress at the post-test compared to their pre-training accuracy. The lexical stress perception scores of

learners in the control condition (who received vocabulary training) remained unchanged from pre- to post-test.

Most significantly, the data has clearly demonstrated that all learners in the experimental condition improved after receiving training and learning gains were not tied to individual differences in their abilities to process auditory information. The results of Experiment 2, therefore, provide initial evidence that the HVPT training approach can be a highly effective tool for teaching non-native prosodic contrasts to L2 learners regardless of differences in their domain-general auditory processing abilities. Furthermore, the training paradigm designed for this study has the potential to be easily adapted to real-world L2 teaching settings. It can be a low-cost and relatively simple training to implement, and similarly to other computer-assisted L2 speech trainings, it can be accessed as part of classroom-based teaching, and remotely through the development of digital applications (Thomson, 2012a), or through the use of more affordable, subscription-based platforms for hosting the training stimuli. Nevertheless, more research is needed to investigate the benefits of HVPT lexical stress training for other L1-L2 language groups. A promising area for future research can examine the effectiveness of the HVPT prosodic paradigm for teaching lexical stress to native speakers of positionally-fixed stress languages (French, Turkish, Finnish), who have well-documented and persistent difficulties with the acquisition of L2 contrastive lexical stress (Dupoux et al., 1997, 2001; Lukyanchenko, et al., 2011; Peperkamp & Dupoux, 2008). Achieving greater understanding of the benefits and limitations of this new application of the HVPT technique will allow us to develop better training paradigms and make more reliable recommendations that will help optimize L2 suprasegmental outcomes for learners.

Methodological limitations

Several methodological limitations should be noted when interpreting the results of the studies presented in this dissertation. Firstly, due to time constraints, participants' English language proficiency was assessed using the LexTALE vocabulary test (Lemhöfer & Broersma, 2012), which has been shown to offer weaker correlations between LexTALE scores and other standardized proficiency measures, particularly for lower proficiency

learners (Puig-Mayenco et al., 2023). Consequently, the reported proficiency levels may not reliably reflect participants' actual English proficiency.

Another limitation relating to participants' English language competence concerns the selection of the minimal stress pairs used for testing in the current experiments. The target words fell within the first 4000-word (K4) families, which might have been unfamiliar to some participants. The task utilized isolated target words, presented without phrasal or sentential context, thereby minimizing reliance on semantic processing. However, while the primary objective of the lexical stress perception task was to assess participants' ability to peceive stress placement rather than their lexical knowledge, unfamiliarity with some target words may have influenced their performance. To address this limitation, future research should aim to recruit participants with a broader range of proficiency levels and explicitly investigate whether lexical knowledge impacts learners' ability to accurately perceive lexical stress location.

Additionally, this study tested learners' lexical stress perception using stimuli recorded by native speakers of Southern British English. However, information regarding the specific variety of English participants had been learning or were predominantly exposed to was not collected. Although lexical stress patterns in English minimal pairs are generally consistent across British and American varieties, subtle differences in vowel quality or prosodic realization may exist (e.g., Wells, 2008). While testing focused on stress placement rather than finer phonetic distinctions, future work should collect detailed information about learners' exposure to different English varieties to better understand how this might influence the perception and acquisition of lexical stress.

A further limitation of the study pertains to the control training condition. Visual-only vocabulary training was chosen to minimize the risk of cross-training transfer effects and to maintain participant engagement. However, this approach introduced a difference in the nature of the stimuli compared to the perceptual training condition, as the control group trained with real words, while the experimental group was exposed to non-words. Although this difference does not directly impact the study's focus on perceptual training, it represents a point of divergence between conditions that should be considered in

interpreting the results. Future studies may benefit from exploring alternative control paradigms that would maintain consistency in stimulus types.

Finally, while multiple regression analysis was used to examine the relationship between auditory processing abilities and lexical stress processing, mixed-effects modelling with item as a random effect might have been a more appropriate alternative. By including item as a random effect, mixed-effects modelling could provide more nuanced insights into how individual differences in auditory processing interact with specific stimuli. Nevertheless, the use of multiple regression was deemed sufficient to address the research questions in this study. Future research should consider employing mixed-effects modelling to better capture item-level variability and its potential effects.

Conclusion

Learning the phonological system of a second language in adulthood presents unique challenges for learners. And while research has mostly centered around addressing perceptual difficulties and remediating pronunciation errors at the segmental level, recent findings have highlighted the fundamental role that prosodic proficiency plays in impacting overall L2 learners' acquisition success. However, L2 speech acquisition outcomes are often characterized by large individual differences and recent studies have revealed a potential link between individual differences in domain-general auditory processing abilities and L2 attainment.

The work in this dissertation, therefore, aimed to contribute to a better understanding of L2 prosodic acquisition by focusing on the lexical stress acquisition of inexperienced native Mandarin learners of English. I approached this research goal from two perspectives: I carried out a cross-sectional and a longitudinal investigation which sought to examine the relationship between individual differences in learners' domain-general auditory perception and variability in English lexical stress processing. The longitudinal experiment was also a training study designed to enhance Mandarin learners' lexical stress perception. The training

was created by modifying an existing high variability phonetic training paradigm predominantly used for teaching segmental contrasts and adapting it to the prosodic domain. Chapter 2 of the current dissertation showed that individual differences in domain-general auditory abilities at the early stages of acquisition have an effect both on lexical stress perception and learners' cue weighting strategies. In Chapter 3, on the other hand, I did not find a relationship between individual differences in auditory processing abilities and learning gains from the HVPT training. In this sense, the two studies reported in this dissertation provide converging evidence that individual differences in domain-general auditory processing abilities can explain some of the variability in L2 speech acquisition success in certain cases, however, their role may depend on the context and stage of L2 acquisition.

Chapter 3 also showed that the modified HVPT training was effective at teaching L2 prosody and trained participants significantly improved their lexical stress perception accuracy at the post-test. Importantly, the training was effective for all trained Mandarin participants regardless of their auditory processing aptitudes. This finding highlights the efficacy of the paradigm and makes it a good candidate for real-world teaching application. Overall, the findings reported here advance our understanding of L2 lexical stress acquisition and hold important theoretical and practical significance for L2 speech research and teaching. The current chapter has also attempted to identify key open questions and outline important avenues for further research that can bring us closer to forming a more comprehensive understanding of the contexts when learner-specific aptitude factors have an effect on L2 speech acquisition.

Future research should be directed towards examining the wider effectiveness of this modified HVPT training paradigm both beyond the perception domain and in different learning contexts. One important area of future investigation concerns the potential transfer of perceptual learning to improvements in production abilities. Future studies should also attempt to determine if learning gains will be retained in time by incorporating delayed post-tests to assess long-term retention. It would also be interesting to test this prosodic paradigm with more experienced L2 learners of English (both learners in instructional settings, and those in immersion contexts), as well as assess its potential

application for the training of children and young learners who experience lexical stress perception difficulties in their L1.

References

Abrahamsson, N., & Hyltenstam, K. (2008). THE ROBUSTNESS OF APTITUDE EFFECTS IN NEAR-NATIVE SECOND LANGUAGE ACQUISITION. Studies in Second Language Acquisition, 30(4), 481–509. https://doi.org/10.1017/S027226310808073X

Adams, C., & Munro, R. R. (1978). In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English. Phonetica, 35(3), 125–156. https://doi.org/10.1159/000259926

Adolphs, S., & Schmitt, N. (2003). Lexical Coverage of Spoken Discourse. Applied Linguistics, 24(4), 425–438. https://doi.org/10.1093/applin/24.4.425

Ahissar, M., Protopapas, A., Reid, M., & Merzenich, M. M. (2000). Auditory Processing Parallels Reading Abilities in Adults. Proceedings of the National Academy of Sciences - PNAS, 97(12), 6832–6837. https://doi.org/10.1073/pnas.97.12.6832

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The Relationship Between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentais, Prosody, and Syllable Structure. Language Learning, 42(4), 529–555. https://doi.org/10.1111/j.1467-1770.1992.tb01043.x

Anisfeld, M., Bogo, N., & Lambert, W. E. (1962). Evaluational reactions to accented English speech. Journal of Abnormal and Social Psychology, 65(4), 223–231. https://doi.org/10.1037/h0045060

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., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2020). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods, 53*(4), 1407–1425. https://doi.org/10.3758/s13428-020-01501-5

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). 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

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. Journal of Phonetics, 32(2), 233–250. https://doi.org/10.1016/S0095-4470(03)00036-6

Archibald, J. (1992). Transfer of L1 Parameter Settings: Some Empirical Evidence from Polish Metrics. Canadian Journal of Linguistics, 37(3), 301–340. https://doi.org/10.1017/S0008413100019903

Archibald, J. (1993). The learnability of English metrical parameters by adult spanish speakers. IRAL : International Review of Applied Linguistics in Language Teaching,31(2), 129. Retrieved from https://www.proquest.com/scholarly-journals/learnability-english-metrical-parameters-adult/docview/1300505204/se-2

ARCHIBALD, J. (1997). The acquisition of English stress by speakers of nonaccentual languages: lexical storage versus computation of stress. _Linguistics_, _35_(1), 167-182. https://doi.org/10.1515/ling.1997.35.1.167

Archibald, J. (1998). SECOND LANGUAGE PHONOLOGY, PHONETICS, AND TYPOLOGY. *Studies in Second Language Acquisition*, *20*(2), 189–211. https://doi.org/10.1017/S0272263198002046

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia. A group study of adults afflicted with a music-specific disorder. Brain (London, England : 1878), 125(2), 238–251. https://doi.org/10.1093/brain/awf028

Barriuso, T. A., & Hayes-Harb, R. (2018). High Variability Phonetic Training as a Bridge From Research to Practice. The CATESOL Journal, 30, 177. http://www.catesoljournal.org/wp-content/uploads/2018/03/CJ30.1_barriuso.pdf

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

Beckman, M. E. (1986). Stress and Non-Stress Accent. In Stress and Non-Stress Accent. DE GRUYTER. https://doi.org/10.1515/9783110874020

Beckman, M. E., Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. Keating (Ed.), Phonological structure and phonetic form: Papers in Laboratory Phonology III (pp. 7–33). Cambridge, England: Cambridge University Press.

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The Influence of Linguistic Experience on the Cognitive Processing of Pitch in Speech and Nonspeech Sounds. *Journal of Experimental Psychology. Human Perception and Performance*, *32*(1), 97–103. https://doi.org/10.1037/0096-1523.32.1.97

Berg, B. G. (1989). Analysis of weights in multiple observation tasks. Journal of the Acoustical Society of America, 86(5), 1743–1746. https://doi.org/10.1121/1.398605

Best, C.T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), Speech Perception and Linguistic Experience. Theoretical and Methodological Issues (pp. 171-203). Baltimore: York Press.

Best, C. T., Halle, P., Bohn, O.-S., & Faber, A. (2003): Cross-language perception of nonnative vowels: phonological and phonetic effects of listeners' native languages, In ICPhS-15, 2889-2892. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/p15_2889.html

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. Journal of the Acoustical Society of America, 109(2), 775–794. https://doi.org/10.1121/1.1332378

Best, C.T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. Journal of Phonetics, 20, 305-330. https://doi.org/10.1016/S0095-4470(19)30637-0

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception Commonalities and complementarities. *Language Learning and Language Teaching*, *17*, 13–34. https://doi.org/10.1075/lllt.17.07bes

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone Language Speakers and Musicians Share Enhanced Perceptual and Cognitive Abilities for Musical Pitch: Evidence for Bidirectionality between the Domains of Language and Music. PLoS ONE, 8(4). https://doi.org/10.1371/journal.pone.0060676

Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In *Language learning and language teaching* (pp. 99–116). https://doi.org/10.1075/lllt.17.12bir

Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquière, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities*, *32*(2), 560–570. https://doi.org/10.1016/j.ridd.2010.12.020

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

Bolinger, D. L. (1958). A Theory of Pitch Accent in English. WORD, 14(2–3), 109–149. https://doi.org/10.1080/00437956.1958.11659660

Bolinger, D. L. (1961). Contrastive Accent and Contrastive Stress. Language, 37(1), 83. https://doi.org/10.2307/411252

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). AGE AND ULTIMATE ATTAINMENT IN THE PRONUNCIATION OF A FOREIGN LANGUAGE. Studies in Second Language Acquisition, 19(4), 447–465. https://doi.org/10.1017/s0272263197004026

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese Listeners to Identify English /r/ and /l/: Long-Term Retention of Learning in Perception and Production. *Perception & Psychophysics*, *61*(5), 977–985. https://doi.org/10.3758/bf03206911

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. The Journal of the Acoustical Society of America, 101(4), 2299–2310. https://doi.org/10.1121/1.418276

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, 104352-. https://doi.org/10.1016/j.jml.2022.104352

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7–9), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Brennan, E. M., & Brennan, J. S. (1981). Accent Scaling and Language Attitudes: Reactions to Mexican American English Speech. Language and Speech, 24(3), 207–221. https://doi.org/10.1177/002383098102400301

Carlet, A., & Cebrian, J. (2019). Assessing the effect of perceptual training on L2 vowel identification, generalization and long-term effects. . In A. M. Nyvad, M. Hejná, A. Højen, A. Bothe Jespersen , & M. Hjortshøj Sørensen (Eds.), A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn. (pp. 91-119). Aarhus University Press. https://doi.org/10.7146/aul.322.218

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), Individual differences and universals in language learning aptitude, (p.83-118). Rowley, MA: Newbury House.

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

Chao, Y. R. (1968). A Grammar of Spoken Chinese. University of California Press, Berkeley, CA.

Chen, M. (1970). Vowel Length Variation as a Function of the Voicing of the Consonant Environment. Phonetica, 22(3), 129-159. https://doi.org/10.1159/000259312

Chen, Y., Robb, M. P., Gilbert, H. R., & Lerman, J. W. (2001). A study of sentence stress production in Mandarin speakers of American English. The Journal of the Acoustical Society of America, 109(4), 1681–1690. https://doi.org/10.1121/1.1356023

Choi, S., & Kang, O. (2023). The roles of suprasegmental features in assessing paired speaking tasks in high-stakes language assessment. System (Linköping), 119, 103183-. https://doi.org/10.1016/j.system.2023.103183

Choudhury, N., & Benasich, A. A. (2011). Maturation of auditory evoked potentials from 6 to 48 months: Prediction to 3 and 4 year language and cognitive abilities. *Clinical Neurophysiology*, *122*(2), 320–338. https://doi.org/10.1016/j.clinph.2010.05.035

Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. Journal of speech, language, and hearing research : JSLHR, 57(4), 1468–1479. https://doi.org/10.1044/2014_JSLHR-L-13-0279

Christensen, L. A., & Humes, L. E. (1997). Identification of multidimensional stimuli containing speech cues and the effects of training. *The Journal of the Acoustical Society of America*, *102*(4), 2297–2310. https://doi.org/10.1121/1.419639

Cobb, T. (2021). The complete lexical tutor. http://www.lextutor.ca/vp/

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (Rev. ed.). Taylor and Francis. https://doi.org/10.4324/9780203771587

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of Lexical Stress on Lexical Access in English: Evidence from Native and Non-native Listeners. *Language and Speech*, *45*(3), 207–228. https://doi.org/10.1177/00238309020450030101

Creel, S. C., Weng, M., Fu, G., Heyman, G. D., & Lee, K. (2018). Speaking a tone language enhances musical pitch perception in 3–5-year-olds. *Developmental Science*, *21*(1), 1–7. https://doi.org/10.1111/desc.12503

Crystal, D. (2009). A dictionary of linguistics and phonetics (6th ed.). Wiley.

Cutler, A. (1986). Forbear is a Homophone: Lexical Prosody Does Not Constrain Lexical Access. Language and Speech, 29(3), 201-220. https://doi.org/10.1177/002383098602900302

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

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. Computer Speech & Language, 2(3), 133–142. https://doi.org/10.1016/0885-2308(87)90004-0

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. Journal of Experimental Psychology: Human Perception and Performance, 14(1), 113–121. https://doi.org/10.1037/0096-1523.14.1.113

Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2009). Singing in congenital amusia. *The Journal of the Acoustical Society of America*, *126*(1), 414–424. https://doi.org/10.1121/1.3132504

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422–433. https://doi.org/10.3758/BF03214546

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. M. (2000). THE ROBUSTNESS OF CRITICAL PERIOD EFFECTS IN SECOND LANGUAGE ACQUISITION. *Studies in Second Language Acquisition*, *22*(4), 499–533. https://doi.org/10.1017/S0272263100004022

Deroche, M. L. D., Lu, H. P., Kulkarni, A. M., Caldwell, M., Barrett, K. C., Peng, S. C., Limb, C. J., Lin, Y. S., & Chatterjee, M. (2019). A tonal-language benefit for pitch in normally-hearing and cochlear-implanted children. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-018-36393-1

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

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in Favor of a Broad Framework for Pronunciation Instruction. Language Learning, 48(3), 393–410. https://doi.org/10.1111/0023-8333.00047

Derwing, T. M., & Rossiter, M. J. (2003). The Effects of Pronunciation Instruction on the Accuracy, Fluency, and Complexity of L2 Accented Speech. Applied Language Learning, 13, 1-17. https://www.researchgate.net/publication/234570703_The_Effects_of_Pronunciation_Instruction_on_the_Accuracy_Fluency_and_Complexity_of_L2_Accented_Speech

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, *22*(6), 680–689. https://doi.org/10.1016/j.lindif.2012.05.005

Dickerson, W. B. (2000 March). Covert rehearsal as a bridge to accurate fluency. Paper presented at International TESOL, Vancouver, BC, Canada.

Dickerson, W. B. (2004). Stress in the speech stream: The rhythm of spoken English. Champaign, IL: University of Illinois Press.

Dickerson, W. B. (2015). Using Orthography to Teach Pronunciation. In The Handbook of English Pronunciation (pp. 488–504). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118346952.ch27

DiStefano, C., Zhu, M., & Mîndrilã, D. (2009). Understanding and Using Factor Scores: Considerations for the Applied Researcher. Practical Assessment, Research & Evaluation, 14, 20. https://www.proquest.com/scholarly-journals/understanding-using-factor-scores-considerations/docview/2366805383/se-2

Doherty, K. A., & Turner, C. W. (1996). Use of a correlational method to estimate a listener's weighting function for speech. The Journal of the Acoustical Society of America, 100(6), 3769–3773. https://doi.org/10.1121/1.417336

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

Douglas, S., & Willatts, P. (1994). The relationship between musical ability and literacy skills. Journal of Research in Reading, 17(2), 99–107. https://doi.org/10.1111/j.1467-9817.1994.tb00057.x

Duanmu, S. (2007). The Phonology of Standard Chinese (2nd ed.). Oxford University Press.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in French? Journal of Memory and Language, 36(3), 406–421. https://doi.org/10.1006/jmla.1996.2500

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness." Journal of the Acoustical Society of America, 110(3,Pt1), 1606–1618. https://doi.org/10.1121/1.1380437

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': the case of French learners of Spanish. Cognition, 106(2), 682–706. https://doi.org/10.1016/j.cognition.2007.04.001

Eady, S. J. (1982). Differences in the F0 Patterns of Speech: Tone Language Versus Stress Language. Language and Speech, 25(1), 29-42. https://doi.org/10.1177/002383098202500103

Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. The Journal of the Acoustical Society of America, 80(2), 402–415. https://doi.org/10.1121/1.394091

Epsy-Wilson C. Y. (1992). Acoustic measures for linguistic features distinguishing the semivowels/wjrl/in American English. The Journal of the Acoustical Society of America, 92(2 Pt 1), 736–757. https://doi.org/10.1121/1.403998

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fayer, J. M., & Krasinski, E. (1987). Native and Nonnative Judgments of Intelligibility and Irritation. Language Learning, 37(3), 313–326. https://doi.org/10.1111/j.1467-1770.1987.tb00573.x

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. Journal of the Acoustical Society of America, 97(3), 1893–1904. https://doi.org/10.1121/1.412063

FIELD, J. (2005). Intelligibility and the Listener: The Role of Lexical Stress. TESOL Quarterly, 39(3), 399–423. https://doi.org/10.2307/3588487

Flege, J. E. (1995a). Second Language Speech Learning: Theory, Findings, and Problems. Speech Perception and Linguistic Experience: Issues in Cross-Language Research, 233–277.

Flege, J. E. (1995). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, *16*(4), 425–442. https://doi.org/10.1017/S0142716400066029

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception [Bookitem]. In Phonetics and Phonology in Language Comprehension and Production (Originally publis..., Vol. 6, pp. 319–358). DE GRUYTER MOUTON. https://doi.org/10.1515/9783110895094.319

Flege, J. E., & Bohn, O.-S. (1989). An Instrumental Study of Vowel Reduction and Stress Placement in Spanish-Accented English. *Studies in Second Language Acquisition*, *11*(1), 35–62. https://doi.org/10.1017/S0272263100007828

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). chapter, Cambridge: Cambridge University Press.

Flege, J.E., Bohn, O.-S. and Jang, S. (1997) Effects of Experience on Non-Native Speakers' Production and Perception of English Vowels. Journal of Phonetics, 25, 437-470. https://doi.org/10.1006/jpho.1997.0052

Flege, J. E., & Hillenbrand, J. (1986). Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. The Journal of the Acoustical Society of America, 79(2), 508–517. https://doi.org/10.1121/1.393538

Flege, J. E., & Liu, S. (2001). THE EFFECT OF EXPERIENCE ON ADULTS' ACQUISITION OF A SECOND LANGUAGE. Studies in Second Language Acquisition, 23(4), 527–552. https://doi.org/10.1017/S0272263101004041

Flege, J. E., & MacKay, I. R. A. (2004). PERCEIVING VOWELS IN A SECOND LANGUAGE. Studies in Second Language Acquisition, 26(1), 1–34. https://doi.org/10.1017/S0272263104026117

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Effects of age of second-language learning on the production of English consonants. Speech Communication, 16(1), 1–26. https://doi.org/10.1016/0167-6393(94)00044-B

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

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age Constraints on Second-Language Acquisition. Journal of Memory and Language, 41(1), 78–104. https://doi.org/10.1006/jmla.1999.2638

Fokes, J., Bond, Z. & Steinberg, M. (1984). Patterns of English Word Stress by Native and Non-native Speakers. In A. Cohen & M. Broecke (Ed.), Proceedings of the Tenth International Congress of Phonetic Sciences (pp. 682-686). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110884685-111

Fonagy, I. (1978). A New Method of Investigating the Perception of Prosodic Features. Language and Speech, 21(1), 34-49. https://doi.org/10.1177/002383097802100102

Fox, R. A., Flege, J. E., & Munro, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: a multidimensional scaling analysis. The Journal of the Acoustical Society of America, 97(4), 2540–2551. https://doi.org/10.1121/1.411974

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics*, *62*(8), 1668–1680. https://doi.org/10.3758/BF03212164

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. Journal of Phonetics, 36(2), 268–294. https://doi.org/10.1016/j.wocn.2007.06.005

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. Journal of Experimental Psychology: Human Perception and Performance, 28(2), 349–366. https://doi.org/10.1037/0096-1523.28.2.349

Fry, D. B. (1955). Duration and Intensity as Physical Correlates of Linguistic Stress. The Journal of the Acoustical Society of America, 27(4), 765–768. https://doi.org/10.1121/1.1908022

Fry, D. B. (1958). Experiments in the Perception of Stress. Language and Speech, 1(2), 126–152. https://doi.org/10.1177/002383095800100207

Fry, D. B. (1965). The Dependence of Stress Judgments on Vowel Formant Structure. Phonetic Sciences, 311, 306–311. https://doi.org/10.1159/000426965

Fu, Q. J., Zeng, F. G., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *The Journal of the Acoustical Society of America*, *104*(1), 505–510. https://doi.org/10.1121/1.423251

Fuhrmeister, P., & Myers, E. B. (2017). Non-native phonetic learning is destabilized by exposure to phonological variability before and after training. The Journal of the Acoustical Society of America, 142(5), EL448–EL454. https://doi.org/10.1121/1.5009688

Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. European Journal of Social Psychology, 42(1), 120–133. https://doi.org/10.1002/ejsp.862

Gallego, J. C. (1990). The Intelligibility of Three Nonnative English-Speaking Teaching Assistants: An Analysis of Student-Reported Communication Breakdowns. Issues in Applied Linguistics, 1(2). https://doi.org/10.5070/L412004998

Gandour, J. (1978). "The perception of tone," in Tone: A Linguistic Survey, edited by V. Fromkin Academy, New York, pp. 41–76.

Gandour, J. (1983). Tone perception in Far Eastern languages. Journal of Phonetics, 11(2), 149–175. https://doi.org/10.1016/s0095-4470(19)30813-7

Gay, T. (1978). Physiological and Acoustic Correlates of Perceived Stress. Language and Speech, 21(4), 347-353. https://doi.org/10.1177/002383097802100409

Ghaffarvand Mokari, P., & Werner, S. (2019). On the Role of Cognitive Abilities in Second Language Vowel Learning. Language and Speech, 62(2), 260–280. https://doi.org/10.1177/0023830918764517

Gibson, L. Y., Hogben, J. H., & Fletcher, J. (2006). Visual and auditory processing and component reading skills in developmental dyslexia. *Cognitive Neuropsychology*, *23*(4), 621–642. https://doi.org/10.1080/02643290500412545

Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S., & Wicha, N. Y. Y. (2011). Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. Frontiers in Psychology, 2(AUG), 1–12. https://doi.org/10.3389/fpsyg.2011.00146

Goffman, L., & Malin, C. (1999). Metrical effects on speech movements in children and adults. Journal of speech, language, and hearing research : JSLHR, 42(4), 1003–1015. https://doi.org/10.1044/jslhr.4204.1003

Goo, J., & Mackey, A. (2013). THE CASE AGAINST THE CASE AGAINST RECASTS. *Studies in Second Language Acquisition*, *35*(1), 127–165. https://doi.org/10.1017/S0272263112000708

Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. Cognitive Psychology, 25(1), 1–42. https://doi.org/10.1006/cogp.1993.1001

Gordon, P. C., Keyes, L., & Yung, Y. F. (2001). Ability in perceiving nonnative contrasts: Performance on natural and synthetic speech stimuli. Perception and Psychophysics, 63(4), 746–758. https://doi.org/10.3758/BF03194435

Gordon, M. & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. Linguistics Vanguard, 3(1), 20170007. https://doi.org/10.1515/lingvan-2017-0007

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., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & Scott, S. K. (2002). Amplitude Envelope Onsets and Developmental Dyslexia: A New Hypothesis. Proceedings of the National Academy of Sciences - PNAS, 99(16), 10911–10916. https://doi.org/10.1073/pnas.122368599

Goswami, U., Wang, H.-L. S., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2011). Language-universal Sensory Deficits in Developmental Dyslexia: English, Spanish, and Chinese. *Journal of Cognitive Neuroscience*, *23*(2), 325–337. https://doi.org/10.1162/jocn.2010.21453

Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. Journal of Phonetics, 12(2), 91–114. https://doi.org/10.1016/S0095-4470(19)30858-7

Gottfried, T. L., Jenkins, J. J., & Strange, W. (1985). Categorial discrimination of vowels produced in syllable context and in isolation. Bulletin of the Psychonomic Society, 23(2), 101–104. https://doi.org/10.3758/BF03329794

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r." Neuropsychologia, 9(3), 317–323. https://doi.org/10.1016/0028-3932(71)90027-3

Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. Journal of Phonetics, 12(2), 91–114. https://doi.org/10.1016/S0095-4470(19)30858-7

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311–343. https://doi.org/10.1177/0267658312461497

Grice, M. (2001). Intonation Systems: A Survey of Twenty Languages [Review of *Intonation Systems: A Survey of Twenty Languages*]. *Journal of Linguistics*, *37*(3), 619–625. Cambridge University Press. https://doi.org/10.1017/S0022226701251354

Grzegorz Dogil, & Briony Williams. (1999). The phonetic manifestation of word stress. In Eurotyp. Mouton de Gruyter. https://doi.org/10.1515/9783110197082.1.273

Guion, S. G., Clark, J. J., Harada, T., & Wayland, R. P. (2003). Factors Affecting Stress Placement for English Nonwords include Syllabic Structure, Lexical Class, and Stress Patterns of Phonologically Similar Words. Language and Speech, 46(4), 403–427. https://doi.org/10.1177/00238309030460040301

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000a). An investigation of current models of second language speech perception: The case of Japanese adults'

perception of English consonants. The Journal of the Acoustical Society of America, 107(5 I), 2711–2724. https://doi.org/10.1121/1.428657

GUION, S. G., FLEGE, J. E., LIU, S. H., & YENI-KOMSHIAN, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, *21*(2), 205–228. https://doi.org/10.1017/S0142716400002034

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. Language Learning and Language Teaching, 17, 57–77.

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. Journal of Phonetics, 32(3), 395–421. https://doi.org/10.1016/S0095-4470(03)00016-0

Hammond, R. M. (1986). Error analysis and the natural approach to teaching foreign languages. Lenguas Modernas, 3, 129–139. https://revistas.uchile.cl/index.php/LM/article/download/45871/47895/

Hazan, V., & Kim, Y. H. (2010). "Can we predict who will benefit from computer-based phonetic training?" in Online Proceedings of the INTERSPEECH 2010 Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology (L2WS 2010), Tokyo, Japan.

Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. The Journal of the Acoustical Society of America, 116(4 I), 2384–2394. https://doi.org/10.1121/1.1783351

Hirata, Y., Whitehurst, E., & Cullings, E. (2007). Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates. The Journal of the Acoustical Society of America, 121(6), 3837–3845. https://doi.org/10.1121/1.2734401

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. The Journal of the Acoustical Society of America, 119(5), 3059–3071. https://doi.org/10.1121/1.2188377

Holt, L. L., & Lotto, A. J. (2008). Speech Perception Within an Auditory Cognitive Science Framework. *Current directions in psychological science*, *17*(1), 42–46. https://doi.org/10.1111/j.1467-8721.2008.00545.x

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

House, A.S. (1961). On Vowel Duration in English. Journal of the Acoustical Society of America, 33, 1174-1178. https://doi.org/10.1121/1.1908941

Howell, P. (1993). Cue trading in the production and perception of vowel stress. Journal of the Acoustical Society of America, 94(4), 2063–2073. https://doi.org/10.1121/1.407479

Howie, J. M. (1976). Acoustical Studies of Mandarin Vowels and Tones. Cambridge University Press, Cambridge.

Hung, T. T. N. (1993). The role of phonology in the teaching of pronunciation to bilingual students. Language, Culture and Curriculum, 6(3), 249–256. https://doi.org/10.1080/07908319309525155

Idemaru, K., & Holt, L. L. (2013). The developmental trajectory of children's perception and production of English /r/-/l/. *The Journal of the Acoustical Society of America*, *133*(6), 4232–4246. https://doi.org/10.1121/1.4802905

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

Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2012). Can native Japanese listeners learn to differentiate /r-l/ on the basis of F3 onset frequency? (vol 15, pg 255, 2012). *Bilingualism (Cambridge, England)*, *15*(2), 434–435. https://doi.org/10.1017/S1366728912000041

Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. Journal of Phonetics, 39(4), 571–584. https://doi.org/10.1016/j.wocn.2011.03.003

Isaacs, T., & Trofimovich, P. (2012). DECONSTRUCTING COMPREHENSIBILITY. *Studies in Second Language Acquisition*, *34*(3), 475–505. https://doi.org/10.1017/S0272263112000150

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. The Journal of the Acoustical Society of America, 126(2), 866–877. https://doi.org/10.1121/1.3148196

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. (1994). Tests of the perceptual magnet effect for American English /r/ and /l/. The Journal of the Acoustical Society of America, 95(5), 2976–2976. https://doi.org/10.1121/1.408983

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., 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

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. Applied Psycholinguistics, 33(1), 145–160. https://doi.org/10.1017/S0142716411000300

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /{eth}/-/θ/ contrast by francophones. *Perception & Psychophysics*, *40*(4), 205–215. https://doi.org/10.3758/BF03211500

Jamieson, D. G., & Morosan, D. E. (1989). Training New, Nonnative Speech Contrasts: A Comparison of the Prototype and Perceptual Fading Techniques. *Canadian Journal of Psychology*, *43*(1), 88–96. https://doi.org/10.1037/h0084209

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, 104481–104481. https://doi.org/10.1016/j.cognition.2020.104481

Jasmin, K., Tierney, A., Obasih, C., & Holt, L. (2023). Short-term perceptual reweighting in suprasegmental categorization. Psychonomic bulletin & review, 30(1), 373–382. https://doi.org/10.3758/s13423-022-02146-5

Juffs, A. (1990). TONE, SYLLABLE STRUCTURE AND INTERLANGUAGE PHONOLOGY: CHINESE LEARNERS' STRESS ERRORS. International Review of Applied Linguistics in Language Teaching, 28(2), 99-118. https://doi.org/10.1515/iral.1990.28.2.99

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

Kager, R. (2007). Feet and metrical stress. In P. Lacy (Ed.), The Cambridge Handbook of Phonology (Cambridge Handbooks in Language and Linguistics, pp. 195-228). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486371.010

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). Sensitivity to amplitude envelope rise time in infancy and vocabulary development at 3 years: A significant relationship. *Developmental Science*, *22*(6), e12836-n/a. https://doi.org/10.1111/desc.12836

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. System, 38(2), 301-315. https://doi.org/10.1016/j.system.2010.01.005

Kang, O., & Johnson, D. (2018). The roles of suprasegmental features in predicting English oral proficiency with an automated system. Language Assessment Quarterly, 15(2), 150–168. https://doi.org/10.1080/15434303.2018.1451531

KANG, O., RUBIN, D., & PICKERING, L. (2010). Suprasegmental Measures of Accentedness and Judgments of Language Learner Proficiency in Oral English. The Modern Language Journal (Boulder, Colo.), 94(4), 554–566. https://doi.org/10.1111/j.1540-4781.2010.01091.x

Kawahara, H., & Irino, T. (2005). Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. *Speech Separation by Humans and Machines*, 167–180. https://doi.org/10.1007/0-387-22794-6_11

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. Behavior Research Methods, 42(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

Kewley-Port, D. (2001). Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context, and training. *The Journal of the Acoustical Society of America*, *110*(4), 2141–2155. https://doi.org/10.1121/1.1400737

Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *The Journal of the Acoustical Society of America*, *122*(1), 418–435. https://doi.org/10.1121/1.2743154

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. Perception & Psychophysics, 50(2), 93–107. https://doi.org/10.3758/BF03212211

Kuhl P. K. (2000). A new view of language acquisition. Proceedings of the National Academy of Sciences of the United States of America, 97(22), 11850–11857. https://doi.org/10.1073/pnas.97.22.11850

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews. Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. Science (New York, N.Y.), 255(5044), 606–608. https://doi.org/10.1126/science.1736364

Ladd, D.R., & Morton, R. (1997). The perception of intonational emphasis: continuous or categorical? Journal of Phonetics, 25, 313-342. https://doi.org/10.1006/jpho.1997.0046

Ladefoged. (2014). *A Course in phonetics* (Seventh edition.). Cengage Learning.

Lamb, S. J., & Gregory, A. H. (1993). The Relationship between Music and Reading in Beginning Readers. Educational Psychology (Dorchester-on-Thames), 13(1), 19–27. https://doi.org/10.1080/0144341930130103

LAMBACHER, S. G., MARTENS, W. L., KAKEHI, K., MARASINGHE, C. A., & MOLHOLT, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*(2), 227–247. https://doi.org/10.1017/S0142716405050150

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. Second Language Research, 24(1), 35–63. https://doi.org/10.1177/0267658307082981

Leather, J. (1987). FO Pattern inference in the perceptual acquisition of second-language tone. In James, A. & Leather, J. (Eds.), Sound patterns in second language acquisition (pp. 59–80). Dordrecht: Foris. https://doi.org/10.1515/9783110878486-005

Lee, M., & Révész, A. (2021). The role of working memory in attentional allocation and grammatical development under textually-enhanced, unenhanced and no captioning conditions. https://www.jpll.org/index.php/journal/article/view/leerevesz/leerevesz

Lehiste, I. (1970). Suprasegmentals. Cambridge, MA: MIT Press.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: a quick and valid Lexical Test for Advanced Learners of English. Behavior research methods, 44(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

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

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2), 467–477. https://doi.org/10.1121/1.1912375

Li, S. (2016). THE CONSTRUCT VALIDITY OF LANGUAGE APTITUDE: A Meta-Analysis. *Studies in Second Language Acquisition*, *38*(4), 801–842. https://doi.org/10.1017/S027226311500042X

Liberman, M. (1975). The intonational system of English. Massachusetts Institute of Technology; Cambridge: 1975. (Unpublished doctoral dissertation).

Lieberman, P. (1960). Some Acoustic Correlates of Word Stress in American English. Journal of the Acoustical Society of America, 32(4), 451–454. https://doi.org/10.1121/1.1908095

Lisker, L. (1986). "Voicing" in English: A Catalogue of Acoustic Features Signaling /b/ Versus /p/ in Trochees. Language and Speech, 29(1), 3-11. https://doi.org/10.1177/002383098602900102

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. Brain (London, England : 1878), 133(6), 1682–1693. https://doi.org/10.1093/brain/awq089

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. Language and speech, 47(Pt 2), 109–138. https://doi.org/10.1177/00238309040470020101

Lively, Scott E., Logan, John, Pisoni, D. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. The Journal of the Acoustical Society of America, 94(1242). https://doi.org/https://doi.org/10.1121/1.408177

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese Listeners To Identify English /R/ And /1/: A First Report. Journal of the Acoustical Society of America, 89(2), 874–886. https://doi.org/10.1121/1.1894649

Lukyanchenko, A., Idsardi, WJ., Jiang, N. (2011). The role of L1 in processing of nonnative prosodic contrasts. Selected Proceedings of the Second Languge Research Forum 2010, 50–62.

Lutfi, R. A. (1992). Informational processing of complex sound. III: Interference. Journal of the Acoustical Society of America, 91(6), 3391–3401. https://doi.org/10.1121/1.402829

Mattys, S. L. (2000). The perception of primary and secondary stress in English. Perception & Psychophysics, 62(2), 253–265. https://doi.org/10.3758/BF03205547

Mattys, S. L., & Samuel, A. G. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. Journal of Memory and Language, 36(1), 87–116. https://doi.org/10.1006/jmla.1996.2472

Mattys, S., & Samuel, A. G. (2000). Implications of Stress-Pattern Differences in Spoken-Word Recognition. Journal of Memory and Language, 42(4), 571–596. https://doi.org/10.1006/jmla.1999.2696

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

MERCIER, J., PIVNEVA, I., & TITONE, D. (2014). Individual differences in inhibitory control relate to bilingual spoken word processing. *Bilingualism (Cambridge, England)*, *17*(1), 89–117. https://doi.org/10.1017/S1366728913000084

Miyake, A., & Friedman, N. P. (1998). Individual Differences in Second Language Proficiency: Working Memory as Language Aptitude. In A. F. Healy, & L. E. Bourne (Eds.), Foreign Language Learning: Psycholinguistic Studies on Training and Retention (pp. 339-364). Mahwah, NJ: Lawrence Erlbaum Associates.

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of (r) and (l) by native speakers of Japanese and English. Perception & Psychophysics, 18(5), 331–340. https://doi.org/10.3758/BF03211209

Monrad-Krohn, G.H. (1947). THE PROSODIC QUALITY OF SPEECH AND ITS DISORDERS: (A BRIEF SURVEY FROM A NEUROLOGIST'S POINT OF VIEW). Acta Psychiatrica Scandinavica, 22. https://doi.org/10.1111/j.1600-0447.1947.tb08246.x

Montgomery, C. R., Morris, R. D., Sevcik, R. A., & Clarkson, M. G. (2005). Auditory backward masking deficits in children with reading disabilities. Brain and Language, 95(3), 450–456. https://doi.org/10.1016/j.bandl.2005.03.006

Mora, J. C. (2022). Aptitude and Individual Differences. In The Routledge Handbook of Second Language Acquisition and Speaking (1st ed., pp. 68–82). Routledge. https://doi.org/10.4324/9781003022497-7

Morrison, G. (2002, April 6–7). Perception of English /i/ and /I/ by Japanese and Spanish listeners: Longitudinal results [Paper presentation]. In G. S.Morrison & L. Zsoldes (Eds.), Proceedings of the North West Linguistics Conference 2002, Simon Fraser University Linguistics Graduate Student Association, Burnaby, BC, Canada (pp. 29–48).

Morrongiello, B. A., Robson, R. C., Best, C. T., & Clifton, R. K. (1984). Trading relations in the perception of speech by 5-year-old children. Journal of Experimental Child Psychology, 37(2), 231–250. https://doi.org/10.1016/0022-0965(84)90002-X

Morton, J., & Jassem, W. (1965). Acoustic Correlates of Stress. Language and Speech, 8(3), 159–181. https://doi.org/10.1177/002383096500800303

Moyer, A. (1999). ULTIMATE ATTAINMENT IN L2 PHONOLOGY: The Critical Factors of Age, Motivation, and Instruction. *Studies in Second Language Acquisition*, *21*(1), 81–108. https://doi.org/10.1017/S0272263199001035

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

Munro, M. J., & Derwing, T. M. (1995). Processing Time, Accent, and Comprehensibility in the Perception of Native and Foreign-Accented Speech. Language and Speech, 38(3), 289–306. https://doi.org/10.1177/002383099503800305

Munro, M. J., Flege, J. E., & Mackay, I. R. A. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics*, *17*(3), 313–334. https://doi.org/10.1017/S0142716400007967

Muñoz, C. (2006). The Effects of Age on Foreign Language Learning: The BAF Project. In Age and the Rate of Foreign Language Learning (Vol. 19, pp. 1–40). Multilingual Matters.

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

Nation, I. S. P. (2006). How Large a Vocabulary Is Needed for Reading and Listening? Canadian Modern Language Review, 63(1), 59–81. https://doi.org/10.1353/cml.2006.0049

Nguyen, ATT, Ingram, J. (2005). Vietnamese Acquisition of English Word Stress. TESOL Quarterly, 39(2), 309. https://doi.org/10.2307/3588314

Nishi, K., & Kewley-Port, D. (2007). Training Japanese Listeners to Perceive American English Vowels: Influence of Training Sets. Journal of Speech, Language, and Hearing Research, 50(6), 1496–1509. https://doi.org/10.1044/1092-4388(2007/103)

Nittrouer S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. The Journal of the Acoustical Society of America, 115(4), 1777–1790. https://doi.org/10.1121/1.1651192

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). PHONOLOGICAL MEMORY PREDICTS SECOND LANGUAGE ORAL FLUENCY GAINS IN ADULTS. *Studies in Second Language Acquisition*, *29*(4), 557–581. https://doi.org/10.1017/S027226310707043X

Okobi, A. (2006). Acoustic correlates of word stress in American English. Massachusetts Institute of Technology; Cambridge: 2006. (Unpublished doctoral dissertation).

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

Omote, A., Jasmin, K., & Tierney, A. (2017). Successful non-native speech perception is linked to frequency following response phase consistency. *Cortex*, *93*, 146–154. https://doi.org/10.1016/j.cortex.2017.05.005

Ortega, M., Mora-Plaza, I., & Mora, J. C. (2021). Differential effects of lexical and non-lexical high-variability phonetic training on the production of L2 vowels. In J. Kirkova-Naskova, A., Henderson, A., & Fouz-González (Ed.), English Pronunciation Instruction: Research- Based Insights. John Benjamins.

Ou, S. C. (2010). Taiwanese EFL learners' perception of English word stress. Concentric: Studies in Linguistics, 36(1), 1-23. http://www.concentric-linguistics.url.tw/upload/articlesfs241402100955133136.pdf

Ou, S. C. (2016). PERCEPTION OF ENGLISH LEXICAL STRESS WITH A MARKED PITCH ACCENT BY NATIVE SPEAKERS OF MANDARIN. Taiwan Journal of Linguistics, 14(2), 1-31. https://doi.org/10.6519/TJL.2016.14(2).1

Parlak, Ö. (2024). The effects of implicit corrective feedback on production of lexical stress in L2 English. *Studies in Second Language Learning and Teaching*. https://doi.org/10.14746/ssllt.38361

Parlak, Ö., & Ziegler, N. (2017). THE IMPACT OF RECASTS ON THE DEVELOPMENT OF PRIMARY STRESS IN A SYNCHRONOUS COMPUTER-MEDIATED ENVIRONMENT. *Studies in Second Language Acquisition*, *39*(2), 257–285. https://doi.org/10.1017/S0272263116000310

PENNINGTON, M. C., & RICHARDS, J. C. (1986). Pronunciation revisited. TESOL Quarterly, 20(2), 207–225. https://doi.org/10.2307/3586541

Peperkamp, S. (2004). Lexical Exceptions in Stress Systems: Arguments from Early Language Acquisition and Adult Speech Perception. *Language (Baltimore)*, *80*(1), 98–126. https://doi.org/10.1353/lan.2004.0035

Peperkamp, S., & Dupoux, E. (2008). A typological study of stress “deafness.” In *Laboratory Phonology 7* (pp. 203–240). https://doi.org/10.1515/9783110197105

Peretz, I., Nguyen, S., & Cummings, S. (2011). Tone language fluency impairs pitch discrimination. *Frontiers in Psychology*, *2*, 145–145. https://doi.org/10.3389/fpsyg.2011.00145

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

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. Journal of the Acoustical Society of America, 32(6), 693–703. https://doi.org/10.1121/1.1908183

Petrova, K., Jasmin, K., Saito, K., & Tierney, A. T. (2023). Extensive residence in a second language environment modifies perceptual strategies for suprasegmental categorization.Journal of Experimental Psychology: Learning, Memory, and Cognition, 49(12), 1943–1955. https://doi.org/10.1037/xlm0001246

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. Attention, Perception, and Psychophysics, 71(6), 1385–1398. https://doi.org/10.3758/APP.71.6.1385

Pike, K. L. (1948). Tone Languages. Ann Arbor : University of Michigan Press.

Pittam, J., & Ingram, J. (1992). Accuracy of perception and production of compound and phrasal stress by Vietnamese-Australians. Applied Psycholinguistics, 13(1), 1–12. https://doi.org/10.1017/S0142716400005397

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. Journal of Experimental Psychology: Human Perception and Performance, 20(2), 421–435. https://doi.org/10.1037/0096-1523.20.2.421

Povel, D.-J., & Essens, P. (1985). Perception of Temporal Patterns. *Music Perception: An Interdisciplinary Journal*, *2*(4), 411–440. https://doi.org/10.2307/40285311

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. The Journal of the Acoustical Society of America, 119(3), 1684–1696. https://doi.org/10.1121/1.2161427

Qin, Z., Zhang, C., & Wang, W. S. (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

Rato, Anabela. (2014). Effects of Perceptual Training on the Identification of English Vowels by Native Speakers of European Portuguese. Concordia Working Papers in Applied Linguistics, 5, 529-546. http://doe.concordia.ca/copal/documents/34_Rato_Vol5.pdf

Rato, A., & Rauber, A.S. (2015). The effects of perceptual training on the production of English vowel contrasts by Portuguese learners. International Congress of Phonetic Sciences. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0656.pdf

R Core Team. (2023) *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.* Retrieved from https://www.R-project.org/

Revesz, A. (2012). Working Memory and the Observed Effectiveness of Recasts on Different L2 Outcome Measures. Language Learning, 62(1), 93–132. https://doi.org/10.1111/j.1467-9922.2011.00690.x

Rietveld, A. C. M., & Koopmans-van Beinum, F. J. (1987). Vowel reduction and stress. Speech Communication, 6(3), 217–229. https://doi.org/10.1016/0167-6393(87)90027-6

Rosen, S., & Manganari, E. (2001). Is There a Relationship Between Speech and Nonspeech Auditory Processing in Children With Dyslexia? Journal of Speech, Language, and Hearing Research, 44(4), 720–736. https://doi.org/10.1044/1092-4388(2001/057)

Rosner, B. S., & Pickering, J. B. (1994). Vowel Perception and Production. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198521389.001.0001

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. Frontiers in Psychology, 5, 1318–1318. https://doi.org/10.3389/fpsyg.2014.01318

Safronova, E., & Mora, J. C. (2013). Attention control in L2 phonological acquisition. In A. Llanes Baró, L. Astrid Ciro, L. Gallego Balsà, & R. M Mateus Serra. (Eds.), Applied linguistics in the age of globaliza- tion (pp. 384–390). Lleida: Edicions de la Universitat de Lleida.

Saito, K. (2013). Age Effects on Late Bilingualism: The Production Development of /ɹ/ by High-Proficiency Japanese Learners of English. *Journal of Memory and Language*, *69*(4), 546–562. https://doi.org/10.1016/j.jml.2013.07.003

Saito, K. (2019). Individual differences in second language speech learning in classroom settings: Roles of awareness in the longitudinal development of Japanese learners' English /ɹ/ pronunciation. Second Language Research, 35(2), 149–172. https://doi.org/10.1177/0267658318768342

Saito, K. (2023). How does having a good ear promote successful second language speech acquisition in adulthood? Introducing Auditory Precision Hypothesis-L2. Language Teaching, 56(4), 522–538. https://doi.org/10.1017/S0261444822000453

Saito, K., & Brajot, F. X. (2013). Scrutinizing the role of length of residence and age of acquisition in the interlanguage pronunciation development of English by late Japanese bilinguals. Bilingualism (Cambridge, England), 16(4), 847–863. https://doi.org/10.1017/S1366728912000703

Saito, K., Hanzawa, K., Petrova, K., Kachlicka, M., Suzukida, Y., & Tierney, A. (2022c). Incidental and Multimodal High Variability Phonetic Training: Potential, Limits, and Future Directions. *Language Learning*, *72*(4), 1049-1091. https://doi.org/10.1111/lang.12503

Saito, K., Kachlicka, M., Sun, H., & Tierney, A. (2020b). Domain-general auditory processing as an anchor of post-pubertal second language pronunciation learning: Behavioural and neurophysiological investigations of perceptual acuity, age, experience, development, and attainment. *Journal of Memory and Language*, *115*, 104168-. https://doi.org/10.1016/j.jml.2020.104168

Saito, K., Kachlicka, M., Suzukida, Y., Mora-Plaza, I., Ruan, Y., & Tierney, A. (2024). Auditory Processing as Perceptual, Cognitive, and Motoric Abilities Underlying Successful Second Language Acquisition: Interaction Model. *Journal of Experimental Psychology. Human Perception and Performance*, *50*(1), 119–138. https://doi.org/10.1037/xhp0001166

Saito, K., Kachlicka, M., Suzukida, Y., Petrova, K., Lee, B. J., & Tierney, A. (2022b). Auditory precision hypothesis-L2: Dimension-specific relationships between auditory processing and second language segmental learning. *Cognition*, *229*, 105236–105236. 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., SUN, H., & TIERNEY, A. (2019). Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental and suprasegmental sensitivity and performance via behavioural and neurophysiological measures. *Bilingualism (Cambridge, England)*, *22*(5), 1123–1140. https://doi.org/10.1017/S1366728918000895

Saito, K., Sun, H., & Tierney, A. (2020a). Domain-general auditory processing determines success in second language pronunciation learning in adulthood: A longitudinal study. *Applied Psycholinguistics*, *41*(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

Sardegna, V. G. (2012). Learner differences in strategy use, self-efficacy beliefs, and pronunciation improvement. In. J. Levis & K. LeVelle (Eds.). Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference https://www.iastatedigitalpress.com/psllt/article/15176/galley/13748/view/

Sardegna, V. G., & Dickerson, W. B. (2023). Improving the Pronunciation of English Polysyllabic Words Through Orthographic Word-Stress Rules. In English Pronunciation Teaching (Vol. 160, pp. 81–97). Multilingual Matters. https://doi.org/10.21832/9781800410503-011

Sardegna, V. G., & Jarosz, A. (2023). Learning English word stress with technology. In R. I. Thomson, T. M. Derwing, J. M. Levis, & K. Hiebert (Eds.), Proceedings of the 13th Pronunciation in Second Language Learning and Teaching Conference, --------held June 2022 at Brock University, St. Catharines, ON. doi: https://doi.org/10.31274/psllt.16142

Schmidt, R. (2001). Attention. In Cognition and Second Language Instruction (pp. 3–32). Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.003

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. Language Teaching, 47(4), 484–503. https://doi.org/10.1017/S0261444812000018

Schonell, F. J., I. G. Meddleton & B. A. Shaw (1956). A study of the oral vocabulary of adults. Brisbane: University of Queensland Press.

Scuffil, M. (1982). Experiments in Comparative Intonation: A Case-Study of English and German. Berlin, New York: Max Niemeyer Verlag. https://doi.org/10.1515/9783111595290

Service, E. (1992). Phonology, Working Memory, and Foreign-language Learning. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *45*(1), 21–50. https://doi.org/10.1080/14640749208401314

SHEN, X. (1990). ABILITY OF LEARNING THE PROSODY OF AN INTONATIONAL LANGUAGE BY SPEAKERS OF A TONAL LANGUAGE: CHINESE SPEAKERS LEARNING FRENCH PROSODY. International Review of Applied Linguistics in Language Teaching, IRAL, 28(2), 119–134. https://doi.org/10.1515/iral.1990.28.2.119

Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. Journal of Phonetics, 66, 242–251. https://doi.org/10.1016/j.wocn.2017.11.002

Sluijter, A. M. C., & van Heuven, V. J. (1996). Acoustic correlates of linguistic stress and accent in Dutch and American English. International Conference on Spoken Language Processing, ICSLP, Proceedings, 2(November), 630–633. https://doi.org/10.1109/icslp.1996.607440

Sluijter, A. M. C., van Heuven, V. J., & Pacilly, J. J. A. (1997). Spectral balance as a cue in the perception of linguistic stress. The Journal of the Acoustical Society of America, 101(1), 503–513. https://doi.org/10.1121/1.417994

Smith, B. L. (1979). A phonetic analysis of consonantal devoicing in children's speech. Journal of Child Language, 6(1), 19–28. https://doi.org/10.1017/S0305000900007595

Sneppe, R. & Wei, V. (1984). F0 Behaviour in Mandarin and French: An Instrumental Comparison. In A. Cohen & M. Broecke (Ed.), Proceedings of the Tenth International Congress of Phonetic Sciences (pp. 299-303). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110884685-042

Stagray, J. R., & Downs, D. (1993). DIFFERENTIAL SENSITIVITY FOR FREQUENCY AMONG SPEAKERS OF A TONE AND A NONTONE LANGUAGE. Journal of Chinese Linguistics, 21(1), 143–163.

Stevens, K. N. (1998). *Acoustic phonetics* (1st ed., Vol. 30). MIT Press. https://doi.org/10.7551/mitpress/1072.001.0001

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. Perception & Psychophysics, 36(2), 131–145. https://doi.org/10.3758/BF03202673

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners The re-education of selective perception. In J. G. H. Edwards & M. L. Zampini (Eds.), Phonology and Second Language Acquisition (pp. 159–198). John Benjamins Publishing Company.

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *The Journal of the Acoustical Society of America*, *64*(6), 1582–1592. https://doi.org/10.1121/1.382142

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

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

Taylor, D. (1981). NON-NATIVE SPEAKERS AND THE RHYTHM OF ENGLISH. International Review of Applied Linguistics in Language Teaching, 19(1-4), 219-226. https://doi.org/10.1515/iral.1981.19.1-4.219

Thomson, R. I. (2012). Improving L2 Listeners' Perception of English Vowels: A Computer-Mediated Approach. Language Learning, 62(4), 1231–1258. https://doi.org/10.1111/j.1467-9922.2012.00724.x

Thomson, R. I. (2016). Does training to perceive L2 English vowels in one phonetic context transfer to other phonetic contexts?. Canadian Acoustics, 44(3). Retrieved from https://jcaa.caa-aca.ca/index.php/jcaa/article/view/2910

Thomson, R. I. (2018). High Variability [Pronunciation] Training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208–231. https://doi.org/10.1075/jslp.17038.tho

Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective. Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference, June, 88--97.

Tierney, A., Gomez, J. C., Fedele, O., & Kirkham, N. Z. (2021). Reading ability in children relates to rhythm perception across modalities. Journal of Experimental Child Psychology, 210, 105196–105196. https://doi.org/10.1016/j.jecp.2021.105196

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

Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., & Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Research*, *1640*(Pt B), 251–263. https://doi.org/10.1016/j.brainres.2015.10.035

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive science*, *34*(3), 434–464. https://doi.org/10.1111/j.1551-6709.2009.01077.x

Tremblay, A., Kim, H., Kim, S., & Cho, T. (2023). Perceptual Training Enhances the Use of Vowel Quality Cues to Lexical Stress: The Benefits of Intonational Variability. *Proceedings of the 20th International Congress of Phonetic Sciences.* 211-215. https://scholar.google.com/scholar?oi=bibs&cluster=13889397327784688675&btnI=1&hl=en

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

Tyler, M. D. (2019). PAM-L2 and phonological category acquisition in the foreign language classroom. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), A Sound Approach to Language Matters: In Honor of Ocke-Schwen Bohn (pp. 607-630). https://doi.org/10.7146/aul.322.218

van Heuven, V. J., & Menert, L. (1996). Why stress position bias?. The Journal of the Acoustical Society of America, 100(4 Pt 1), 2439–2451. https://doi.org/10.1121/1.417952

Vuvan, D. T., Nunes-Silva, M., & Peretz, I. (2015). Meta-analytic evidence for the non-modularity of pitch processing in congenital amusia. Cortex, 69, 186–200. https://doi.org/10.1016/j.cortex.2015.05.002

Wang, Q. (2008) Perception of English stress by Mandarin Chinese learners of English: An acoustic study [Doctoral dissertation, University of Victoria]. UVicSpace Repository http://hdl.handle.net/1828/1282

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. The Journal of the Acoustical Society of America, 106(6), 3649–3658. https://doi.org/10.1121/1.428217

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese Listeners to Perceive Thai Tones: A Preliminary Report. Language Learning, 54(4), 681–712. https://doi.org/10.1111/j.1467-9922.2004.00283.x

Wayland, R., Landfair, D., Li, B., & Guion, S. G. (2006). Native Thai speakers' acquisition of english word stress patterns. _Journal of Psycholinguistic Research, 35(3), 285–304. https://doi.org/10.1007/s10936-006-9016-9

Wells, J. C. (2008). *English Intonation: An Introduction*. Cambridge University Press.

Wennerstrom, A. K. (2001). The music of everyday speech : prosody and discourse analysis / Ann Wennerstrom. Oxford University Press.

Werker, J. F. (1989). Becoming a Native Listener. American Scientist, 77(1), 54–59. http://www.jstor.org/stable/27855552

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. Child Development, 52(1), 349–355. https://doi.org/10.2307/1129249

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior & Development, 7(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. Phonetica, 49(1), 25–47. https://doi.org/10.1159/000261901

White, C. M. (1981). Tonal perception errors and interference from English intonation. *Journal ofChinese Language Teachers* Association, 16(2), 27-56.

White-Schwoch, T., Woodruff Carr, K., Thompson, E. C., Anderson, S., Nicol, T., Bradlow, A. R., Zecker, S. G., & Kraus, N. (2015). Auditory processing in noise: A preschool biomarker for literacy. PLoS Biology, 13(7), 17-. https://doi.org/10.1371/journal.pbio.1002196

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org

Winter B. (2014). Spoken language achieves robustness and evolvability by exploiting degeneracy and neutrality. BioEssays : news and reviews in molecular, cellular and developmental biology, 36(10), 960–967. https://doi.org/10.1002/bies.201400028

WONG, P. C. M., & PERRACHIONE, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. Applied Psycholinguistics, 28(4), 565–585. https://doi.org/10.1017/S0142716407070312

Wright, C. M., & Conlon, E. G. (2009). Auditory and Visual Processing in Children With Dyslexia. Developmental Neuropsychology, 34(3), 330–355. https://doi.org/10.1080/87565640902801882

Wright, B. A., Lombardino, L. J., King, W. M., Puranik, C. S., Leonard, C. M., & Merzenich, M. M. (1997). Deficits in auditory temporal and spectral resolution in language-impaired children. Nature (London), 387(6629), 176–178. https://doi.org/10.1038/387176a0

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. (2010). Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-language Users of English. Journal of Cognitive Neuroscience, 22(6), 1319–1332. https://doi.org/10.1162/jocn.2009.21272

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, K., Peng, G., Li, Y., Minett, J. W., & Wang, W. S. Y. (2018). The effect of speech variability on tonal language speakers' second language lexical tone learning. Frontiers in Psychology, 9, 1982–1982. https://doi.org/10.3389/fpsyg.2018.01982

Zhang, R., & Yuan, Z. (2020). EXAMINING THE EFFECTS OF EXPLICIT PRONUNCIATION INSTRUCTION ON THE DEVELOPMENT OF L2 PRONUNCIATION. Studies in Second Language Acquisition, 42(4), 905–918. https://doi.org/10.1017/S0272263120000121

Zhang, X. (2012). A Comparison of Cue-Weighting in the Perception of Prosodic Phrase Boundaries in English and Chinese (doctoral dissertation). University of Michigan. Retrieved from https://deepblue.lib.umich.edu/handle/2027.42/96107

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

Zhang, X., & Lu, X. (2014). A longitudinal study of receptive vocabulary breadth knowledge growth and vocabulary fluency development. Applied Linguistics, 35(3), 283–304. https://doi.org/10.1093/applin/amt014

Zhang, Y., Nissen, S. L., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. The Journal of the Acoustical Society of America, 123(6), 4498–4513. https://doi.org/10.1121/1.2902165

Zheng, Y., & Samuel, A. G. (2018). The effects of ethnicity, musicianship, and tone language experience on pitch perception. *Quarterly Journal of Experimental Psychology*, *71*(12), 2627–2642. https://doi.org/10.1177/1747021818757435

APPENDIX A

Lexical stress test stimuli

Full list of the lexical stress minimal pairs used in the lexical stress perception task in Experiment 1 and Experiment 2 (section 2.2.4.1). The words were elicited in question-answer pairs and later extracted for testing. Forty-seven (47) minimal pairs were used for testing, and 1 minimal pair was played in the practice trial (*insert*).

Word	**Stress pattern**	**Part of speech**	**Question-answer pair**	**Word frequency families**
abstract	trochaic	noun	Q: Is it included in the paper? A: No, it's just in the abstract.	K3
abstract	iambic	verb	Q: Is the summary ready? A: No, I will abstract it now.	
address	trochaic	noun	Q: Will they send it by post? A: Yes, I gave them my address.	K1
address	iambic	verb	Q: Did you see the president? A: Yes, he even address**ed** us.	
compact	trochaic	noun	Q: Did you buy a new camera? A: Yes, but it's a small compact.	K4
compact	iambic	verb	Q: I have almost no space left. A: You should compact the data.	
compound	trochaic	noun	Q: Is it spacious enough? A: Yes, it's a large compound.	K3
compound	iambic	verb	Q: He has just lost his job.	

			A: This will compound his problems.	
compress	trochaic	noun	Q: Did you hurt your knee? A: Yes, I'll get a warm compress.	K4
compress	iambic	verb	Q: Should I get more storage? A: No, just compress the files.	
conduct	trochaic	noun	Q: Are they investigating the company? A: Yes, for unethical conduct.	K3
conduct	iambic	verb	Q: Is this the new teacher? A: Yes, he will conduct the training.	
confines	trochaic	noun	Q: Does he feel safe? A: Yes, in the confines of his home.	K3
confines	iambic	verb	Q: Is this the main point? A: Yes, this confines the talk nicely.	
conflict	trochaic	noun	Q: Have they signed the agreement? A: Yes, they settled the conflict.	K3
conflict	iambic	verb	Q: Why don't you work together anymore? A: Her ideas conflict with mine.	
console	trochaic	noun	Q: Can we try the new game? A: No, the console is broken.	K4
console	iambic	verb	Q: Did you see how upset she was?	

			A: Yes, I tried to console her.	
content	trochaic	noun	Q: Did you like the new website? A: The content can be improved.	K3
content	iambic	verb	Q: Will he get first place? A: He will content himself with third.	
contest	trochaic	noun	Q: Why do you need a printer? A: For a photo contest.	K3
contest	iambic	verb	Q: Have you paid the fine? A: No, I will contest it.	
contract	trochaic	noun	Q: Did she accept the job? A: Yes, she signed the contract.	K2
contract	iambic	verb	Q: Are you doing the repairs alone? A: No, I will contract a plumber.	
contrast	trochaic	noun	Q: Did you like the phone? A: Yes, I liked the color contrast.	K3
contrast	iambic	verb	Q: Should I choose the white frame? A: Yes, to contrast the dark wall.	
convert	trochaic	noun	Q: What is this programme for? A: It will convert the signal.	K3
convert	iambic	verb	Q: Is he Christian? A: Yes, he is a new convert.	
decrease	trochaic	noun	Q: House prices have really dropped.	K3

			A: Yes, it's a big decrease in price.	
decrease	iambic	verb	Q: Will these exercises help? A: Yes, they help decrease the pain.	
desert	trochaic	noun	Q: Are there any towns nearby? A: No, they live in the desert.	K2
desert	iambic	verb	Q: Are the soldiers still at their post? A: No, they had to desert it.	
detail	trochaic	noun	Q: It's a good report, right? A: Yes, it's full of detail.	K2
detail	iambic	verb	Q: I can't wait to hear more A: I will detail the plan later.	
discharge	trochaic	noun	Q: How was he injured? A: From an electric discharge.	K3
discharge	iambic	verb	Q: Is he still in hospital? A: They discharge him today.	
discount	trochaic	noun	Q: Did you pay full price? A: No, they gave me a discount.	K3
discount	iambic	verb	Q: Is this possible? A: Yes, I won't discount it.	
escort	trochaic	noun	Q: Will he be protected? A: Yes, he will have police escort.	K4
escort	iambic	verb	Q: Did they call security? A: Yes, they had to escort her out.	

export	trochaic	noun	Q: Do they sell abroad? A: Yes, coffee is a big export.	K3
export	iambic	verb	Q: What do I do next? A: You will have to export the file.	
fragment	trochaic	noun	Q: Did you cut yourself? A: Yes, with a fragment of glass.	K3
fragment	iambic	verb	Q: Is the material durable? A: No, it will will fragment.	
impact	trochaic	noun	Q: Was it a good speech? A: Yes, it had quite an impact.	K3
impact	iambic	verb	Q: Is it an important decision? A: Yes, it will impact his future.	
import	trochaic	noun	Q: What are these forms for? A: They're for my import license.	K3
import	iambic	verb	Q: Do they sell cars? A: Yes, they import them.	
increase	trochaic	noun	Q: Was business better last month? A: Yes, there was an increase in sales.	K2
increase	iambic	verb	Q: Did you read the latest news? A: Yes, they will increase taxes.	
insert	trochaic	noun	Q: What about the side effects? A: They're on the package insert.	K3
insert	iambic	verb	Q: How can I get a drink? A: You have to insert a coin.	

insult	trochaic	noun	Q: How did they react? A: They took it as an insult.	K4
insult	iambic	verb	Q: Was she in the wrong? A: She had no right to insult him.	
intern	trochaic	noun	Q: Is Bob still a student? A: No, he works as an intern.	K6
intern	iambic	verb	Q: Will he go to prison? A: Yes, they will intern him for years.	
invite	trochaic	noun	Q: Are you coming to the wedding? A: Yes, we received the invite.	K2
invite	iambic	verb	Q: We haven't seen them for so long. A: We can invite them to dinner.	
perfect	trochaic	noun	Q: This necklace is so beautiful? A: Yes, it's the perfect gift.	K1
perfect	iambic	verb	Q: He is still very young. A: Yes, he will perfect his art.	
perfume	trochaic	noun	Q: What did you get for your birthday? A: A bottle of perfume.	K5
perfume	iambic	verb	Q: What a wonderful garden! A: Yes, the flowers perfume the air.	
permit	trochaic	noun	Q: Can I park here? A: Yes, but you need a permit.	K3

permit	iambic	verb	Q: Can I copy the files? A: The law does not permit it.	
present	trochaic	noun	Q: What did you get at the store? A: I bought you a present.	K1
present	iambic	verb	Q: What about the new project? A: I will present it today.	
proceeds	trochaic	noun	Q: Is it a charity concert? A: Yes, all proceeds are donated.	K3
proceeds	iambic	verb	Q: Is this the right way? A: Yes, the road proceeds north.	
produce	trochaic	noun	Q: Do you like this market? A: Yes, it's all local produce.	K2
produce	iambic	verb	Q: Do they work in the theatre? A: Yes, they produce plays.	
progress	trochaic	noun	Q: How is Steven doing at school? A: He is making good progress.	K2
progress	iambic	verb	Q: Will this affect the project? A: Yes, it will progress slowly.	
protest	trochaic	noun	Q: Did he agree to it? A: No, he left in protest.	K2
protest	iambic	verb	Q: Have they cut the budget? A: Yes, but we plan to protest.	
rebel	trochaic	noun	Q: Isn't her style very unusual? A: Yes, she is quite the rebel.	K3
rebel	iambic	verb	Q: He doesn't like authority, right? A: No, he tends to rebel.	

record	trochaic	noun	Q: Is he a marathon runner? A: Yes, he holds a record.	K1
record	iambic	verb	Q: Should I take notes? A: No, we record all meetings.	
refund	trochaic	noun	Q: How can I return the product? A: It's in our refund policy.	K6
refund	iambic	verb	Q: The phone is damaged. A: They will refund you in full.	
refuse	trochaic	noun	Q: Should I use this bin? A: Yes, it's for garden refuse.	K2
refuse	iambic	verb	Q: Are you coming then? A: Yes, I couldn't refuse.	
reject	trochaic	noun	Q: Why is this tire so cheap? A: It's a factory reject.	K3
reject	iambic	verb	Q: What did he say? A: He rejected our offer.	
research	trochaic	noun	Q: Is he a scientist? A: Yes, he works in research.	K2
research	iambic	verb	Q: Do you know a lot about the market there? A: No, we will research more.	
rewrite	trochaic	noun	Q: Have they started rehearsals? A: No, the play needs a rewrite.	K1
rewrite	iambic	verb	Q: Did he submit the essay? A: No, he had to rewrite it.	
subject	trochaic	noun	Q: Did she tell you anything? A: No, she changed the subject.	K1

subject	iambic	verb	Q: So what happens now? A: They will subject him to a test.	
survey	trochaic	noun	Q: Where did you read that? A: In a recent survey.	K3
survey	iambic	verb	Q: Why is the insurance company coming? A: To survey the damage.	
suspect	trochaic	noun	Q: Have they arrested him? A: Yes, he is their new suspect.	K2
suspect	iambic	verb	Q: Do you think she knows? A: Yes, I suspect she does.	
upset	trochaic	noun	Q: Did your team lose? A: Yes, that was a big upset.	K2
upset	iambic	verb	Q: Why is she crying? A: The phone call upset her.	

APPENDIX B

LexTALE Language proficiency test: Word list

Full word list presented in the LexTALE test (section 2.2.6.1). The first 3 items in the list are dummies used for practice trials. The word status designates the type of word, 0 = nonword, 1 = word. The test items and instructions presented below were taken from www.lextale.com. For more on the test and its validity see Lemhöfer & Broersma (2012).

Number	Item	Word status
0	platery	0
0	denial	1
0	generic	1
1	mensible	0
2	scornful	1
3	stoutly	1
4	ablaze	1
5	kermshaw	0
6	moonlit	1
7	lofty	1
8	hurricane	1
9	flaw	1
10	alberation	0
11	unkempt	1

12	breeding	1
13	festivity	1
14	screech	1
15	savoury	1
16	plaudate	0
17	shin	1
18	fluid	1
19	spaunch	0
20	allied	1
21	slain	1
22	recipient	1
23	exprate	0
24	eloquence	1
25	cleanliness	1
26	dispatch	1
27	rebondicate	0
28	ingenious	1
29	bewitch	1
30	skave	0
31	plaintively	1
32	kilp	0
33	interfate	0
34	hasty	1
35	lengthy	1

36	fray	1
37	crumper	0
38	upkeep	1
39	majestic	1
40	magrity	0
41	nourishment	1
42	abergy	0
43	proom	0
44	turmoil	1
45	carbohydrate	1
46	scholar	1
47	turtle	1
48	fellick	0
49	destription	0
50	cylinder	1
51	censorship	1
52	celestial	1
53	rascal	1
54	purrage	0
55	pulsh	0
56	muddy	1
57	quirty	0
58	pudour	0
59	listless	1

60	wrought	1

Instructions

The following instructions were shown to participants:

Recognising *English* words

This test consists of about 60 trials, in each of which you will see a string of letters. Your task is to decide whether this is an existing English word or not. If you think it is an existing English word, you click on "yes", and if you think it is not an existing English word, you click on "no".

If you are sure that the word exists, even though you don't know its exact meaning, you may still respond "yes". But if you are not sure if it is an existing word, you should respond "no".

In this experiment, we use British English rather than American English spelling.

For example: "realise" instead of "realize"; "colour" instead of "color", and so on. Please don't let this confuse you. This experiment is not about detecting such subtle spelling differences anyway.

You have as much time as you like for each decision. This part of the experiment will take about 5 minutes.

If everything is clear, you can now start the task.

APPENDIX C

Relationship between domain-general auditory processing abilities and lexical stress perception at Time 1

Analysis with N = 96 dataset

Regression model: This appendix reports the output from a multiple regression analysis with participants' lexical stress scores at Time 1 as an outcome variable, and the following variables as predictors: auditory discrimination thresholds (for pitch, formant, and risetime), dimension selective attention scores (for pitch, formant, and amplitude risetime), auditory-motor integration scores (for rhythmic memory, and melodic memory), and demographic measures (age of acquisition, pronunciation training, and weekly English use).

Dataset: The data of 96 participants were included in this regression analysis. The following participants were excluded from the original N = 102 dataset: the raw data of participants identified as outliers in the auditory discrimination tests (defined as performance that was more than 2 standard deviations away from the sample mean) were examined manually and those who showed anomalous response behaviour (i.e., consistent evidence for misinterpreting the task instructions), were excluded (N = 1). Additionally, participants who failed to perform above chance level (< .55) on any of the three dimension selective attention tasks were also excluded (N = 5).

Results:

The model was significant, $F(11, 84) = 5.25$, $p < .001$, $R^2 = .33$ (Table 16). Pitch discrimination abilities emerged as a significant negative predictor ($t = -2.60$, $p = .01$, Figure 24).

Table 16

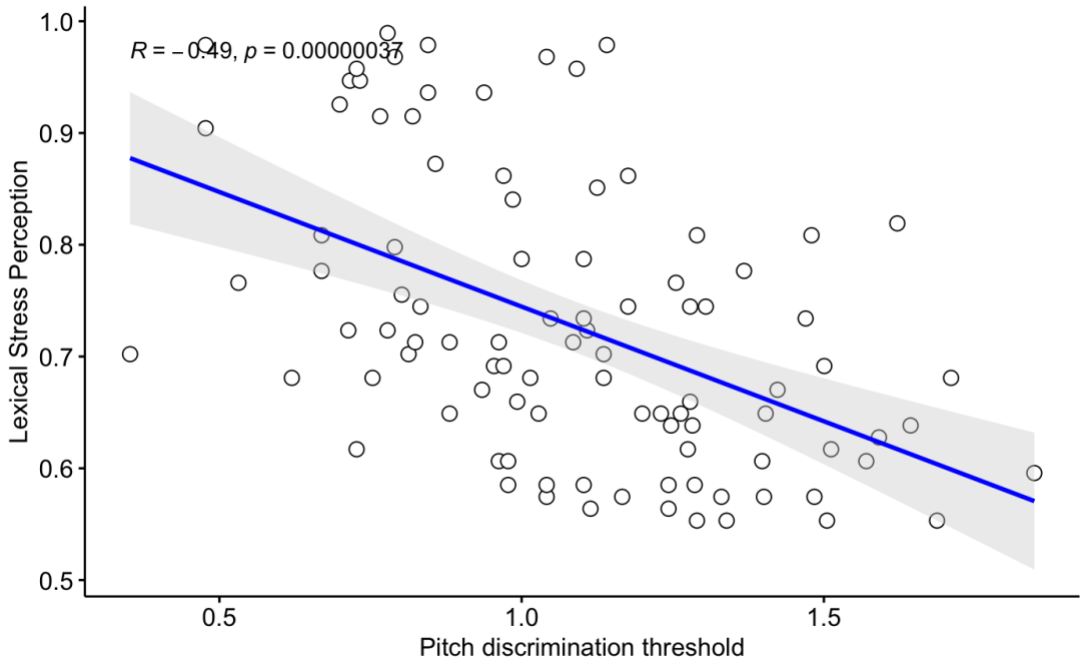
Multiple regression model output predicting lexical stress perception from a number of potential predictors (auditory discrimination (formant, pitch, risetime), dimension selective attention (formant, pitch, risetime), auditory-motor integration (rhythm memory, melody memory), and demographics (age of acquisition, pronunciation training, English use)).

Outcome variable	Predictor variables	Estimate	*SE*	95% CI		*t*	*p*
				LL	*UL*		
Lexical stress perception	Formant auditory discrimination thresholds	0.00	0.00	-0.00	0.00	0.53	0.60
	Pitch auditory discrimination thresholds	-0.12	0.04	-0.20	-0.03	-2.60	**0.01**
	Risetime auditory discrimination thresholds	-0.00	0.00	-0.00	0.00	-1.40	0.16
	Formant dimension selective attention	0.02	0.05	-0.09	0.13	0.43	0.67
	Pitch dimension selective attention	0.08	0.05	-0.03	0.19	1.46	0.15
	Risetime dimension selective attention	-0.03	0.11	-0.25	0.18	-0.31	0.76
	Rhythm memory	0.00	0.00	-0.00	0.00	1.56	0.12
	Melody memory	0.08	0.06	-0.05	0.20	1.19	0.24
	Age of acquisition	0.00	0.00	-0.00	0.01	1.01	0.31
	Pronunciation training	0.05	0.03	-0.01	0.10	1.72	0.09
	Weekly English use	0.00	0.03	-0.06	0.06	0.04	0.97

Note. SE - standard error, CI = confidence interval, *LL* - lower limit, *UL* - upper limit, *t* - t value. Statistically significant *p* values are bolded.

Figure 24

Negative linear relationship between pitch discrimination abilities and lexical stress perception scores. Participants with better discrimination acuity for pitch showed better English lexical stress perception.



APPENDIX D

Language and background questionnaire

In the following section you will be asked questions concerning your demographic background and language use.

1. What is your age in years?
2. What is your sex?
3. What is the highest level of education you have completed?
    a. No education
    b. Primary education
    c. Secondary education
    d. Bachelor's degree
    e. Master's degree
    f. Doctoral or professional degree
4. What is your occupation?
5. If you are studying at university, please indicate your current year of study.
6. What is your student status?
    a. Full-time
    b. Part-time
    c. Not applicable
7. Which university programme are you enrolled on?
8. Have you ever been diagnosed with any of the following:
    a. a vision problem
    b. a hearing impairment
    c. a language disability
    d. a learning disability
    e. none of the above
9. If you selected any of the above, please specify:
10. Please list all the languages you know in order of dominance from the most dominant to the least dominant.

a. Language 1
    i. At what age did you learn it?
    ii. Where did you learn it?
        1. Home
        2. School
        3. Community
        4. Other (please specify)
b. Language 2
    i. At what age did you learn it?
    ii. Where did you learn it?
        1. Home
        2. School
        3. Community
        4. Other (please specify)
c. Language 3
    i. At what age did you learn it?
    ii. Where did you learn it?
        1. Home
        2. School
        3. Community
        4. Other (please specify)
d. Language 4
    i. At what age did you learn it?
    ii. Where did you learn it?
        1. Home
        2. School
        3. Community
        4. Other (please specify)

11. Give the total number of years you received English language instruction.
12. Have you ever lived in place where English was the dominant communicative language?
    a. Yes

b. No

13. If you answered yes to the above question, please supply the following information:

    Which country did you stay in?

    How old were you? (in years)

    How long was your stay? (days, months, or years)

    What was the purpose of your stay? (language course, student exchange programme, summer language school, etc.)

14. Rate your English proficiency level on a scale of 0 - no proficiency to 10 - high proficiency for the following activities:
    a. Speaking
    b. Understanding
    c. Reading
    d. Writing
15. Have you taken any English language proficiency exams or standardised tests (e.g., IELTS, TOEFL, etc.)?
    a. No
    b. Yes, please specify which exam(s) you took and what scores you obtained
16. Have you taken the College English Test (band 4 or band 6) as part of the Chinese national exams?
    a. No
    b. Yes, please specify which exam(s) you took and what scores you obtained
17. Have you ever received training in English pronunciation?
    a. No
    b. Yes (please specify what type and length of training
18. Have you ever received training in English word stress (e.g., reCORD vs REcord)?
    a. No
    b. Yes
19. Are you confident in your perception of English word stress (e.g. reCORD vs REcord)?
    a. Not at all confident
    b. Very confident
20. Have you ever taken any linguistics or linguistics-related classes?
    a. No

b. Yes

21. Have you ever taught English as a foreign language?
    a. No
    b. Yes (please specify for how long)
22. If you are currently a university student, please indicate the following:
    a. How many hours a week do you have English language classes? (enter only numbers; if not applicable, enter "0")
23. How many hours a week do you have classes where you are taught the subject matter in English? (enter only numbers; if not applicable, enter "0")
24. Outside the classroom, please indicate how many hours per week you do the following:
    a. How many hours per week do you read in English? (enter only numbers)
    b. How many hours per week do you spend writing in English? (enter only numbers)
    c. How many hours per week do you spend listening to content in English? (enter only numbers)
    d. How many hours per week do you spend speaking in English? (enter only numbers)
    e. In your English language learning experience, have you received any other training outside of regular classroom settings?
25. Have you ever received any music training (playing an instrument (e.g. guitar/piano, singing, etc.)?
    a. No
    b. Yes (please specify how long in years)
26. Give the age at which you started your music training:
    a. Not applicable
    b. Specify the age at which you started your music training
27. In which musical instrument did you receive training?
    a. Not applicable
    b. Please specify type of musical instrument

APPENDIX E

Improvements in lexical stress perception from pre-post test

This appendix shows the output of a two-way analysis of variance of lexical stress perception scores run on the full dataset (N = 102). Here you will find summary statistics (Table 17), the ANOVA (Table 18), pair-wise comparisons (Table 19), and participants' lexical stress performance plots (Figure 25).

Note the ANOVA was run with lexical stress perception scores as an outcome variable, Group was the between-subjects factor, and Time - the within-subjects factor.

Summary statistics

Table 17

Group lexical stress performance (N = 102)

Group	Time	Variable	n	M	SD
Experimental	Time1	Portion correct	51	0.68	0.14
Experimental	Time2	Portion correct	51	0.78	0.16
Control	Time1	Portion correct	51	0.66	0.14
Control	Time2	Portion correct	51	0.66	0.15

ANOVA table

Table 18

ANOVA on pre-post test lexical stress perception scores

Effect	*F*(1, 100)	ηp2

Group	6.11*	0.06
Time	25.31***	0.20
Group:Time	24.69***	0.20

Note. Statistical significance levels: *p < 0.05. ** p < 0.01. *** p < 0.001

Pairwise comparisons

Table 19

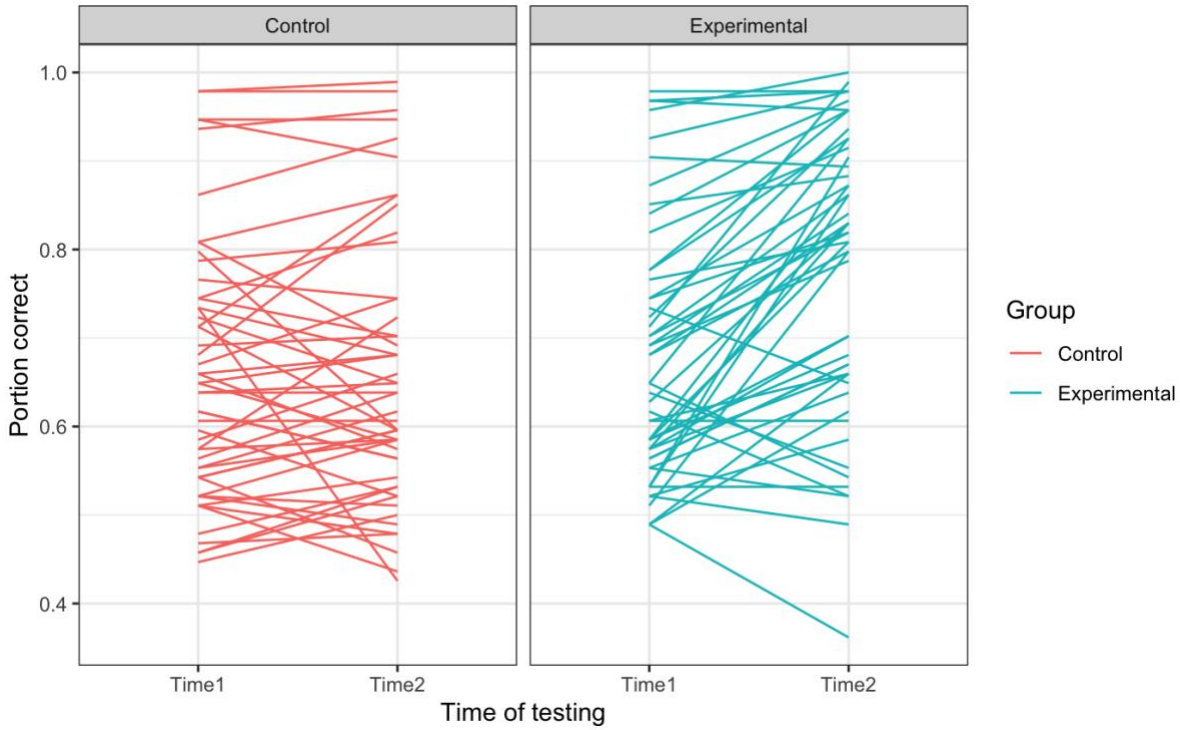
Pair-wise comparisons

For	Measure	Comparison	*p*
Time 1	Portion correct	Experimental group - Control group	ns
Time 2	Portion correct	Experimental group - Control group	***
Experimental group	Portion correct	Time 1 - Time 2	ns
Control group	Portion correct	Time 1 - Time 2	**

Note. Statistical significance levels: ns - not significant *p < 0.05. ** p < 0.01. *** p < 0.001

Figure 25

Lexical stress identification accuracy for the two training conditions (left - control group, right - experimental group), at Time 1 (pre-test), and Time 2 (post-test) for individual participants' data points (lines).


Control
Experimental
1.0
0.8
0.6
0.4
Portion correct
Time1
Time2
Time1
Time2
Time of testing
Group
Control
Experimental

APPENDIX F

HVPT training stimuli

List of training stimuli* that were used in the lexical stress identification and the category discrimination training task in the experimental training. To obtain the recordings, each word was embedded in the two carrier sentences presented below.

N	**Non-word**
1	abstroll
2	blotets
3	compold
4	condyle
5	conflave
6	demell
7	detiff
8	eltrave
9	flolest
10	fresment
11	imbert
12	insove
13	mihest
14	objorn
15	perfort
16	premress

17	projill
18	rebace
19	regand
20	reslind
21	subjall
22	sumrey
23	trassrer
24	udgate
25	ugbet

Carrier sentences:

I need to apgress. (*example)*

I need an apgress. *(example)*

**Note that only 25 minimal pairs from this list were used for training purposes. The final list of stimuli will be determined after all speaker recordings are obtained.*

APPENDIX G

Vocabulary training lists

The full list of the vocabulary items included in our online training can be found below. The words are grouped according to their word frequency level.

2000 Word Frequency Level	
Chinese	**English**
灰尘	dust
胜利	victory
选择	choice
旅途	journey
规模	scale
饿的	hungry
自豪	pride
勇敢的	brave
噪音	noise
苦的	bitter
责怪	blame
出生	birth
珍宝	treasure
学生	pupil
巨款	fortune
帝国	empire
倾斜	lean
（动物或人的）肉	flesh
倾倒	pour

游荡	wander
改进	improve
安排	arrange
受欢迎的	popular
袭击	attack
操作	operation
温度	temperature
税	tax
介绍	introduce
融化	melt
教育	education
当地的	local
慢的	slow
抱怨	complain
红酒	wine
薪水	salary
古老的	ancient
钢笔	pen
工厂	factory
发动机	motor
正义	justice
缓解	relief
制造	manufacture
财富	wealth
巨款	fortune
体育运动	sport
神圣的	holy
细小的	slight

爆裂	burst
秘密	secret
连接	connect
皇家的	royal
末端，尖端	tip
薪资	wage
询问	inquire
烘烤	bake
起初的	original
雇佣	hire
跳	jump
整个的	entire
欠	owe
固定	fix
更喜欢	prefer
好奇的	curious
独立的	independent
愉快的	merry
攀登	climb
匮乏	lack
诡计	trick
牺牲	sacrifice
检查	examine
复制品	copy
咖啡	coffee

3000 Word Frequency Level	
Chinese	**English**
受害者	victim
山脊	ridge
靴子	boot
蜡烛	candle
排列	dispose
编织	knit
居住	dwell
堡垒	fort
评论	comment
毯子	blanket
级别较低的	junior
掩盖的	concealed
博物馆	museum
鸡肉	muscle
放弃	quit
魔法的	magic
不顾一切的	desperate
使害怕	scare
每年的	annual
先前的	previous
地狱	hell
牧场	pasture
尖叫	scream
意识到的	aware
慈善	charity
情节	plot

引用	quote
注册	register
最高的	supreme
迫使	oblige
因素	factor
正常的	normal
公寓	apartment
中尉	lieutenant
（用图表等）说明	illustrate
谴责	condemn
凉意	chill
修剪	trim
野蛮的	savage
畜群	herd
漂移	drift
短语	phrase
竖起	erect
密封	seal
冠军	champion
传统	tradition
池塘	pond
忍受	endure
气候	climate
伙伴	mate
拥抱	embrace
疲惫的	weary
气氛	atmosphere
杰出的	brilliant

抛	toss
母亲的	maternal
确定的	definite
宣布	proclaim
省	province
木材	timber
进口商品	import
母鸡	hen
握紧	grasp
稳定的	stable
引人注目的	striking
大理石	marble
转换	shift
比赛	contest
霜冻	frost
附上	attach
明显的	distinct
空白的	blank

5000 Word Frequency Level	
Chinese	**English**
病房	ward
碎片	scrap
环状物	loop
总的	gross
沉思	contemplate
阶段	phase
老兵	veteran
独一无二的	unique
空着的	vacant
锻造车间	forge
（使）复原	revive
内部的	internal
胯部	hip
军团	legion
骑兵	cavalry
凳子	stool
放松	relax
无力的	limp
逗弄	tease
安抚	soothe
通信	correspond
悲惨的	tragic
酒精	alcohol
新奇的事物	novelty
桶	pail
灯泡	bulb

流血	bleed
肉排	steak
发起	launch
脉搏	pulse
称赞	compliment
对健康有益的	wholesome
香的	fragrant
坍塌，倒下	collapse
热情	zeal
苗条的	slim
浴缸	tub
火腿	ham
猪肉	pork
演示	demonstrate
穿透	penetrate
先于	precede
废除	abolish
出租	lease
混合	blend
时代	era
预测	predict
财政收入	revenue
仪器	apparatus
土堆	mound
氢	hydrogen
刺绣	embroider
市政的	municipal
充足的	adequate

愤恨	resent
志愿者	volunteer
纲领	creed
前夜	eve
按揭贷款	mortgage
随意的	casual
布道	sermon
公馆，宅邸	mansion
用力举起	heave
脆弱的	frail
独自的	solitary
气球	balloon
杂乱	mess
沙砾	gravel
丛林	jungle
分析	analysis
体面的	decent
成熟的	mature

10000 Word Frequency Level	
Chinese	**English**
牧师	vicar
触手	tentacle
美味的	luscious
拆毁	demolish
幽灵	apparition
闷烧	smolder
（失去平衡而）倒下	topple
飕飕作响	whiz
辅助的	auxiliary
坦率的	candid
亵渎，咒骂	blaspheme
晒太阳，取暖	bask
手脚架	scaffold
兵工厂	arsenal
养育	nurture
辅音	consonant
抢劫	loot
推动力	impetus
等级，层，排	tier
假释	parole
扣押权	lien
八度音节	octave
从事某项活动/工作的特定时间	stint
粗鲁的	impudent
行家	connoisseur
幸福	felicity

匆匆记下	jot
织锦	brocade
药膏	salve
废弃的	obsolete
热烈的	torrid
透明的	translucent
脱口而出	blurt
涉猎	dabble
使凹陷	dent
打趣	banter
口水	saliva
栓绳	leash
打滑	skid
蛆	maggot
巨大的	mammoth
前奏	prelude
水晶吊灯	chandelier
驱散	dissipate
潦草地写	scrawl
污染	contaminate
懒散的	indolent
非法的	illicit
人质	hostage
水坑	puddle
小桶	keg
马赛克	mosaic
夜间生活的	nocturnal
车辙	rut

先兆	foreboding
杂七杂八的	motley
自大的	pompous
雪花石膏	alabaster
（控制发动机油量的）节流阀	throttle
壁龛	alcove
泡沫	froth
植物学	botany
（教会的）执事	deacon
孢子	spore
舰队	convoy
滑稽可笑的举止	antics
伤亡	casualty
窥视	peek
抚慰	pacify
大摇大摆地走	swagger
狡猾的	wily
时间的	temporal

Academic Level	
Chinese	**English**
基础设施	infrastructure
中立的	neutral
趋势	trend
侵蚀	erosion
以实验为依据的	empirical
积累	accumulation
版本	edition
保证	guarantee
证据	evidence
方法	method
角色	role
辩论，争论	debate
性别	gender
强调，突出	highlight
层次	layer
想出（主意）	conceive
金融的	financial
即将发生的	forthcoming
首要的	primary
随机的	random
视觉的	visual
可供选择的事物	alternative
模棱两可的	ambiguous
与…相符合	correspond
心理学	psychology
计划表	schedule

全球的	global
总数	sum
转化	convert
预料	anticipate
编纂	compile
说服	convince
表示	denote
操纵	manipulate
出版	publish
对等的	equivalent
合同	contract
否定的	negative
哲学	philosophy
文件	file
连接	link
份额	proportion
技巧	technique
同意	consent
实施	enforcement
调查	investigation
规范	parameter
话题	topic
定义	definition
成年人	adult
百分比	percent
费用	fee
活下来	survive
援引	invoke

保持	retain
黏合	bond
引导	channel
估计	estimate
识别	identify
相邻的	adjacent
执照	license
媒体	media
车辆，交通工具	vehicle
使减少到最低限度	minimize
专题讨论组	panel
暴露	exposure
整合，一体化	integration
选择	option
计划	scheme
否认	deny
致力于	devote
释放	release

APPENDIX H

Post-test survey

1. Please answer the following questions about your experience taking part in this training study. On a scale of 1 to 10 please indicate how true these statements are:
    a. Do you think this online training has improved your English listening?
    b. Do you think this online training has improved your English pronunciation?
2. Were you paying attention during your online training sessions? Please answer honestly as this will help us to improve the quality of our data collection. Select from the below list the description most representative of your training experience.
    a. I was distracted most of the time
    b. I was distracted some of the time
    c. I was paying attention most of the time
    d. I was paying attention all the time
3. The below questions relate to the equipment you used while taking this experiment.
    a. Please indicate the type of audio device you used to complete the experiment:
        i. Earphones
        ii. Headphones
        iii. Other (please specify0
    b. If you were using earphones/headphones, were they:
        i. Wired
        ii. Wireless/Bluetooth-connected
    c. Were you using your earphone/headphones for the full duration of these experiments and your training sessions?
        i. Yes
        ii. No (please specify)

APPENDIX I

Changes in pre/post-test lexical stress cue weighting strategies - duration dimension

This appendix shows the output of a two-way analysis of variance of normalized duration cue weights run with N = 55 who had a significant relationship between at least one of the stimulus dimensions (pitch, or duration) and their lexical stress categorization responses.

Summary statistics

Table 20

Summary statistics of participants' lexical stress cue weightings at Time 1 and Time 2 for the experimental and control groups

Group	Time	Variable	n	M	SD
Experimental	Time1	NDCW*	30	0.20	0.19
Experimental	Time2	NDCW	30	0.24	0.27
Control	Time1	NDCW	25	0.17	0.18
Control	Time2	NDCW	25	0.18	0.19

*NDCW - Normalized duration cue weights

ANOVA table

Table 21

ANOVA table

Effect	*F*(1, 53)	*p*	ηp2

Group	0.78	0.38	0.01
Time	1.20	0.28	0.02
Group:Time	0.17	0.63	0.00

Figure 26

Participants' normalized duration cue weights at Times 1 and 2. There were no statistically significant changes in cue reliance between the Time 1 and Time 2 testing for either of the groups. Mean normalized duration cue weights are shown with error bars indicating 95% confidence intervals.