# An acoustic study of the production of English stress levels by English and Mandarin first-language speakers

by

**Magdalena Ivok**

Bachelor of Arts, University of Glasgow, 2021

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts

in the
Department of Linguistics
Faculty of Arts and Social Sciences

## Declaration of Committee

| | |
|---|---|
| **Name:** | **Magdalena Ivok** |
| **Degree:** | **Master of Arts** |
| **Title:** | **An acoustic study of the production of English stress levels by English and Mandarin first-language speakers** |
| **Committee:** | **Chair: Ashley Farris-Trimble**<br>Associate Professor, Linguistics |
| | **Henny Yeung**<br>Supervisor<br>Associate Professor, Linguistics |
| | **John Alderete**<br>Committee Member<br>Professor, Linguistics |
| | **Hua Lin**<br>Examiner<br>Associate Professor, Linguistics<br>University of Victoria |

## Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

## Abstract

Acquiring proficiency in the production of English lexical stress, crucial for effective communication in English as a second language (L2), involves mastering acoustic correlates such as fundamental frequency (F0), duration, intensity, and vowel quality. This study investigates how first language (L1) Mandarin speakers who are highly proficient in English, alongside L1 English speakers, produce distinct stress levels: While previous research focused on primary-stressed syllables (PS) and unstressed syllables with reduced vowels (UR), this study newly explores secondary-stressed syllables, or unstressed syllables without with vowel reduction (U). Acoustic analyses reveal that English and Mandarin L1 speakers largely employ similar cues, with duration crucial for distinguishing PS from U and UR syllables, and F0 playing a role in distinguishing UR from PS and U syllables. These findings challenge binary stress models and offer new insights into acoustic cues across English stress levels. Furthermore, the study sheds light on L1 Mandarin speakers' English stress production, demonstrating slight differences in their use of intensity and F0 peak compared to L1 English speakers, potentially due to L1 influences.

**Keywords**: lexical stress; second language acquisition; speech production; acoustics; English; Mandarin

# Table of Contents

# List of Tables

# Chapter 1. Literature Review

The second language acquisition (SLA) literature has extensively documented the influence of a learner's first language (L1) on the acquisition of their second language (L2). For example, in early stages of SLA, learners tend to rely on the phonological, morphological, and syntactic rules of their L1 to both perceive and produce sounds, words, and sentences in their L2 (Stefanich & Cabrelli, 2021; Mattys, 2000). This L1 effect has also been observed in L2 reading and writing (Derakhshan & Karimi, 2015; Karim, 2003). However, the precise nature and the extent of the influence of L1 on L2 acquisition are still being researched.

A better understanding of this relationship is essential for a more effective and informed approach to second language instruction. In today's globalised world, where English is a lingua franca, applying a uniform instructional approach might not be equally effective for learners from diverse L1 backgrounds. Therefore, to make L2 instruction more language-specific, it is crucial to identify those features that L1 speakers rely on the most for language learning. Here, I explore L2 acquisition of English phonological patterns with a particular focus on lexical stress: In Chapter 1, I review the broader SLA literature on these topics, and in Chapter 2, I report results from a novel study that explores the acoustic cues that Mandarin versus English L1 speakers use to produce stress in multisyllabic English words. Finally, in Chapter 3, I summarize the findings and discuss the implications of this work for theories of SLA and applied linguistics, as well as discuss the limitations of this work, with a view towards identifying directions for future research.

## 1.1. Background

Speech models explain that the production of speech involves a series of stages beginning with lexical preparation and lemma selection, followed by phonological and phonetic encoding and articulation (Levelt et al., 1999; Kröger et al. 2022). Prior research has established that an L2 learner's phonological processing and encoding is influenced by their L1, which entails the processing of both segmental (units of sound such as vowels and consonants) and suprasegmental information, which refers to phonological features that can extend over several segments, syllables, words, or phrases in running speech. Suprasegmental features provide information on syllable structure and language prosody, and are associated with acoustic information such as pitch, intensity, and (vowel) duration.

A substantial body of research has demonstrated the significance of segmental features for L2 speech perception (Flege et al., 1999; Guion, 2003). More recently, however, greater emphasis has been placed on suprasegmental features and their importance in L2 speech perception. In a study on the perception of Russian-accented speech by L1 American English listeners, for example, Banzina et al. (2016) discovered that L1 English listeners rely not only on segmental features, but also suprasegmental features such as the degree of lexical stress in their perception of accentedness. Kang (2004) collected speech samples from international teaching assistants in the US and presented them to 58 native American English speakers to make foreign accent judgements. She found that suprasegmentals such as pitch and lexical stress alone could predict how strong L1 English speakers judged non-L1 accent to be. Furthermore, suprasegmental features can, at times, have a greater effect on the perception of L2 speech than segmental ones (Derwing, 2008, as cited in Cheng, 2017; Munro & Derwing, 1999). To summarize, the failure to acquire not only segmental, but also suprasegmental features, might make L2 speech less comprehensible and can increase the perception of a foreign accent (Flege & Bohn, 1989; Zhang et al., 2008). This highlights the importance of suprasegmental features in L2 acquisition.

One of the most important suprasegmental features for English learners is word stress: in disyllabic and multisyllabic words, certain syllables are pronounced with greater prominence compared to others (e.g., the first syllable in *marker*, and the second syllable in *biology*), thereby considered as ‘strong’ syllables, whereas the unstressed syllables are considered as ‘weak.’ In the following sections, I will provide an overview of word stress production by L1 English speakers, highlighting acoustic characteristics that distinguish different syllable types. I will then discuss the production of English word stress by non-L1 speakers. These discussions will provide groundwork for the subsequent analysis of the current study, which will be discussed in the final section of this chapter.

## 1.2. Production of word stress in L1 English speakers

Numerous studies have highlighted the significance of stress in speech perception and production for L1 speakers, with a specific emphasis on its role in word segmentation. In perception, L1 English listeners rely on the distinction between strong and weak syllables to perceive boundaries in a continuous stream of speech, often identifying them as word onsets (Cutler & Norris, 1988; Cutler, 1995). Furthermore, research has demonstrated that even when segmenting artificial languages, listeners whose L1 languages have lexical stress tend to rely on stress cues for speech segmentation (Vroomen et al., 1998; Johnson & Jusczyk, 2001). In production, English speakers tend to pause before weak syllables to make articulation clearer, which indicates that they expect listeners to have more trouble hearing weak syllables compared to strong ones (Cutler & Butterfield, 1992). Overall, the correct stress production in English has been shown to enhance speech processing, and can affect speaker intelligibility (Field, 2005; Mattys & Samuel, 1997).

Two of the most studied aspects of stress in English are a) stress placement within a word based on syllable structure and b) acoustic realization of based on the interaction of different acoustic correlates, such as pitch, intensity, vowel duration, and vowel quality. While the present study focuses on the latter aspect of stress production, a brief overview of stress placement and syllable structure will contextualize stress categorization in English.

### *1.2.1. Stress placement and syllable structure*

Stress placement in English primarily depends on vowel quality and number of coda consonants within a syllable. Syllables can be classified as light (L) or heavy (H) based on their phonological structure, where heavy syllables contain a coda consonant or a diphthong. Weight-sensitive languages like English predict stress placement based on the word’s phonological structure (Hayes, 1995), resulting in primary stress (PS) typically falling on the penultimate syllable in multisyllabic non-verbs (e.g., ci'lantro, a'ssassin) when a heavy syllable is present. If a word does not contain a heavy syllable, PS falls on the antepenultimate syllable (e.g., 'resume, A'merica). However, PS can also fall on the final syllable of non-verbs if it is heavy and contains a full vowel (e.g. sar'dine).

For disyllabic verbs, stress falls on the final syllable with a long vowel or at least two consonants (e.g., co'llapse); otherwise, stress falls on the first syllable (Guion, 2005).

Such syllable-weight patterns predict that most disyllabic and trisyllabic words in English will have PS on the first syllable. This was confirmed by Garcia (2020) who, using a sample (*n* = 4573) from the CMU Dictionary that contained trisyllabic words with only one heavy syllable, had calculated that weight sensitivity correctly predicts stress position (penultimate or antepenultimate) in 98% of the trisyllabic words following HLL and LLL patterns (2020). Cutler and Carter (1987) similarly found that 70% of English content words contain PS on the first syllable, increasing to 90% when for frequently used words. Garcia (2020) estimated that 70% of a sample of 133,557 monosyllabic, disyllabic and trisyllabic words from the CMU Dictionary were initially stressed. If considering the initial SS syllable, almost two thirds of all words were initially stressed. Research on L1 English speakers' familiarity with these distributional stress patterns indicates that significant predictors of stress position include syllable structure, lexical class, and phonological similarity to real words (Guion et al., 2003).

Mattys (2000) suggests that speaker's familiarity with these stress patterns aids speech segmentation. Indeed, Field (2005) demonstrated that incorrect stress distribution impairs listeners' segmentation of words in an ongoing speech flow. Most content words in English contain primary stress on the first syllable, and syllables that receive stress are considered to be 'strong'. Consequently, in running speech, word onsets will be marked around strong syllables. Mattys (2000), however, highlights that speech segmentation based on distributional stress patterns may lead to false word boundary detections, especially for multisyllabic words, prompting the need for a finer segmentation process based on the degree of stress.

In summary, there is a notable bias for stress in English words to be positioned on the first syllable. However, disyllabic nouns are commonly initially stressed, contrasting with disyllabic verbs that typically have final stress (Guion, 2005; Kelly & Bock, 1988; Liberman & Prince, 1977). L1 English speakers are sensitive to these distributional patterns and use their intuitions for word segmentation in running speech. Recognizing the impact of this positional bias on stress production could be important for understanding differences in stress patters between L1 and L2 English speakers, aligning with the present study's exploration of the intricacies of English word stress.

### *1.2.2. Acoustic correlates for the production of English stress*

***Correlates for stressed and unstressed syllables.*** Stressed syllables display higher pitch, higher intensity and longer duration than unstressed syllables (Fry, 1955). Furthermore, stressed syllables always contain a full vowel, while unstressed syllables may have full or more centralized, reduced vowels. Some linguists argue that the term 'prominence' is more encompassing than 'stress' because it includes both acoustic attributes associated with stress and vowel quality (Banzina, 2012). However, this study will continue to use the term *stress* while also examining how vowel quality interacts with stress, given the historical and common use of this term.

There is a debate among researchers regarding which acoustic ccorrelate is the most reliable indicator of stress. Fry (1958, 1965) and Morton et al. (1965) investigated the perception of stress by manipulating various acoustic properties, such as duration, pitch, and intensity. Their results indicated that English speakers heavily rely on raised pitch to determine stressed syllables. However, the notion that pitch is the primary acoustic cue for stress has later been disputed because word stress can interact with phrase-level intonation, causing syllables within accented words to be naturally more prominent (Beckman, 1986; Chrabaszcz et al. 2014). This made some researchers argue that, in natural speech, pitch is less reliable than a combination of intensity, vowel duration, and vowel quality as a collection of acoustic cues to stress (Sluijter & van Hueven, 1996, as cited in Chrabaszcz et al. 2014).

Vowel duration has been found to be a reliable acoustic correlate for stress, with stressed syllables showing significantly longer vowel duration than unstressed syllables (Zhang et al., 2008). Similarly, vowel quality has been suggested to be a strong indicator of stress, as stressed syllables contain full vowels while unstressed syllables often have more centralized vowels reduced to schwa (Beckman & Edwards, 1994). However, both vowel duration and vowel quality may also be influenced by the syllable's position in a word or sentence. In contrast, intensity has been shown to be the least reliable and least important correlate to English stress. Although stressed syllables generally have higher intensity than unstressed syllables, this difference is not as effective an indicator of stress as pitch, vowel quality and vowel duration (Mattys, 2000).

***Correlates for different stress levels.*** Previous research on English stress levels has presented evidence for a clear distinction between primary stressed (PS) and unstressed reduced (UR) syllables as separate stress levels. However, the categorization of intermediate stress levels remains a subject of discussion. It is suggested that English contains two additional stress levels situated between PS and UR syllables, namely secondary-stressed syllables (SS) and unstressed unreduced syllables (UU). While empirical support exists for acoustic differences between PS and UR syllables and other syllables within a (multisyllabic) word (Plag et al., 2011; Okobi, 2006; Fear et al., 1995), the acoustic evidence for treating SS and UU as separate stress level categories is a matter of ongoing debate.

The existing literature on acoustic cues to word stress in English often compares 'stressed' syllables, which may include both PS and SS syllables unless specified, to 'unstressed' syllables, typically including only UU and UR syllables. Alternatively, this literature sometimes also compares unreduced PS, SS and UU syllables against reduced UR syllables, showcasing the inconsistency in grouping, which again reflects the lack of consensus on defining distinct stress levels. This binary categorization is also evident in the strong-weak syllable distinction, which some argue is based on the presence of a full vowel (as in PS, SS, and UU syllables) versus a reduced vowel (UR syllables) (Cutler & Butterfield, 1992). Other studies have shown that L1 English speakers perceive intermediate SS and UU syllables differently from PS and UR syllables, challenging the idea of binary distinction (Fear et al., 1995).

A limited number of studies have simultaneously examined the acoustic correlates of all stress levels. Studies by Mattys (2000) and Fear et al. (1995) found differences in vowel duration, but not in pitch and intensity, between PS and SS syllables. Comparing PS and UU syllables, Cutler et al. (2007) revealed that UU syllables had lower pitch, lower intensity, and shorter duration than PS syllables. Fear et al. (1995) found that UU syllables had lower intensity and shorter duration, but not lower pitch, than SS syllables, and a higher pitch, higher intensity and longer vowel duration than UR syllables. Consequently, this work implies three distinct levels of English stress: PS, UR, and the intermediate level that combines SS and UU syllables.

Overall, previous research suggests that pitch is the primary stress correlate in English, followed by vowel duration and vowel quality. Notably, these acoustic rankings are based on studies that have primarily compared stressed to unstressed syllables in disyllabic words (Fry, 1958; Okobi, 2006; Campbell & Beckman, 1997). Additionally, the disagreement among researchers on the weight of acoustic cues suggests that no single measurement can accurately predict lexical stress in natural speech. It is more likely that stress is realized through the interaction of different acoustic cues.

## 1.3. Acquisition of stress correlates in production by L2 English learners

Studies on L2 stress acquisition have reported that English L2 learners tend to transfer patterns from their L1 to L2, which means that they rely on their knowledge of L1 stress patterns to produce stress in English. In particular, an L2 learner's acoustic realization of an L2 stress contrast is influenced by acoustic cues associated with stress patterns in their L1 (Altman, 2006; Nguyen & Ingram, 2005; Cooper et al., 2002; see Van der Pas & Zonnevald (2004) for a detailed literature overview on L2 stress acquisition). Building on the acoustic cues identified in stress production of L1 English speakers, this section overviews how L2 learners of diverse linguistic background learn how to reproduce these cues in their L2 speech.

When producing English stress, L2 learners tend to employ acoustic cues that are familiar to them from their L1. For example, Zhang et al. (2008) investigated stress correlates used by L1 Mandarin learners of English when producing PS and UR syllables. They found that while both L1 Mandarin and English speakers used higher pitch, intensity and vowel duration to differentiate between PS and UR syllables, Mandarin learners had a significantly higher pitch than L1 English speakers. The study also found that Mandarin speakers were either not reducing vowels in UR syllables, or they were reducing them to a non-schwa sound. This is likely because Mandarin, unlike English, is a tonal language, where pitch is the primary acoustic correlate for tones, followed by duration. Furthermore, some vowels used in the study are not part of the Mandarin phonological inventory, and it is likely the reason why Mandarin L1 speakers had difficulty producing their reduced forms. These findings are supported by other studies that have shown that L1 Mandarin speakers rely significantly more on pitch than other acoustic cues when perceiving and producing English stress (Yu & Andruski, 2010; Rob et al., 2001; Corley, 2019). It is important to note, however, that perception and production should be treated separately, as research is still unclear about the extent to which one affects the other in L2 learning (Flege, 1995; Flege & Bohn, 2021; Nagle & Baese-Berk, 2021).

Vowel reduction was also investigated in a study by Flege and Bohn (1989). They presented L1 Spanish speakers with 4 English word pairs and found that even though Spanish speakers were able to make the contrast between stressed and unstressed syllables, they had difficulty reducing unstressed vowels. Their study further highlights the influence of the L1 on the acquisition of L2 stress patterns, as Spanish vowels do not undergo reduction regardless of their stress position. Conversely, Banzina (2012) explored the production of English stress by L1 Russian speakers and found that Russian speakers tend to over-reduce vowels in syllables that are not primary-stressed. This is consistent with earlier studies that found that Russian, unlike English, allows only one syllable containing a full vowel per word, which Russian learners of English may transfer to English, leading to incorrect reduction of full vowels in SS and UU syllable. Finally, in the study by Yakup (2018) that compared Kazakh-Russian bilinguals' production of English stress using 10 English minimal pairs, it was found that speakers utilized vowel duration and intensity as main stress cues, and not pitch. However, the study did not investigate vowel quality.

Ye-Jee (2018) conducted an acoustic analysis of English stress patterns produced by L1 speakers of Korean, Japanese and Taiwanese-Mandarin using 16 disyllabic words. Although all three groups a similar stress pattern to L1 English speakers, duration was identified as the main acoustic cues for stress. The results are consistent with the findings of Lee et al. (2006), who argued that Japanese speakers rely on duration more because vowel duration is phonemic in Japanese. Both studies also found that Korean speakers could differentiate duration in stressed and unstressed syllables, even though these differences are not as significant as those found in the production of L1 English speakers. However, Lee, Guion and Harada (2006) found that Korean speakers duration differences were smaller than pitch differences, arguing that Korean mainly uses pitch to realize its rhythm patterns.

The overarching patterns in the literature indicate that L2 learners predominantly rely on primary acoustic cues from their L1 to both perceive and produce L2 English stress, as well as their experience with L1 phonology and vowel reduction rules. However, little is known about the extent to which these L1 cues are used to distinguish between more than two stress levels. Additionally, it is pertinent to explore whether L1 acoustic cues can facilitate the acquisition of more than two distinct L2 stress levels. Rather than treating the differences in acoustic realization of L2 stress patterns between native and non-native speakers as L1 interference, a nuanced perspective might involve differentiating between the acquisition and acoustic realization of these categories as two separate domains, as was implied in Banzina's (2012) research.

## 1.4. Influence of L1 on L2 acquisition of multiple stress levels: An overview of the current study

The preceding sections have offered an overview of English word stress and the acoustic cues associated with English stress production. Additionally, the acquisition and use of these acoustic cues by English L2 learners is influenced by their familiarity with L1 acoustic cues. This suggests that, depending on their L1, certain cues may be more difficult to acquire, and that L2 learners rely on the L1 primary cues when acquiring English stress patterns. Several influential theoretical frameworks within the field of L2 acquisition aim to explain and predict the impact of L1 on L2 acquisition. These frameworks differ in their assumptions about the base for difficulties in L2 learning, but they all assume the strong influence of L1, whether perceptual or articulatory, and whether differences in the segmental or suprasegmental properties of the L1 and L2 matter more. The following section will overview some of the most influential models in this literature, including Best's Perceptual Assimilation Model (Best, 1995) and Flege's Speech Learning Model (Flege, 1995).

### *1.4.1. L2 Acquisition Models*

Perceptual Assimilation Model (PAM) (Best, 1995) was first developed to predict L2 sound perception based primarily on segmental properties, and from the perspective of a beginning L2 learner (i.e., someone hearing L2 speech sounds without knowing that language). However, the model could also be extended to suprasegmentals, and specifically predicts that learners will assimilate new sounds to existing L1 suprasegmental categories based on articulatory similarities. There are three ways in which new sounds can be interpreted: the L1 phonological influence can be facilitating, neutral or hindering (see also Banzina, 2012). Facilitation happens when two newly contrastive sounds assimilate to two different L1 phonological categories, or when one sound is clearly similar to L1 sound, and the other one is clearly different and cannot be assimilated to any of the L1 categories. Neutral influence happens when the new sound is placed somewhere in-between two L1 phonological categories. Finally, hindering effect happens when two L2 sounds assimilate to a single L1 phonological category.

Best and Tyler (2007) later extended this model to account for L2 speech perception by more experienced L2 learners and named it PAM-L2. PAM-L2 assumes that more experienced L2 learners can perceive phonetic differences between L1 and L2 sounds, but they will still assimilate them to the same phonological categories.

Unlike PAM, which focuses only on perception from the perspective of inexperienced learners, Flege's Speech Learning Model (SLM) was initially created to account for both L2 perception and production from the perspective of experienced L2 learners. The model postulates that difficulties that L2 learners encounter come from their experience with their L1. The more advanced the L1 system is at the start of L2 acquisition, the more influence it will have over L2 system. This means that the earlier the L2 acquisition starts, the less dissimilarity between phonological features is needed to perceive differences and form new phonological categories. Overall, L2 sounds that are similar to L1 sounds are harder to acquire than dissimilar sounds because L2 learners perceive them as equivalent.

Building on these theories that describe phonological and phonetic effects of the L1 on L2 acquisition, this thesis focuses on English stress by looking at production of English stress by speakers whose L1 languages differ in interesting ways from English. Specifically, I will investigate how English L2 learners who are L1 speakers of Mandarin, a language that has a completely different metrical and syllable structure compared to English, produce stress. Broadly, I will explore what acoustic cues and prosodic dimensions they use while producing stress. Two prominent models, PAM and SLM, offer predictions about the effect of L1 phonological inventory on perception and production of L2. PAM suggests that similarities between L1 and L2 suprasegmental features facilitate L2 stress production, while SLM predicts an increased likelihood of production errors in case of L1 and L2 similarities.

One other aspect of this present study is that it also explores multisyllabic words. Recent research, often guided by Liberman's (1975) argument that syllables are stressed by their association with non-stressed syllables within the same word, overwhelmingly focuses on within word syllable comparisons, particularly within disyllabic words. Consequently, there is very limited research on comparisons of stress levels across multisyllabic words, necessary for a more comprehensive understanding of more complicated English stress levels than just a binary stress/unstressed contrast. Additionally, apart from Banzina's (2012) study involving L1-Russian L2 learners, no study has directly compared acoustic cues to production of more than two English stress levels between L1 and L2 speakers. While previous studies have compared L1 English and L1 Mandarin speakers’ production of English stress and found Mandarin L1 acoustic correlates to be a significant predictor of English stress, this has not been done with all stress levels, which would provide another dimension to L2 acquisition models.

# Chapter 2. Acoustic study

## 2.1. Motivation for the present study

Although prior research has extensively investigated the production of English stress by English learners (Lee et al., 2006; Flege & Bohn, 1989; Zhang et al., 2008), most studies have focused on the binary stressed-unstressed syllable distinction. Major focus has been given to PS syllables and their contrast with an unstressed syllable, usually UR syllables. However, those intermediate syllables have received limited attention in previous research, particularly in terms of how they are produced by English learners. Relatedly, with a few notable exceptions (Fear et al., 1995; Banzina, 2012), there is a lack of research comparing more than two stress types in a single experiment. Another problem identified in previous literature is the choice of stimuli and how they were presented. Most studies presented participants with a list of disyllabic words, where the one syllable would be considered as 'stressed' and the other one as 'unstressed' syllable (e.g., Zhang et al., 2008; Li & Grigos, 2021; Guo & Chen, 2022). Because the stressed syllable is always a PS syllable, the specification of the unstressed syllable is often overlooked. Stimuli presentation can also affect the validity of the results in more complex contexts, because most studies present words to participants in isolated condition only, which does not accurately reflect natural speech.

Additionally, a significant number or studies have relied on L1 speakers' subjective perceptions of non-native speech stimuli without conducting an acoustic analysis to provide objective quantitative information on how non-native production differs from that of a L1 speaker. To the best of my knowledge, no study has yet explored the acoustic cues to production of primary stressed, unstressed reduced, and intermediate English stress levels by L1 Mandarin speakers in the same experiment. The aim of this study is to investigate acoustic cues to English stress levels by L1 speakers of Mandarin, a language that has a distinctly different stress system to that of English.

## 2.2. Acoustic correlates for English word stress by L1 Mandarin L2 speakers

Mandarin belongs to a group of tonal languages where syllables can either be ‘light’ and carry a neutral tone, or ‘heavy’ and carry one of the four lexical tones. Some researchers have argued that the distinction between the heavy and light syllables in Mandarin is similar to English stress in English, especially because heavy syllables have higher pitch, longer duration and higher intensity than light syllables, while light syllable also undergo reduction (Chao, 1968; Chrabaszcz et al., 2014; Yuan & Cheng, 2017). However, light syllables commonly occur in Mandarin words with two (or more) syllables, which are common in Mandarin. Many such words were historically formed by adding affixes to monosyllabic words, and so when a prefix is added, the second syllable is heavy and considered as ‘stressed’. Conversely, when a suffix is added, the second syllable is usually light (Tiee, 1977). Duanmu (2007) elaborated on that pattern and suggested that 70% of Mandarin Chinese disyllabic words follow the heavy-heavy syllable pattern, with the second syllable being acoustically more prominent. Drawing from Duanmu (2007), Yuan and Cheng (2017) further argued that speakers of Mandarin "have the knowledge of stress in their L1 grammar, […] but they do not “use” such knowledge as often as English speakers because [Mandarin] does not have many polysyllabic words". However, Garcia (2020) argued that "contrary to popular belied, most common words in the language are disyllabic, not monosyllabic", suggesting that Mandarin L2 learner of English might have more experience with disyllabic ‘stress’ than has been previously assumed.

Studies that have investigated acoustic correlates to Mandarin lexical tones have found pitch to be the primary acoustic correlate because tones primarily depend on the pitch contour, and there can be four types: steady, raising, falling-raising, and falling (Duanmu, 2000; Lia & Samuel, 2004). Results on intensity and duration vary – while different tones may exhibit distinct duration and intensity measures, they are generally considered to be less important than pitch. However, they can be effective cues when pitch information is absent (Altman, 2006; Liu & Samuel, 2004; Zhang et al., 2008).

Previous literature reports that L1 Mandarin speakers have difficulty manipulating acoustic cues to English stress and producing it in native-like manner. Even though it is possible that Mandarin speaker’s lack of experience with multisyllabic words might be the cause of their difficulty with English stress, research has shown that even when L1

Mandarin speakers are instructed on the correct stress placement, they still have difficulties using the correct acoustic cues and they produce stress differently from L1 English speakers. Specifically, Chen et al. (2001) found that Mandarin speakers tended to produce stressed words with a significantly higher pitch compared to L1 English speakers. This was further supported by Zhang et al. (2008) who presented L1 Mandarin speakers with seven pairs of disyllabic words that differed in stress placement only (e.g., 'desert and de'sert). The words were presented both in isolation and in sentence-medial position, and participants were informed about stress position. The results showed that L1 Mandarin speakers were able to use similar acoustic cues to that of L1 English speakers in both stressed and unstressed syllables (PS and UR in this case), but they still produced stressed syllables with a significantly higher pitch than L1 English speakers. These findings demonstrate Mandarin speaker's reliance on pitch to produce tonal categories in their L1 language translates to their overuse of pitch in their production of English stress.

## 2.3. Research objectives

The primary goal of the present study is to investigate acoustic cues that English learners use in their production of English word stress levels. Given the previous overview of English and Mandarin, I investigate any cross-language differences in acoustic cues that could be explained by language-specific prosodic patterns. Specifically, I sought to answer the following two questions:

1. Do English L1 and Mandarin L1 speakers make acoustic distinctions between different levels on English stress?
2. Do Mandarin L1 speakers employ similar acoustic cues as English L1 speakers to distinguish stress levels? If not, could these acoustic differences be attributed to their familiarity with L1 acoustic cues?

Additionally, since the evidence for acoustic distinction of SS and UU syllables is inconsistent, I will primarily treat the two stress types as one syllable type, unreduced (U), although I have also analyzed my results retaining the SS and UU distinction in a supplementary file that accompanies these results [https://osf.io/s6b5k/]. This analysis revealed a significant interaction effect of stress type and language only for intensity. While there were no significant intensity differences between SS and UU syllables within each language group, English speakers exhibited significantly higher intensity for SS syllables compared to Mandarin speakers ($p$ = <*.001)*.

### *Summary of predictions*

If Mandarin speakers transfer their L1 acoustic characteristics to English, they are predicted to primarily use pitch and vowel duration to produce different stress levels. While vowel quality and intensity are not expected to be as significantly different across language groups, vowel quality is expected to be more significant than intensity. Across the two groups, Mandarin speakers are predicted to use pitch significantly more than English speakers as acoustic correlate to stress levels. These predictions are based on previous research (Zhang et al., 2008; Chen et al., 2001; Liu & Samuel, 2004; Li & Grigos, 2021) that has reported that Mandarin L1 - English L2 learners differ in their realization of English stress in 6 different acoustic correlates from L1 English speakers: vowel duration, pitch, maximum pitch location, intensity and vowel quality (F1 and F2). The following section will discuss specific predictions about each of these acoustic correlates. Table 1 provides a general overview of stress patterns in English and Mandarin, and Table 2 provides a summary of predictions.

***Pitch.*** Based on previous research (Chen et al., 2001; Zhang et al. 2008), Mandarin speakers are predicted to use pitch to produce different stress levels similar to L1 English speakers. However, they are also expected to show significantly different pitch patterns from L1 English speakers. They are expected to have a higher pitch than L1 English speakers in all stress positions, and they should primarily use pitch to produce stress level differences.

***Vowel duration.*** Based on previously reported findings (Li & Grigos, 2021; Zhang et al., 2008), Mandarin speakers are predicted to have significant durational differences between PS and UR syllables. However, they are not expected to have significant durational differences between PS and U, and UR and U syllables.

***Vowel quality.*** Previous research has suggested that Mandarin speakers can produce reduced English vowels but have difficulty to do so when the target vowel (or the unreduced target vowel) does not exist in Mandarin vowel system (Zhang et al., 2008; Flege et al., 1997; Chen et al., 2001). Hence, their reduced vowels show unexpected F1 and F2 values. Here, I use a vowel that is present in the Mandarin vowel system, so Mandarin speakers in this study are predicted to produce reduced ([schwa])

and unreduced forms ([u]) of the vowel /u/ with F1 and F2 values similar to those of L1 English speakers. However, they are expected to reduce the vowel to a lesser degree than English L1 speakers, as the vowel /u/ does not usually exhibit reduction in Mandarin. There should also be significant vowel quality differences between unreduced and reduced syllables, however, mostly seen in values for PS and UR syllables.

***Intensity.*** Mandarin speakers are expected to show differences in intensity values between PS and other syllables – a pattern similar to L1 English speakers. Namely, intensity should be significantly higher in PS syllables compared to other stress levels for both groups. However, Mandarin speakers are expected to use intensity as stress correlate significantly less than L1 English speakers. Given the inconsistent research on the role of intensity in Mandarin speech, it is expected that L1-Mandarin speakers will show no significant differences in intensity between U and UR syllables.

***Maximum F0 location.*** This measure is included because a previous study reported that the maximum F0 value occurs earlier in stressed syllables than in unstresses syllables for Mandarin L1 speakers while L1-English speakers show no such difference (Zhang et al., 2008). Consequently, Mandarin speakers in the present study are expected to have their maximum F0 location earlier in PS syllables when compared to UR syllables, while this difference is not expected to occur with English L1.

**Table 1 Comparison of English and Mandarin stress patterns based on past studies.**

| | English | Mandarin |
|---|---|---|
| **Stress type** | PS, U, UR | Tonal language (full and light syllables) |
| **Primary stress cue** | F0, Duration | F0 |
| **Secondary stress cues** | Vowel quality | Duration, Vowel quality |
| **Least important stress cue** | Intensity | Intensity |
| **Multisyllabic stress** | PS, U and UR | Only disyllabic words |

Note: PS = Primary Stress, U = Unstressed, UR = Unstressed-reduced

**Table 2 Comparison of Predicted Acoustic Cues in Stress Type Distinctions Between English and Mandarin L1 Speakers**

| | PS-U | | U-UR | | PS-UR | |
|---|---|---|---|---|---|---|
| | Mandarin | English | Mandarin | English | Mandarin | English |
| **F0** | Higher for PS | Higher for PS | Higher for U | Higher for U | Higher for PS | Higher for PS |
| | M > E | | | | | |
| **Duration** | **No difference** | **PS longer** | **No difference** | **U longer** | PS longer | PS longer |
| | No group difference | | | | | |
| **Vowel quality** | No difference | No difference | Lower F1 and higher F2 than U | Lower F1 and higher F2 than U | Lower F1 and higher F2 than PS | Lower F1 and higher F2 than PS |
| | Less reduction for M | | | | | |
| **Intensity** | Higher for PS | Higher for PS | **No difference** | **Higher for U** | Higher for PS | Higher for PS |
| | M < E | | | | | |
| **Maximum F0 location** | No difference | No difference | No difference | No difference | **Lower for PS** | **No difference** |
| | No group difference | | | | | |

Note: PS = Primary Stress, U = Unstressed, UR = Unstressed-reduced. For each acoustic measurement, bold font in the first row indicates predictions for potential interactions involving stress-comparisons and language groups, which are denoted by M for Mandarin L1 and E for English L1. The second row for each acoustic measurement indicates the predictions for the main effect of language (i..e, group predictions).

In accordance with theoretical frameworks previously discussed in Chapter 1, it is possible to make predictions regarding the effect of L1 on L2 acquisition based on similarities and differences between the languages. Generally, differences between L1 and L2 may facilitate L2 production following the SLM framework. For example, Mandarin speakers may notice phonetic dissimilarities between their L1 and L2, which present challenges for their English production: In particular, L1 Mandarin speakers may struggle with distinguishing between PS and U syllables in English, as full tones are usually distinguished on the basis of F0 contour, and both of these syllables contain a full vowel, have higher pitch and longer duration than UR syllables. However, if significant differences are observed between PS and U syllables in the same word, they should associate them with different L1 categories and produce them differently. Furthermore, they may observe dissimilarities in vowel reduction between U and UR syllables because that feature is highly dissimilar to Mandarin phonology, which shows no vowel reduction for the target vowel in the present study, the vowel /u/.

Alternatively, in line with PAM, it is possible that similarities between L1 and L2 (instead of differences) may facilitate L2 acquisition. For instance, Mandarin speakers may associate English stress with Mandarin tones, relying primarily on pitch to produce

English stress and overlooking other important acoustic cues such as vowel quality and intensity. Specifically, if English stress levels are perceived different enough in terms of familiar acoustic cues from their L1, Mandarin L1 speakers are expected to rely on those L1 acoustic cues to differentiate stress levels. In that case, pitch is expected to be the primary predictor of stress levels for L1 Mandarin speakers, followed by duration.

In summary, this study investigates the production of English stress levels by Mandarin L1 speakers, particularly looking at acoustic cues associated with English stress. By exploring the differences in the use of acoustic cues across English and Mandarin L1 speakers, the research aims to uncover potential acoustic patterns influenced by one's L1. Based on prior research, it is anticipated that Mandarin L1 speakers will primarily use F0 and duration to differentiate stress levels, with vowel quality and intensity expected to exhibit lesser significance.

## 2.4. Methods

### *2.4.1. Subjects*

This study involved two groups of participants: 20 L1 speakers of Canadian English (13 female, 4 male and 3 non-binary) and 20 L1 Mandarin speakers (13 female and 7 male), who were recruited after obtaining a written informed consent form. L1 English speakers ranged in age from 19 to 38 years (*M* = 24 years), and the age of L1 Mandarin speakers ranged from 19 to 26 years (*M* = 23 years). Prior to participating in the experiment, all participants were asked to complete a short questionnaire on their linguistic background (Appendix B). All L1 English speakers were monolingual speakers and have lived in Anglophone Canada since they were born. Consistent with previous research (Zhang et al., 2008; Banzina, 2012), all L1 Mandarin speakers have lived in Canada for a minimum of 1 year and a maximum of 5 years (*M* = 3.1). Except for two participants, none of the Mandarin speakers lived in an English-speaking environment prior to arriving to Canada, and they received formal English instruction in China. The other two participants had, prior to moving to Canada, lived in the US for 3 and 8 years. The average number of years spent learning English for Mandarin participants was 13.6 years. Most of the participants were students at Simon Fraser University at the time of testing. The rest were friends of participants who shared information about the study. All participants reported to have normal hearing, speech, and language abilities. Participants could choose between course credits or cash as compensation.

### *2.4.2. Stimuli*

All experimental stimuli in Task 1 were disyllabic and multisyllabic English words that contained the vowel /u/ in one of the 3 stress positions: primary stressed (PS), intermediate unreduced (U), and unstressed-reduced (UR) syllables. As this vowel is present in the phonetic inventory of both English and Mandarin in its full form, and reduced only in English, it was chosen as the target vowel to ensure that the speakers of both languages are familiar with the sound and have no difficulty producing it.

Target words were selected from the Longman Pronunciation Dictionary (Wells, 2000), and their frequency was measured using the SubtlexUS database, which contains 51 million words from English and American movies and TV series. Only words with frequency value higher than 3 were included since words with frequency value 1-3 are considered low frequency words, and 4-7 as high frequency words. Initially, only words with a frequency value 4-7 were considered, but due to a limited list (10 words in total), the cut off was adjusted to include words with a frequency value 3-7, adding 11 more words. The final stimuli list included 22 words: 6 with the vowel /u/ in PS position, 10 in intermediate, and 6 in UR position, included in Table 1. Of the 10 'intermediate' words in this study, the dictionary indicated that 3 of them contained the vowel /u/ in a SS syllable, and 7 in a UR syllable. They were subsequently coded as one single 'intermediate' level. Words were presented in isolation and in sentence-medial position. The latter condition was added to mimic how they would be produced in a natural speech flow. Since previous research suggests that sentential pitch can affect sounds found in sentence-initial and sentence-final positions (Pater, 1997), the sentences were constructed so the target syllable was always the 6$^{th}$ syllable in the sentence, and it was followed by at least 3 syllables.

**Table 3 List of words containing the target vowel /u/.**

| Primary Stressed (PS) | Intermediate (U) | Unstressed Reduced (UR) |
|---|---|---|
| 'music | uni'versity | par'ticular |
| oppor'tunity | uni'versal | popu'lation |
| 'funeral | super'vision | 'instrument |
| un'usual | u'nited | 'volume |
| 'nuclear | 'schedule | i'rregular |
| im'prove | 'perfume | 'monument |
| | 'prostitute | 'educate |
| | circu'lation | |
| | refu'gee | |

### *2.4.3. Procedure*

All recordings were made in a sound-attenuated room at Simon Fraser University. Participants were asked to comfortably sit on a chair in front of a computer screen and wear a headset. All recordings were made using a Logitech USB Headset H390 that includes a high-quality microphone. The mouth-to-microphone distance was 10 cm. Speech tokens were sampled at a rate of 44.1 kHz and each token was saved as an individual .wav sound file to the SFU's server using the open-source experiment builder OpenSesame (Mathot et al., 2012).

Participants were first given verbal instructions, and written instructions were provided on the screen once the experiment started. The experiment was presented on a computer screen using OpenSesame. The experiment was divided into 2 parts, where words in Part 1 were presented in the isolated condition, and words in Part 2 in the sentence-medial condition. Within each part, the order of presentation was randomized across participants.

Once participants were ready to start Part 1, a target word appeared at the center of the screen in black font. Following a similar method described in Banzina (2012), participants were instructed to first *silently* pronounce the word that appeared on the screen twice to themselves, and then pronounce it aloud once the word turned red. They were asked to speak at a typical speech rate and loudness level. When the word turned black again, they could proceed to the next trial by pressing 'space' on the keyboard that was placed in front of them. The timing between the word appearing on the screen and font colour changing to red was 8 seconds, and then it took 8 seconds more for the word font colour to change to black again. The experiment was programmed so keyboard input was not accepted until word changed to black again. To ensure that the full audio was recorded even on occasions when someone would start saying the word too early (before it changed to red), the recording was taken from the time each word appeared on the screen until 'space' was pressed. Part 2 followed the same procedure, except that words were presented in sentence-medial position.

Each condition began with two practice trials; practice trials in Part 1 contained multisyllabic words with the vowel /u/ in PS position, and in Part 2 those same words were presented in sentence-medial position (see Appendix C). Practice trials were not included in the analysis. After completing the two practice trials participants were given the option to do 2 more practice trials by pressing 'enter' or proceed to the experimental trial by pressing 'space'. The average time to complete the full experiment was 15 minutes.

## 2.5. Analysis

After excluding practice recordings, a total of 44 recordings per speaker were analyzed. The recordings were first examined using Praat (Boersma & Weeknik, 2004). In instances where speakers had two or three productions of the same word or sentences, only the final instance was selected. Furthermore, the recordings underwent manual cleaning to remove excessive silences and noises (such as coughing, shuffling, etc.) before and after an utterance. Only the target sound, with 50 - 300 ms of silence on either end of the target word or sentence, was retained. To specifically isolate the target vowel /u/, all recordings underwent processing with the Montreal Forced Aligner (McAuliffe et al., 2017), a tool for forced speech alignment. After running the aligner, the recordings were manually adjusted in Praat to ensure that the target vowel was isolated consistently across all participants. More than half of the L1 English participants pronounced 'perfume' with the stress on the second syllable. Since the vowel in that word was originally coded to carry the intermediate stress, and not primary stress, the word was not included in the analysis. There was a total of 42 tokens per speaker that were included in the analysis, or 1680 tokens overall.

Following the approach outlined in Banzina (2012), the onset and offset of the vowel /u/ were identified by changes in F2 and F3 energy in the spectrogram. When preceded or followed by a stop or a fricative, boundaries were determined by the onset or offset of waveform periodicity. For approximants, vowels may have lower F3 than /l/, and higher F3 than /r/, while intensity may also indicate the boundary. Consequently, if the vowel was preceded or followed by an approximant, changes in the F3 and/or intensity were observed, and the boundary was placed at the midpoint of the transition.

Acoustic measurements of fundamental frequency (F0), F1, F2, vowel duration, intensity, and the time of maximum F0 were collected and analysed. F0 values were taken at vowel midpoint for each token, where the autocorrelation method set the F0 range for male speakers to be 75 – 600 Hz, and for female speakers 100 – 600 Hz. To account for the inherent individual variation among speakers, F0 values were normalized by converting them to semitones. This involved calculating the semitones with respect to each speaker's lowest frequency. The time of maximum F0 was automatically extracted from the F0 contour and converted to a proportion of the syllable duration, following the approach outlined in Zhang et al. (2008). Similar to F0, F1 and F2 measurements for each vowel were taken at the vowel midpoint for each token. Intensity values were taken at midpoint for each vowel and were normalized. The average intensity for each participant was calculated by dividing the sum of all intensity values by the total number of tokens (21), and intensity values were subsequently divided by this average for each speaker, separately for each condition.

Vowel duration, measured between vowel boundaries, was also normalized across participants to accurately reflect individual speech rates. Average speech rate was calculated separately for each participant and condition. First, the total duration of all the tokens containing the target vowel in the sentence condition was calculated and divided by the number of tokens (21), resulting in the average speech rate. Normalized vowel durations were obtained by dividing the vowel duration for each token by the participant's average speech rate. The same procedure was done for the word-only condition. The resulting normalized values were then multiplied by 100 and were displayed without a specified unit. Results

In accordance with preregistration on the Open Science Framework (OSF), the original study design included a between-subjects factor 'Group Type' (Language) with three levels (English, Mandarin, Russian) and a within-subject factor 'Stress Type' with four levels (PS, SS, UU, UR). Additionally, the original study design included two additional tasks that focused on the stress placement accuracy. However, due to logistical constraints and challenges in recruiting Russian L1 speakers, only English and Mandarin L1 speakers were included, and the study focus shifted to analysis of acoustic correlates of stress in Task 1.

Initially, a sample size of 15 participants per language group was determined based on a power analysis conducted using the software program G*Power (Faul et al., 2007). The goal was to obtain 0.95 power to detect a medium effect size at the 0.05 alpha error probability. However, this number was increased to 20 participants per group to ensure robustness in our findings, allowing for greater statistical power and increased reliability of the results. Additionally, while previous studies investigating acoustic correlates of stress have utilized smaller sample sizes (e.g., 10 to 15 participants per language group), the larger sample size provides a more comprehensive dataset for examining acoustic correlates of word stress in English and Mandarin L1 speakers, enhancing the validity of our findings and contributing to a more robust understanding of L1 and L2 English stress production.

A minor adjustment to the preregistered statistical model involved Language having two levels instead of three, and Stress Type having three levels instead of four. The latter change was based on inconsistent results around the differences between SS and UU syllables in previous research. Given evidence for the existence of an 'intermediate' stress type, the main focus shifted to investigating acoustic cues used by Mandarin and English L1 speakers to produce three different stress levels, while potential differences between SS and UU stress types were explored in a post-hoc analysis. Despite these adjustments, the general analysis steps remained consistent with the preregistered plan.

The acoustic measurements employed in production of different stress levels by L1 English and L1 Mandarin speakers were analyzed using six linear mixed-effects models. The analysis was done in the RStudio statistical computing software (Posit Team, 2023) running R (R Core Team, 2023) using the lme4 package (Bates & Maechler, 2010). Each model focused on a specific level of the dependent variable, incorporating fixed effects of stress level (PS, U, UR), language (English, Mandarin), the interaction of stress level and language, and presentation context (word-only, sentence). Additionally, the models included random effects of word and gender. Random intercepts were included to account for the differences in the baseline levels of the dependent variable across different words and participants. Random slopes were included to account for individual differences in the effect of language and stress type on acoustic measurements, by specifying the interaction between language and stress type among participants of different genders (see Table 4). In cases where the models failed to converge, they were reduced systematically. The process involved experimenting with different optimizers, decorrelating, and eliminating random slopes or intercepts that made minimal contribution to AIC/BIC levels. In situations where these adjustments did not improve the model, as was the case with the intensity measure, the decision was made to exclude random effects. Additionally, post-hoc analyses were conducted using the emmeans package (Lenth, 2023) to investigate any interactions and main effects through pairwise comparisons of estimated marginal means. Pairwise comparisons were conducted for the significant main effects or interaction effects identified in the linear mixed effects model, and p-values were adjusted for family-wise error rate using Tukey's method, where the significance level (alpha) was set at 0.05. The detailed analyses are available on the OSF repository [https://osf.io/s6b5k/].

Four participants, consisting of two L1 English and two L1 Mandarin speakers, were excluded from the analysis of fundamental frequency due to missing more than half of fundamental frequency values for U and UR stress types. Consequently, the total participant number in this particular analysis was 36. Similarly, the data from 1 L1 English speaker and 1 L1 Mandarin speaker were omitted from the analysis of maximum F0 time as they were missing more than half of the maximum F0 time values for U and UR syllables, bringing the number of participants for this specific analysis to 38.

**Table 4 Statistical model summary**

| **Model** | |
|---|---|
| DV ~Group Type * Stress Type + Context (1 + Group Type \| Word) + (1 + Stress Type\| Gender:Participant) | |
| **Variable** | **Description** |
| **DV** | Acoustic measurement of each of the 6 levels of the dependent variable (F0, Duration, Intensity, F1, F2, Peak F0 location) |
| **Group Type** | Independent variable with 2 levels (English, Mandarin) |
| **Stress Type** | Independent variable with 3 levels (PS, U, UR) |
| **Context** | Independent variable with 2 levels (word, sentence) |
| **Gender** | Random variable with 3 levels (Female, Male, Non-Binary) |

### *2.5.1. Acoustic results*

***Fundamental frequency (F0).*** The results showed no interaction effects of language and stress type [$F(2, 1219) = 1.07$, $p = .344$). There were significant main effects of stress type [$F(2, 17) = 10.24$, $p = .001$] and context [$F(1, 1236) = 47.01$, $p < .001$]. For both groups, syllables produced in word-only condition on average had higher F0 compared to syllables presented in a sentence. There were marginally significant main effects of language [$F(1, 34) = 3.70$, $p = .063$], with Mandarin participants showing a trend of having higher F0 than English participants.

To further understand the differences between different stress type levels, a post-hoc pairwise comparison was conducted using estimated marginal means. The pairwise comparison revealed that both L1 English and Mandarin speakers produced PS syllables with a significantly higher F0 than UR syllables ($p < .001$), while UR syllables exhibited a significantly lower F0 than U syllables ($p = .033$). In summary, PS had the highest F0, and UR the lowest F0, as shown in Figure A1. Interestingly, there was no difference between PS and U syllables ($p = .153$).

***Duration.*** The results for normalized duration revealed no interaction effects of language and stress type [$F(2, 21) = 0.96$, $p = .400$). There was a significant main effect of language [$F(1, 46) = 4.44$, $p = .041$], where L1 Mandarin speakers exhibited longer vowel duration for all stress types compared to L1 English speakers. There were also significant main effects of stress type [$F(2, 19) = 16.93$, $p = .001$] and context [$F(1, 1454) = 101.16$, $p = < .001$]. A *post-hoc* pairwise comparison of marginal means revealed that the vowel /u/ in the PS position was significantly longer than U ($p = .001$) and UR ($p < .001$) vowels, shown in Figure A2. However, there were no durational differences between U and UR vowels ($p = .252$).

***Intensity.*** The analysis of normalized intensity showed an interaction effect of language and stress type [$F(2, 1427) = 9.18$, $p < .001$], suggesting that patterns of intensity use across stress type differed in two language groups. While there were no significant intensity differences observed for pairwise comparisons within each language group, English speakers did produce UR syllables with significantly lower intensity than Mandarin speakers ($p = .007$), shown in Figure A3. This suggests that intensity was not a critical correlate for Mandarin speakers. Importantly, there were no significant main effects of language [$F(1, 1427) = 0.02$, $p = .896$], stress type [$F(2, 18) = 2.18$, $p = .551$]

and context [$F(1, 1433) = 0.01$, $p = .904$]. These results interestingly differ from the previous literature that reported higher intensity values for stressed syllables compared to unstressed syllables in English and Mandarin speakers, as well that L1 Mandarin speakers have, on average, lower intensity than L1 English speakers (Zhang et al., 2008).

***Formant frequencies.*** The results of F1 analyses showed no significant interaction effects of language and stress type [$F(2, 1473) = 0.77$, $p = .463$], and no main effects of language [$F(1, 38) = 0.64$, $p = .428$], stress type [$F(2, 18) = 1.62$, $p = .225$] or context [$F(1, 1488) = 1.39$, $p = .239$]. These results align with previous reports on Mandarin and English vowel space, where both groups had similar F1 values when producing English words (Zhang et al., 2008). However, my results show no significant differences in F1 for PS and UR vowels. Likewise, the analysis of F2 revealed no interaction effects of language and stress type [$F(2, 19) = 1.15$, $p = .339$], and there were no significant main effects of stress type [$F(2, 18) = 0.95$, $p = .405$]. However, there was a significant main effect of language [$F(1, 46) = 18.38$, $p < .001$] and context [$F(1, 1457) = 8.14$, $p = .004$]. Mandarin speakers, on average, produced the vowel /u/ with a lower F2 than English speakers, as shown in Figure A4. This aligns with previous findings where Mandarin speakers were reported to have a lower F2 for the vowel /u/ than L1 English speakers (Zhang et al., 2008).

***Time of maximum F0.*** The analysis of the time of maximum F0 revealed a significant interaction effect of language and stress type [$F(2, 21) = 7.96$, $p = .003$]. Furthermore, significant main effects were observed for stress type [$F(2, 19) = 5.30$, $p = .015$] and context [$F(1, 1205) = 16.99$, $p < .001$]. The main effect of language was not significant [$F(1, 33) = 3.01$, $p = .092$]. As shown in Figure A5, the interaction effect indicated that L1 Mandarin participants produced maximum F0 significantly later in PS syllables compared to UR syllables within their group ($p = .002$), but that was not the case for L1 English speakers, who had no significant maximum F0 differences between stress types ($p = .740$).These findings are consistent with prior research that reported that Mandarin speakers produce maximum F0 earlier in unstressed syllables compared to stressed syllables, while English speakers show no such differences.

# Chapter 3. Discussion and Conclusion

The primary goal of this study was to investigate the acoustic cues to English stress produced by L2 speakers, with a particular focus on examining whether L1 acoustic patterns could be observed in the realization of English stress patterns. As has been described in Chapter 1, the focus was centered around the effect of L1 on the acquisition of L2 suprasegmental properties, and this study aimed to expand on existing literature by comparing the acoustic realization of not two, but three different stress categories by introducing a less-studied, 'intermediate' stress level and comparing the acoustic realization of three stress categories. To do this, an acoustic study compared the production of three stress levels between L1 English speakers and L1 Mandarin speakers' production of English stress, analyzing the acoustic measurements of fundamental frequency, vowel duration, vowel quality, intensity, and maximum F0 location.

## 3.1. Summary of results

In the following sections, I will discuss the main findings concerning the acoustic measurements analyzed in this study. To provide a clear overview, Table 5 presents a summary that outlines the key results for the use of F0, duration, intensity, vowel quality, and maximum F0 location by both English and Mandarin L1 speakers in producing stress type distinctions. Subsequently, the following subsections provide a detailed discussion of the findings for each acoustic measurement.

**Table 5** **Comparison of Main Results for Acoustic Cues in Stress Type Distinctions Between English and Mandarin L1 Speakers**

| | PS-U | | U-UR | | PS-UR | |
|---|---|---|---|---|---|---|
| | Mandarin | English | Mandarin | English | Mandarin | English |
| F0 | Higher for PS | Higher for PS | Higher for U | Higher for U | Higher for PS | Higher for PS |
| | M > E (instead, both groups similar) | | | | | |
| Duration | No difference | PS longer | No difference | U longer | PS longer | PS longer |
| | No group difference (instead, M > E, likely to due to lower L2 proficiency) | | | | | |
| Vowel quality | No difference | No difference | Lower F1 and higher F2 than U | Lower F1 and higher F2 than U | Lower F1 and higher F2 than PS | Lower F1 and higher F2 than PS |
| | Less reduction for M (instead, E F2 > M F2, likely due to /u/-fronting) | | | | | |
| Intensity | Higher for PS | Higher for PS | **No difference** | **Higher for U** | Higher for PS | Higher for PS |
| | M < E (instead, both groups similar) | | | | | |
| Maximum F0 location | No difference | No difference | No difference | No difference | **Lower for PS** | **No difference** |
| | No group difference | | | | | |

Note: PS = Primary Stress, U = Unstressed, UR = Unstressed-reduced. For each acoustic measure, the first row indicates results for separate language groups (bold indicates an interaction). The second row for each acoustic measurement indicates the group predictions. Supported predictions are denoted by underlined cells, while unsupported predictions are marked with faded, non-underlined cells.

### *3.1.1. Results summary for pitch*

The results showed that Mandarin speakers were able to produce three distinct stress categories, and their production was overall comparable to L1 English speakers on measures of duration, mean fundamental frequency, and formant measures. These results partially align with previous research by Zhang et al. (2008) who found that L1 Mandarin speakers' use of fundamental frequency and vowel duration was comparable to L1 English speakers, though Mandarin speakers' production of stressed syllables with a higher F0 than English speakers was not replicated here. In their study on English sentence stress, Chen et al. (2001) also found that L1 Mandarin speakers produced both stressed and unstressed words with a higher F0 than L1 English speakers, suggesting that even though there may be L1 interference, it is not critically affecting the implementation of stress. These differences in results could be due to the experimental design; participants in the previously mentioned studies were told which syllables and words were stressed, which does not reflect the natural occurring speech. In contrast, Li and Grigos (2021) compared the production of stressed and unstressed syllables between L1 Mandarin and L1 English speakers using a nonword repetition tasks, and no significant differences in F0 were reported between the two groups. In a related stress perception study, Chrabaszcz et al. (2014) also found that Mandarin and English participants used F0 similarly to distinguish between stressed and unstressed syllables, revealing no significant group differences.

Although it has been suggested that higher F0 in L1 Mandarin speakers compared to L1 English speakers is because their L1 has a higher frequency range than L1 English speakers (Zhang et al., 2008), the lack of such results in the current study would suggest the alternative view: Mandarin L1 speakers' experience with four 'heavy' L1 syllable types based on tonal categories related to the F0 contour, in addition to the existence of the 'light' syllable carrying the neutral tone that contrasts the other 4 tones, aids Mandarin speakers' production of L2 stress patterns such that they can accurately approximate English L1 speakers' use of stress. Moreover, given that Mandarin speakers were able to distinguish between PS and UR, and UU and UR syllables using F0, it shows that F0 is used to make a two-way distinction between reduced and unreduced syllables, partially supporting the prediction that F0, along with duration, was used as primary stress cue. In short, L1 Mandarin speakers were familiar with F0 manipulation from their L1 and were able to produce comparable stress patterns to English speakers. This phenomenon suggests that similarities between L1 and L2 facilitate L2 acquisition and aligns with the prediction described under PAM (Best, 1995).

Additionally, previously reported results for cross-linguistic influences on the location of maximum F0 in English stress were replicated, as I also showed differences across language groups. Specifically, I also found that the maximum F0 in L1 Mandarin speakers occurring later in PS syllables than UR, compared to L1 English speakers. Zhang et al. (2008) suggested that later F0 peak might be due to L1 transfer and longer duration of stressed syllables in their Mandarin-speaking participants. However, Mandarin participants in the current study had longer duration across all syllable types, and still produced F0 peak later in the stressed syllable compared to their English counterparts. In their recent research on F0 contour in Mandarin speakers, Yeung et al. (2020) argued that the later peak is not due to the L1 transfer but is a result of general inability to navigate multiple L2 acoustic cues to produce native-like stress patterns. Therefore, the results in the current study suggest that, while English L2 learners generally used higher or lower F0 to produce differences between syllable types, they had not yet learned the finer-grained cues like the location of peak F0. While Mandarin

L1 speakers used F0 to differentiate stress contrasts similarly to English L1 speakers, they were unable to locate maximum F0 in a native-like manner, suggesting an interference from their L1.

Furthermore, Mandarin L1 speakers did not exhibit differences in the location of maximum F0 between U and other syllables. While maximum F0 location in Mandarin has been linked to syllable duration, with longer syllables exhibiting later maximum F0, (Xu, 1999), which Zhang et al. (2008) later used to explain the differences in max F0 between stressed and unstressed syllables produced by their Mandarin L1 participants, this explanation does not apply directly to the present findings. Despite PS syllables being notably longer than both U and UR syllables, significant differences in maximum F0 location were observed only between PS and UR syllables. If Mandarin L1 learners were transferring maximum F0 patterns from their L1 based on syllable duration, differences between PS and U syllables would also be expected. The absence of such differences could suggest and interaction between duration and pitch mandated by their L1, which may not readily U syllables as distinct from PS or UR syllables.

### *3.1.2. Results summary for duration*

In terms of duration, it was predicted that durational differences for L1 Mandarin speakers will be observed only between PS and UR syllables. This was partially supported by the results, as PS syllables were indeed longer than UU and UR syllables for both groups. L1 Mandarin speakers exhibited longer durations across all three stress levels compared to L1 English speakers, but this did not interact with stress type. While some studies have suggested that duration serves as the primary correlate for tone contrasts in Mandarin (Li & Grigos, 2021; Duanmu 2007), there is no evidence in the literature to suggest that Mandarin syllables are, on average, longer than English syllables. Hence, these durational group differences in stress are novel. Mandarin speakers' longer syllable duration is more likely due to their English proficiency levels. Several studies that measured English syllable duration and speech rate in L1 English speakers and Mandarin speakers of intermediate (IP) and high proficiency (HP) found that L1 Mandarin speakers, particularly of IP, had a significantly slower articulation rate than L1 English speakers, as well as HP Mandarin speakers (Anderson-Hsieh & Venkatagiri, 1994; Chen & Robb, 2004). Even though the proficiency level of participants in the present experiment was not tested, they were all international students in Canada who moved to the country within the past 5 years. This aligns with the demographic characteristics of previous studies. Therefore, it is possible that the longer duration is the result of non-L1 speech rate, as suggested in previous research (Guion et al., 2000).

### *3.1.3. Results summary for intensity and vowel quality*

The intensity results revealed differences between Mandarin and English L1 speakers in their production of UR syllables. While the role of intensity as a correlate to English stress remains debated, previous studies generally indicate that English speakers produce UR syllables with lower intensity compared to PS syllables (Lehiste, 1970; Fear et al., 1995), suggesting a consistent trend of lower intensity for UR syllables. In contrast, research on Mandarin tones has presented conflicting findings regarding intensity patterns. While some studies suggest that 'light' syllables (carrying the neutral tone) exhibit lower intensity than 'heavy' syllables, others argue against such patterns, suggesting that light syllables do not consistently have lower intensity than heavy syllables. This discrepancy places intensity as the least reliable correlate for Mandarin tones (Duanmu, 2007; Cao, 1986). I found that English L1 speakers adhere to a downward pattern in using intensity as a correlate for UR syllables, while Mandarin L1 speakers do not exhibit such a pattern and produce UR syllables with higher intensity compared to English L1 speakers. However, the lack of between and within-group differences in the use of intensity as correlate to stress levels aligns with some studies that reported intensity to be unreliable marker of stress in English, secondary to F0 and duration, and others where Mandarin speakers have been shown to rely on intensity more (second to duration) when F0 is removed from speech stimuli. This suggests that both groups primarily rely on F0 and duration in their perception, and possibly production of stress levels (Mattys, 2000, Chrabaszcz et al., 2014; Shen, 1993; Cutler & Darwin, 1981).

In contrast to the prediction that Mandarin L1 speakers would exhibit lower vowel reduction (as measured in formant values) for UR syllables compared to English L1 speakers, the analysis revealed no interaction effects between language and stress type for F1 and F2 measures. Additionally, there were no vowel quality differences between UR and other unreduced syllable types for both language groups, contrasting the prediction that UR syllables would exhibit a reduced vowel space compared to PS and U syllables. Notable, while there were no group differences for F1 measures, while

Mandarin speakers demonstrated lower F2 values compared to English speakers. These results are surprisingly confounding given that vowel quality has occasionally been argued as the most important stress correlate for English speakers (Braun et al., 2011; Cutler et al., 2007). The lack of vowel reduction for UR syllables may be due to the experimental design; Kuo and Weismer (2016) analysed vowel reduction across four types of experimental tasks and found that reduction (measured in F1 – F2 distance) was least evident in the tasks were target words were presented in a carrier phrase as opposed to passage reading or conversations. It is likely that participants emphasized the target words and showed less reduction.

However, Mandarin speakers were found to have lower F2 values than English speakers for all syllable types, aligning with previous research that reported lower F2 for Mandarin speakers when compared to English speakers. Zhang et al. (2008) attributed these results to the historical fronting of American English vowels, while Chen et al. (2001) suggested that L1 Mandarin speakers may produce vowels with greater posterior tongue constriction, or tongue backing, lowering the second formant frequency. The current study involved participants who were L1 speakers of Canadian English, and recent studies on the Vowel shift in Canadian English found that three back vowels, including /u/, are moving frontwards, leading to higher F2 values (Bobert, 2010; Bobert 2019). Therefore, as all English participants were monolingual and only ever lived in Canada, it is likely that group differences in F2 values are due to the fronting off the Canadian /u/ vowel, a feature that has not been picked up by L2 speakers.

## 3.2. Responses to Research Questions

### *3.2.1. Do English L1 and Mandarin L1 speakers make acoustic distinctions between different levels on English stress?*

The first goal of this study was to understand what acoustic measurements were to produce different stress levels and provide further evidence for the distinct acoustic differences between primary stressed and unstressed reduced syllables, as well as the intermediate syllables. In both language groups, PS vowels exhibited longer duration and higher fundamental frequency (F0), but not higher intensity, compared to UR vowels. This partially aligns with previous research (Fry, 1955; Lieberman, 1960; Okobi, 2006; Li & Grigos, 2023). The absence of an intensity difference suggests that intensity might not be as reliable an acoustic cue as F0 and duration, as reported in some perceptual studies (Mattys, 2000; Chrabaszcz et al., 2014). However, an alternative explanation could be the influence of experimental methods. Banzina (2012) highlighted the potential confounding effect of phrase-level prominence in previous studies, where words were often presented in isolation or in sentence-final positions, potentially confounding the effects of stress level with tonic or phrase-level stress.

Moreover, the intermediate U syllables differed from PS syllables in shorter duration, and from UR syllables in higher F0. This suggests the existence of an intermediate stress level with acoustic characteristics distinct from both PS and UR syllables. Notably, the intriguing aspect lies in the fact that both duration and F0 play roles in distinguishing between PS and UR, but not U syllables. While U syllables were shorter than PS syllables, they did not demonstrate a different F0. Conversely, U syllables had a higher F0 than UR syllables, yet they did not exhibit a different duration. While this nuanced acoustic variation may emphasize the presence of an intermediate stress level with acoustic characteristics lying between those of PS and UR syllables, it may be influenced by the coding of U syllables in this study. Specifically, U syllables contain both secondary-stressed (SS) and unstressed-unreduced (UU) syllables – two stress types whose distinct acoustic correlates have been a subject of debate in previous research. For example, Fear et al. (1995) found that SS and PS syllables only differed in duration, while UU syllables differed from PS, UR, as well as SS syllables in terms of duration, vowel quality and intensity. UU syllables also had lower F0 than PS syllables. These acoustic discrepancies between SS, UU and other syllables might explain the current differences in F0 and duration.

To address the debate around the acoustic distinctiveness of SS and UU syllables, a post-hoc analysis on secondary stressed (SS) and unstressed unreduced (UU) syllables was conducted. It is important to note that the primary goal was to compare SS and UU syllables to each other to analyse any acoustic differences, and not to the rest of the syllables. The results revealed no significant acoustic differences between the two stress levels across any of the acoustic measurements, contrary to findings in some previous studies that reported differences in intensity, F0, and duration (Banzina, 2012; Fear et al., 1995). The detailed results of this post-hoc analysis is available in the supplementary materials provided on the OSF for further reference [https://osf.io/s6b5k/].

These differences in results could be due to the stimuli presentation and the position of the target syllable. Plag et al. (2011) and Banzina (2012) suggested that a prominence of a syllable and its acoustic correlates are related to the syllable position within the word or sentence. Namely, SS and UU syllables that occur before the PS syllable may be acoustically more prominent than in a position after the PS syllable. In the current study, the criteria for selecting target vowels constrained word options, resulting in six out of eight words containing the intermediate stress level with the target vowel positioned before the PS syllable. This positioning might have contributed to the less pronounced differences previously reported. Furthermore, many studies employed isolated words, and when sentences were used, words were often presented in sentence-final positions, attracting sentential stress (Pater, 1997).

### *3.2.2. Do Mandarin L1 speakers employ similar acoustic cues as English L1 speakers to distinguish stress levels? If not, could these acoustic differences be attributed to their familiarity with L1 acoustic cues?*

Lastly, this study aimed to investigate the effect of L1 acoustic cues on the acoustic realization of L2 stress patterns. While some previous research has argued that F0 serves as the primary cue for Mandarin tones (Liu & Samuel, 2004), others have suggested that, in addition to F0, cues such as duration and vowel quality are important in distinguishing between heavy syllables carrying tones, from light syllables that are tone neutral, shorter, and have reduced vowel quality (Duanmu, 2007)

Mandarin participants demonstrated acoustic patterns comparable to those of English L1 speakers, primarily using F0 and duration to distinguish stress levels, suggesting a positive transfer of their L1 features. Specifically, F0 was used to differentiate between UR and PS, as well as U and UR stress levels, so the UR can be considered similar to the neutral tone in Mandarin. However, different use of intensity by Mandarin and English L1 speakers in producing UR syllables may indicate challenges in employing L2 acoustic cues that are not as consistently used in their L1. This discrepancy could potentially signify a negative L1 transfer, where dissimilarities hinder the acquisition of L2 stress patterns. Such disparities were also observed in the location of maximum F0, utilized by Mandarin but not English L1 speakers to produce stress contrasts between PS and UR syllables. Overall, these results align with PAM, suggesting that, that similarities between L1 and L2 suprasegmental features may facilitate the acquisition of L2 suprasegmentals.

Additionally, similar use of duration to distinguish PS from U and UR syllables suggests that U and UR syllables were not perceived as sufficiently different from each other on the durational continuum. This observation may be attributed to the phonological structure of Mandarin words, where most words are monosyllabic and disyllabic, while multisyllabic words are rare. In disyllabic words, ‘stress’ occurs on the first syllable, while the second syllable is shorter (Duanmu, 2007). Consequently, Mandarin speakers used duration as a cue to distinguish between PS and other stress levels, suggesting that the difference between U and UR syllables was not perceived as sufficiently different.

## 3.3. Final remarks

The present study provides valuable insights into the production of three distinct stress levels by L1 speakers of Mandarin, contributing to the field of second language acquisition. The findings demonstrate that L1 Mandarin speakers can employ the same acoustic cues as L1 English speakers to produce English stress levels. Namely, primary stressed, unstressed reduced, and the intermediate stress levels that encompasses secondary stressed and unstressed unreduced syllables. Both duration and F0 were used as primary acoustic cues by both groups, revealing that Mandarin speakers transfer their L1 acoustic correlates to produce L2 stress patterns. This reinforces the notion that the manipulation of shared cues in L2 facilitates the acquisition of L2 stress contrasts, offering important implications for language acquisition research.

Additionally, the study challenges the adequacy of a binary stress contrast model, which fails to account for more nuanced stress types differing from primary stressed (PS) and unstressed reduced (UR) syllables. Ladefoged (2005) argued that syllables are either stressed or unstressed, with any acoustic differences between primary and secondary stressed syllables attributed to their position within the word. In contrast, the binary contrast could be based on vowel quality alone, distinguishing reduced from unreduced syllables (Beckman & Edwards, 1994). The present study, therefore, provided evidence that does not support such binary categorization. However, while it reveals that the distinction between intermediate syllables and PS and UR syllables is less well-defined than the difference between PS and UR, it also highlights the limitation of using one acoustic measurement alone as a predictor for all three stress levels. This suggests that stress levels are realized through the interaction of multiple acoustic cues, in addition to considering syllable position within the word, particularly in multisyllabic words.

### *3.3.1. Limitations*

There were several limitations to this study. First, the absence of a formal language proficiency assessment raises questions about the potential influence of different proficiency levels on the acoustic measurements. While the information about the participants' length of stay in Anglophone Canada and total number of years spent learning English was collected, a comprehensive assessment of language proficiency was not conducted, following a similar approach as previous studies in the field (Banzina, 2012; Zhang et al., 2008). Including proficiency assessments in future research could help prevent potential confounding of results arising from differing levels of proficiency. Additionally, while the study focused on Mandarin and English L1 speakers, further research involving speakers of other tonal and non-tonal languages is warranted to broaden our understanding of L1 influence on the acquisition of L2 acoustic cues to stress.

Furthermore, the choice of the vowel /u/ in this study, while phonemically contrastive in both languages, may also present limitations. Notably, /u/ does not typically undergo reduction in Mandarin, and Canadian speakers produce it with fronting. Choosing different vowels in the future may provide more reliable results in terms of vowel quality. Another limitation was the number and type of stimuli used. While comparing two syllables in disyllabic words gives a binary description of the two syllables (i.e., high versus low F0 or intensity), multisyllabic approach provides a more nuanced view of acoustic cues and their interactions. The current study did not control for the syllable structure and position, including the phonological environment of the target vowel, and the patterns observed here could represent the overarching acoustic cues across words of varying lengths and syllable structures. For example, the absence of intensity differences may suggest that intensity is not as strong of a correlate compared to duration and F0. However, accounting for syllable structure and stress position could yield different results, for instance, non-primary stressed syllables preceding the primary stressed syllable in multisyllabic words may be acoustically more prominent than those following it (Plag et al., 2011). Additionally, vowel /u/ that is preceded by a palatal /j/ may have a longer duration than that same vowel preceded by a stop (Padgett, 2008; Peterson & Lehiste, 1960). While these patterns are inherently constrained by the lexicon, and while they were distributed across each stress level, a more nuanced selection of stimuli items could perhaps offer different insights into the relationship between syllable structure and acoustic characteristics across different stress levels. For example, by including a selection of multisyllabic words with SS and UU syllables occurring both before and after the primary stressed syllable, while controlling for the number of syllables, and including a broader range of vowels found in both L1 and L2 phonetic inventories, as well as only L1 phonetic inventory, future research could offer a clearer understanding of how syllable structure and stress levels influence acoustic cues in stress production. This would nevertheless require the use of more low-frequency words in English, or perhaps the use of nonword stimuli.

Additionally, Banzina (2012) points out that the SS syllables analyzed in a study by Cutler et al. (2007) were, in fact, UU syllables given them being immediately adjacent to PS syllables. This calls for improved consistency and transparency in research on stress levels. Another direction for future research is to present an equal number of different stress levels within multisyllabic words, as well as analyse any differences between same syllables positioned before and after a PS syllable.

### *3.3.2. Conclusions*

Overall, this study contributes to advancing our understanding of L2 stress acquisition, as well as L1 and L2 stress production. In particular, the findings show that English L1 and Mandarin L1 speakers make acoustic distinctions between three different levels of English stress, which represents a significant contribution to the field. Specifically, the results indicate a greater consistency in acoustic correlates marking contrasts between primary stressed and unstressed reduced syllables compared to contrasts between intermediate and other syllables. Moreover, the findings offer insights into the acoustic cues utilized by L1 Mandarin speakers in acquiring English stress patterns, revealing the potential influence of shared acoustic cues from their native language in shaping their approach to L2 stress acquisition. This suggests that similarities in acoustic patterns between L1 and L2, particularly in suprasegmentals like lexical stress and tones, may facilitate the acquisition of different suprasegmentals in L2.

Stepping back, this study holds practical implications for teaching and language learning. Educators and language instructors can leverage the recognition of shared acoustic cues between L1 and L2 stress patterns to design targeted instruction that focuses on L2 learner's existing knowledge of L1 suprasegmental features that will facilitate L2 acquisition. Specifically, focusing on acoustic features such as fundamental frequency (F0) and duration may enhance Mandarin L2 learners' acquisition of English stress patterns. Furthermore, identifying differences in the acoustic cues between Mandarin and English L1 speakers provides valuable insights into areas that require additional attention in L2 instruction for Mandarin L1 speakers. Notably, acoustic features like intensity, less important correlate in their L1, and peak F0 location, more familiar in their L1, warrant focused instruction to improve L2 stress production. Of particular interest is the challenge with peak pitch alignment differences in primary stressed syllables and unstressed unreduced syllables, despite Mandarin L1 speakers' ability to use F0 for L2 stress contrasts in a native-like manner. Consequently, instructors may need to provide clear guidance on intonation and stress patterns.

Furthermore, this study contributes to theoretical frameworks regarding the acquisition of L2 suprasegmental features. By examining the influence of shared acoustic cues from L1 on L2 stress production, this research provides empirical support for theories of L2 acquisition suggesting that L1-L2 similarities facilitate L2 acquisition, notably the Perceptual Assimilation Model. These findings provide further support for extending the model beyond L2 speech perception of segmental properties to encompass L2 production of suprasegmentals.

# References


Altmann, H. (2006). *The perception and production of lexical stress: A cross-linguistic experimental study* [Unpublished doctoral dissertation]. University of Delaware.

Anderson-Hsieh, J., & Venkatagiri, H. (1994). Syllable Duration and Pausing in the Speech of Chinese ESL Speakers. TESOL Quarterly, 28(4), 807–812. doi:10.2307/3587566

Banzina, E. (2012). *The Role of Secondary-stressed and Unstressed-unreduced Syllables in Word Recognition: Acoustic and Perceptual Studies with Russian Learners of English* [Doctoral dissertation, Bowling Green State University]. Communication Disorders Ph.D. Dissertations. 3.

Banzina, E., Dilley, L. C., & Hewitt, L. E. (2016). The Role of Secondary-Stressed and Unstressed-Unreduced Syllables in Word Recognition: Acoustic and Perceptual Studies with Russian Learners of English. *Journal of psycholinguistic research*, *45*(4), 813–831. doi:10.1007/s10936-015-9377-z

Beckman, M. E. (1986). *Stress and Non-Stress Accents*. Dordrecht, the Netherlands: Foris Publications. doi:10.1515/9783110874020

Beckman, M. E. & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form* (7-33). Cambridge: Cambridge University Press.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp.167–200). New York, NY: Timonium

Best, C., & Tyler M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. Munro (Eds.), *Language experience in second language speech learning: in honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins Publishing.

Boberg, C. (2010). *The English language in Canada: Status, history and comparative analysis*. Cambridge University Press.

Boberg, C. (2019). A Closer Look at the Short Front Vowel Shift in Canada. *Journal of English Linguistics*, 47(2), 91-119. doi:10.1177/0075424219831353

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* (Version 6.4.05) [Computer program]. Retrieved January 27, 2024, from http://www.praat.org/

Braun, B., Lemhöfer, K., & Mani, N. (2011). Perceiving unstressed vowels in foreign-accented English. *The Journal of the Acoustical Society of America*, 129, 376-387. doi:10.1121/1.3500688

Campbell, N., & Beckman, M. (1997). “Stress, prominence, and spectral tilt,” in *Proceedings of ESCA Workshop on Intonation: Theory, Models and Applications*, edited by A. Botinis, G. Kouroupetroglou, and G. Carayiannis, Athens, pp. 67–70.

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

Chen, Yang & Robb, Michael. (2004). A study of speaking rate in Mandarin speakers of American English. *Asia Pacific Journal of Speech, Language and Hearing*. 9. 173–188. doi:10.1179/136132804805575813.

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. doi:10.1121/1.1356023

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. doi:10.1044/2014_JSLHR-L-13-0279

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*, 207–28.

Corley, G. A. (2019). *Stress production by Mandarin speaking learners of English* (Publication No. 13427306) [Doctoral dissertation, University of Wisconsin--Madison]. Linguistics and Language Behavior Abstracts (LLBA); ProQuest Dissertations & Theses A&I.

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

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

Cutler, A., & Darwin, C. J. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception and Psychophysics*, 29, 217–224.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology*, 14(1), 113–121.

Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). Dutch listeners' use of suprasegmental cues to English stress. In J. Trouvain, & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetics Sciences* (ICPhS 2007) (pp. 1913–1916). Dudweiler: Pirrot.

Derakhshan, A., & Karimi, E. (2015). The Interference of First Language and Second Language Acquisition. *Theory and Practice in Language Studies*, 5, 2112–2117. doi:10.17507/TPLS.0510.19

Derwing, T. M. (2008). Curriculum issues in teaching pronunciation to second language learners. In J. G. Hansen-Edward, & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 347-369). Amsterdam: John Benjamins. doi:10.1075/sibil.36.17der

Duanmu S. (2007). *The phonology of standard Chinese*. New York: Oxford University Press.

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, 175–191.

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

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, 39(3), 399–423.

Flege, J. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233–277). Timonium, MD: York Press.

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, 35–62.

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). Cambridge: Cambridge University Press.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, 10, 2973–2987.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 27, 765–768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1(2), 126–152.

Fry, D.B. (1965). The dependence of stress judgments on vowel formant structure. In E. Zwirner and W. Bethge (Eds.), *Proceedings of the 5th International Congress of Phonetic Sciences, Münster, 16–22 Aug., 1964* (pp. 306–311). Basel: S. Karger.

Garcia, G. D. (2020). Language transfer and positional bias in English stress. *Second Lang. Res.* 36 (4), 445–474. doi:10.1177/0267658319882457

Guion, S. G. (2005). Knowledge of English word stress patterns in early and late Korean-English bilinguals. *Studies in Second Language Acquisition, 27*(4), 503–533. doi:10.1017/S0272263105050230

Guion S. G. (2003). The vowel systems of Quichua-Spanish bilinguals. Age of acquisition effects on the mutual influence of the first and second languages. *Phonetica*, *60*(2), 98–128. doi:10.1159/000071449

Guo, X. & Chen, X. (2022). Perception of English Stress of Synthesized Words by Three Chinese Dialect Groups. *Frontiers in Psychology*, 13. doi:10.3389/fpsyg.2022.803008

Hayes, B. (1995). *Metrical stress theory: principles and case studies*. University of Chicago Press, Chicago.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*(4), 548–567. doi:10.1006/jmla.2000.2755

Jung, Ye-Jee & Rhee, Seok-Chae. (2018). Acoustic analysis of English lexical stress produced by Korean, Japanese and Taiwanese-Chinese speakers. *Phonetics and Speech Sciences.* 10. 15-22. doi:10.13064/KSSS.2018.10.1.015.

Karim, K. (2010). First Language (L1) Influence on Second Language (L2) Reading: The Role of Transfer. *WPLC,* 17, 49–54.

Kelly, M. H. & Bock, J. K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 389–403.

Kröger, B. J., Bekolay, T., & Cao, M. (2022). On the emergence of phonological knowledge and on motor planning and motor programming in a developmental model of speech production. *Front. Hum. Neurosci*. 16, 844529. doi:10.3389/fnhum.2022.844529

Kuo, C., & Weismer, G. (2016). Vowel reduction across tasks for male speakers of American English. *The Journal of the Acoustical Society of America*, *140*(1), 369. doi:10.1121/1.4955310

Ladefoged, P. (2005). *A course in phonetics* (5th ed.). Boston: Heinle.

Lee, B., Guion, S. G., & Harada, T. (2006). Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition*, 28, 487–513.

Lehiste I. (1970). *Suprasegmentals*, Cambridge, Massachusetts & London, UK. United States: MIT Press

Lenth, R. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 1.8.7) [R package]. Retrieved from https://CRAN.R-project.org/package=emmeans

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and brain sciences*, *22*(1), 1–75. doi:10.1017/s0140525x99001776

Li, J. J., & Grigos, M. I. (2023). Production of English lexical stress by Mandarin speakers: Acoustics and kinematics. *Second Language Research*, *39*(3), 667-695. doi:10.1177/02676583211066290

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

Liberman, A. M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249–336.

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

Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Attention Perception & Psychophysics*, 62(2), 253–265.

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. doi:10.1006/jmla.1996.2472

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. In *Proceedings of the 18th Conference of the International Speech Communication Association*.

Morton, J., & Jassem, W. (1965). Acoustic correlates of stress. *Language and speech*, *8*(3), 159–181. doi:10.1177/002383096500800303

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49(1), 285–310. doi:10.1111/0023-8333.49.s1.8

Nagle, C. L., & Baese-Berk, M. M. (2022). Advancing the state of the art in l2 speech perception-production research: Revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition*, 44(2), 580–605. doi:10.1017/S0272263121000371

Nguyen, A.T.T., & Ingram, J. (2005). Vietnamese acquisition of English word stress. *TESOL Quarterly*, 39(2), 309–319.

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

Pater, J. (1997). Metrical Parameter Missetting in Second Language Acquisition. In S.J. Hannahs and Martha Young-Scholten (eds.). *Focus on Phonological Acquisition*. Amsterdam: John Benjamins.

Pas, B., & Zonneveld, W. (2004). L2 parameter resetting for metrical systems. *Linguistic Review*, 21(2). 125-170. doi:10.1515/tlir.2004.003.

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

Plag, I., Kunter G., & Schramm, M. (2011). Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics*, 39(3), 362–374.

Shen X. S. (1993). Relative duration as a perceptual cue to stress in Mandarin. *Language and speech*, *36 (4)*, 415–433. doi:10.1177/002383099303600404

Sluijter, A. M. C., & Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 4, 2471–2485.

Stefanich, S., & Cabrelli, J. (2021). The Effects of L1 English Constraints on the Acquisition of the L2 Spanish Alveopalatal Nasal. *Frontiers in psychology*, 12. doi:10.3389/fpsyg.2021.640354

Tiee, H. H.-Y. (1979). The productive affixes in Mandarin Chinese morphology. *Word*, 30(3), 245-255. doi:10.1080/00437956.1979.11435670

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language, 38*(2), 133–149. doi:10.1006/jmla.1997.2548

Wells, J.C. (2008). *Longman Pronunciation Dictionary.* Pearson Longman.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, 27. 55-105.

Yakup, M. (2018). The acquisition of English stress by Kazakh-Russian bilinguals: The role of dominant language. *The Journal of the Acoustical Society of America, 144*(3), 1726-1726. doi:10.1121/1.5067659

Yeung, A.H., Baek, H., Takahashi, C., Buttner, S., Hwang, J., & Broselow, E.I. (2020). Too little, too late: A longitudinal study of English corrective focus by Mandarin speakers. *Proceedings of the Linguistic Society of America*, 5(1), 270. doi:10.3765/plsa.v5i1.4727

Yu, V., & Andruski, J. (2010). A cross-language study of perception of lexical stress in English. *Journal of Psycholinguistic Research*, 39(4), 323–344.

Yuan, Man & Cheng, Wei. (2017). The Acquisition of English Lexical Stress by Chinese-speaking Learners: An OT Account. *English Language Teaching*, 10, 140. doi:10.5539/elt.v10n7p140.

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

# Appendix A: Figures

**Figure A1 Mean fundamental frequency (F0) across PS, U, and UR stress types.**

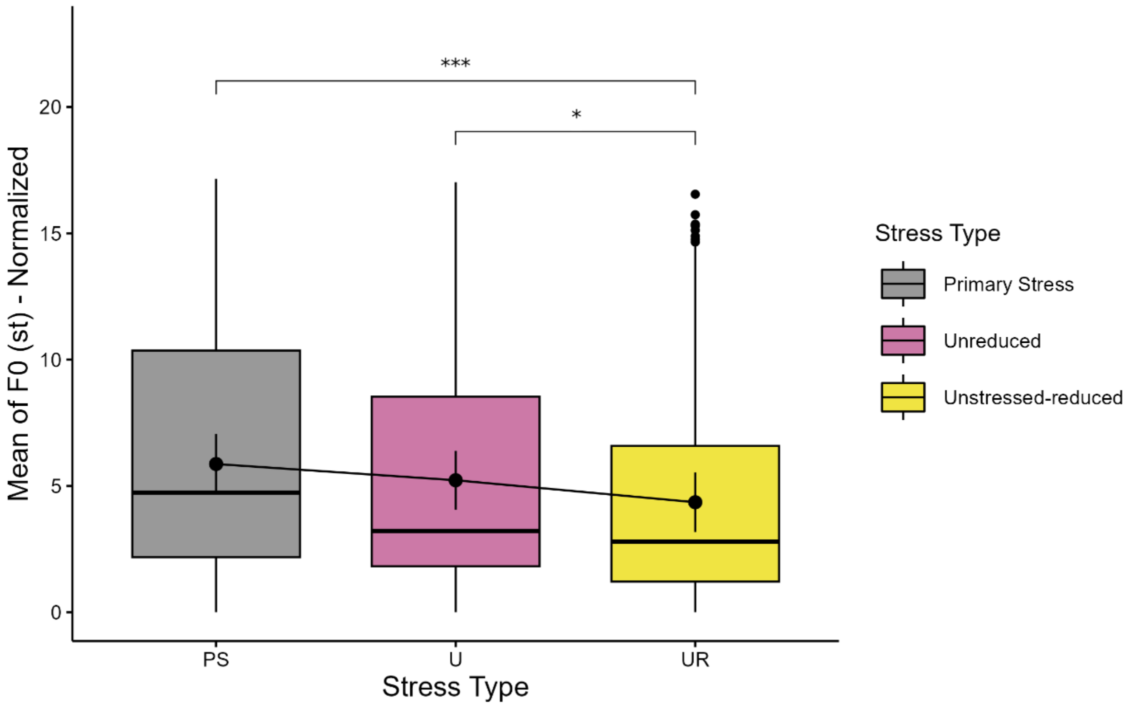


Figure A1 visualizes main effects of stress type on F0, showing significant differences between PS (higher F0) and UR (lower F0), as well as U (higher F0) and UR (lower F0) syllables. Mean F0 values are normalized for speaker-specific differences across PS, U, and UR stress types: Each boxplot represents the distribution of F0 values within each stress type, with the center line indicating the median, the box edges representing the interquartile range (IQR), and the whiskers extending to 1.5 times the IQR. The dots represent the estimated mean F0 values from the model, and error bars denote the 95% confidence intervals around these estimates. Pairwise comparisons are depicted with asterisks indicating significance levels (**p* < .05, ****p* < .001).

**Figure A2 Mean duration across three stress types and two language groups.**

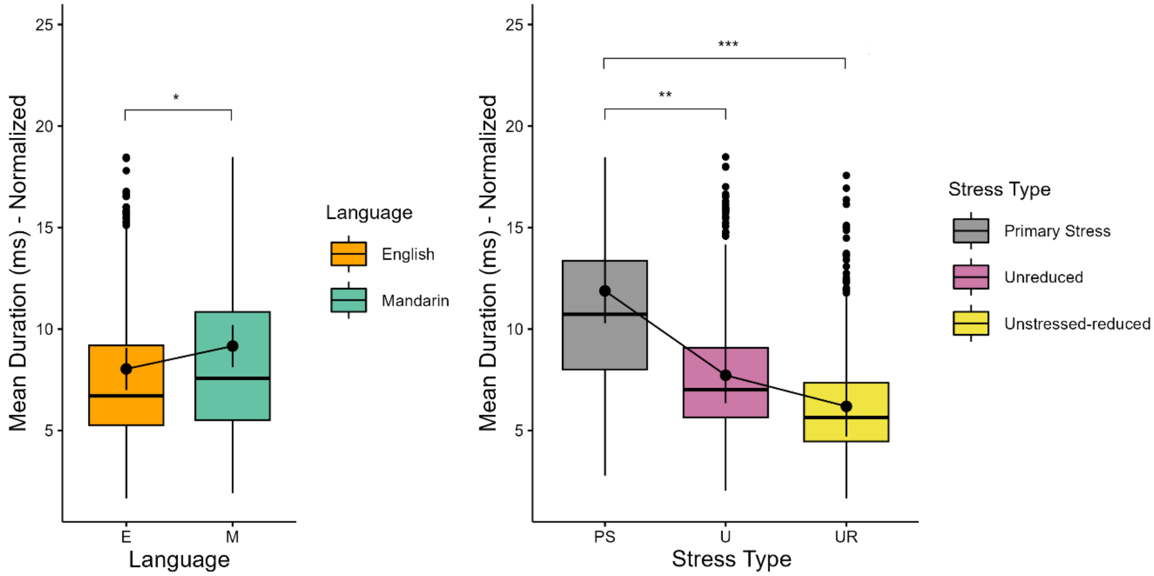


Figure A2 visualizes main effects of language and stress type on duration. It shows longer duration for Mandarin L1 speakers across all stress types when compared to English L1 speakers. The figure also shows that PS syllables were significantly longer than U and UR syllables. Each boxplot represents the distribution of duration values within each language group or stress type, with the center line indicating the median, the box edges representing the interquartile range (IQR), and the whiskers extending to 1.5 times the IQR. Dots represent the estimated mean duration values, while error bars denote the 95% confidence intervals around these estimates. Pairwise comparisons of mean duration values between languages are depicted with asterisks indicating significance levels (**p* < .05, ***p* < .01, ****p* < .001).

**Figure A3 Mean intensity across PS, UU, and UR syllables spoken by L1 English speakers and Mandarin L2 speakers.**

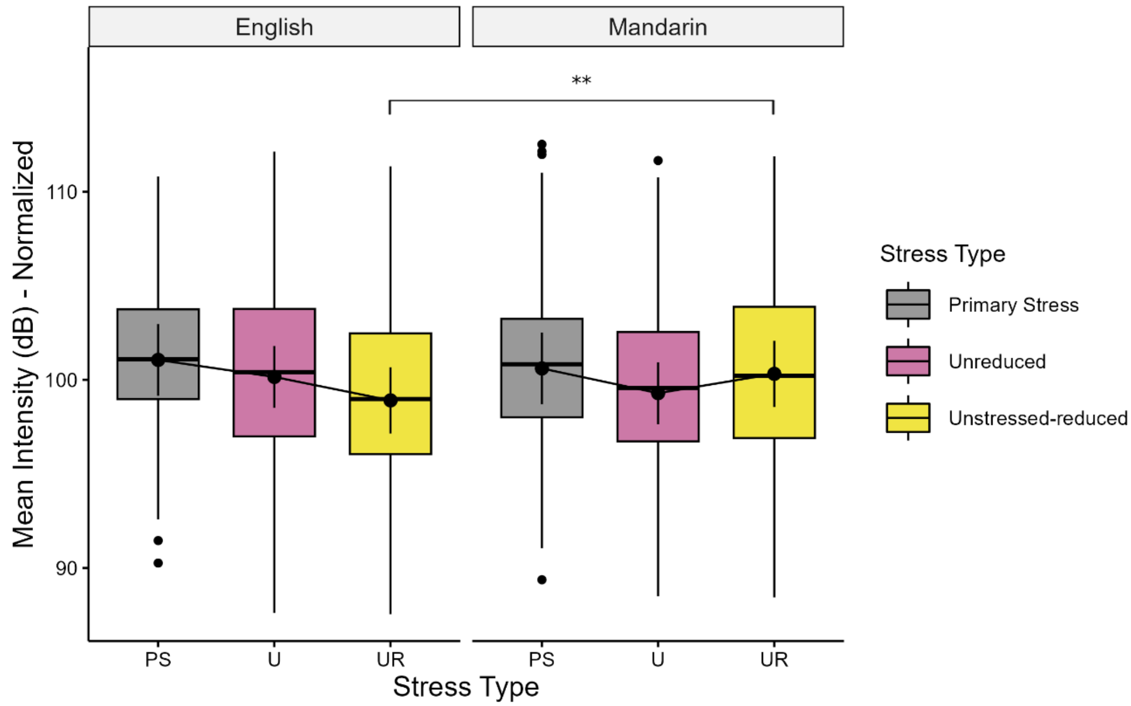


Figure A3 illustrates interaction effects of stress type and language on intensity, showing that English L1 speakers produced UR syllables with lower intensity than Mandarin L1 speakers. Each boxplot represents the distribution within each stress type and language group, with the center line indicating the median, the box edges representing the interquartile range (IQR), and the whiskers extending to 1.5 times the IQR. Dots represent the estimated mean values, and error bars denote the 95% confidence intervals around these estimates. Pairwise comparisons between stress types and language groups are depicted with asterisks indicating significance levels (***p* < .01).

**Figure A4 F2 at midpoint across L1 English L1 Mandarin speakers.**

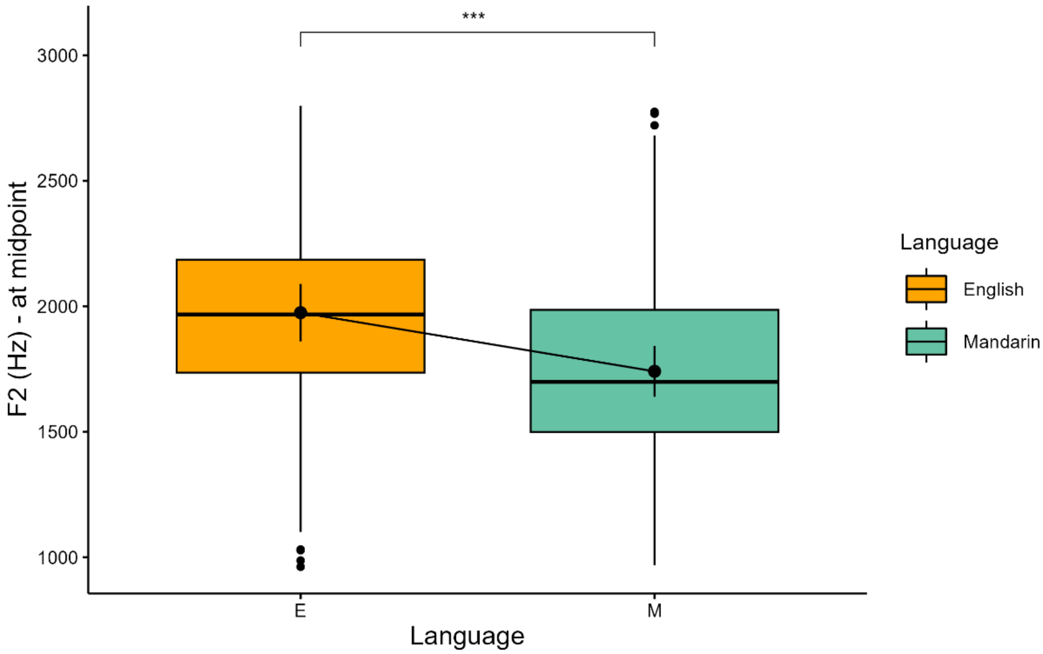


Figure A4 illustrates main effects of language on F2, showing that Mandarin L1 speakers produced the vowel /u/ with a lower F2 compared to English L1 speakers. Each boxplot represents the distribution within each language group, with the center line indicating the median, the box edges representing the interquartile range (IQR), and the whiskers extending to 1.5 times the IQR. Dots represent the estimated mean values, and error bars denote the 95% confidence intervals around these estimates. Pairwise comparison between languages is depicted with asterisks indicating significance level (***$p$ < .001).

**Figure A5** **Location of maximum F0 as syllable percentage across PS, UU, and UR syllables spoken by L1 English speakers and Mandarin L2 speakers.**

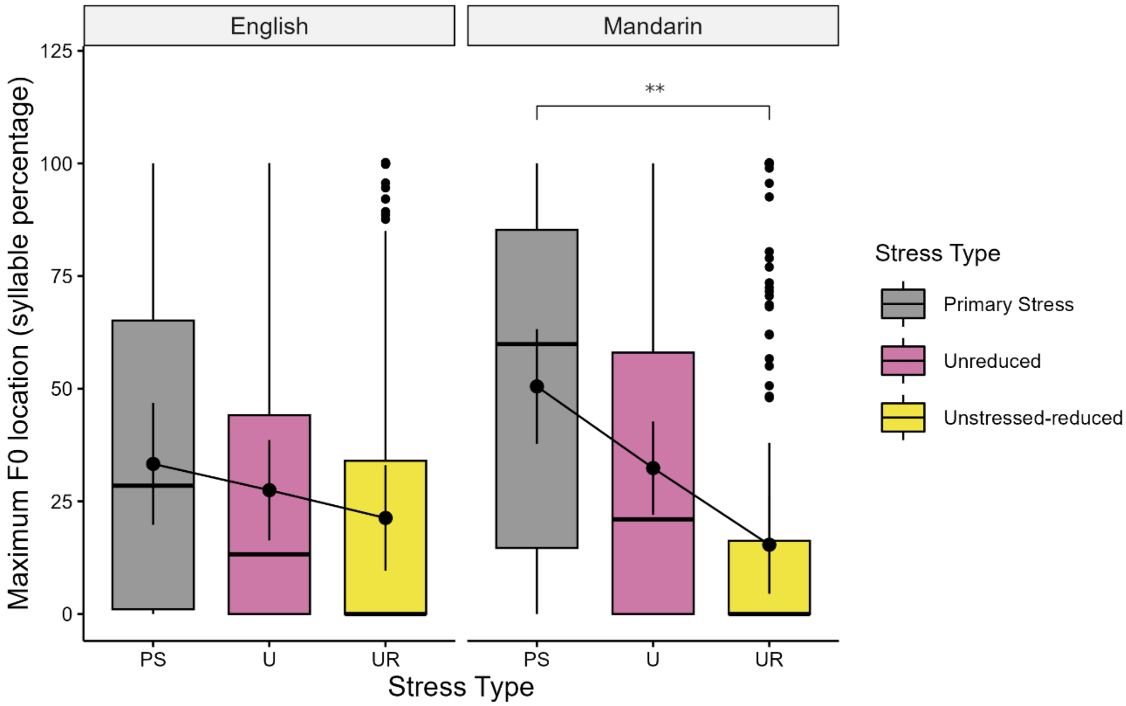


Figure A5 displays the distribution of maximum F0 location values expressed as a percentage of the syllable duration across three stress types for L1 English and L1 Mandarin participants. The figure visualizes the main effects of stress type, and interaction effects of stress type and language. For Mandarin L1 speakers, maximum F0 occurs later in PS syllables when compared to UR syllables. It also shows maximum F0 occurring earlier PS syllable produced by English L1 speakers compared to Mandarin L1 speakers. Each boxplot represents the distribution within each stress type, with the center line indicating the median, the box edges representing the interquartile range (IQR), and the whiskers extending to 1.5 times the IQR. Dots represent the estimated mean values, and error bars denote the 95% confidence intervals around these estimates. Pairwise comparisons between stress types and interactions of stress type and language are depicted with asterisks indicating significance levels (**$p < .01$).

# Appendix B: Linguistic Background Questionnaire

FACULTY OF ARTS AND SOCIAL SCIENCES | DEPARTMENT OF LINGUISTICS

LANGUAGE LEARNING & DEVELOPMENT LAB
QUESTIONNAIRE

**Demographic and linguistic background Questionnaire**
***Acoustic cues to production of English stress by non-native speakers of English***
**(ORE Study #30001628)**

1. What is your age?

2. How would you describe your gender?
   a. Man
   b. Woman
   c. Non-binary
   d. Prefer to self-describe: ___________
   e. Prefer not to answer

3. How long have you lived in Canada for? Please answer in years.

4. How much time have you spent living in any English-speaking countries/regions, including Canada? Please use months and/or years.

5. Other than your native language, what other languages do you speak? Please list them below

6. How many years have you been learning English?

7. How old were you when you started learning English?

8. Do you have any speech or hearing difficulties?

   a. No
   b. Yes

# Appendix C: Stimuli

| Word | IPA | Stress type | Sentence |
|---|---|---|---|
| music | ˈmjuːzɪk | PS | I don't enjoy the music she's playing. |
| opportunity | ˌɑpərˈtuːnət̬i | PS | She had the opportunity to learn a new language in college. |
| funeral | ˈfjuːnərəl | PS | Mary attended a funeral service this afternoon. |
| unusual | ʌnˈjuːʒuəl | PS | Today's weather is unusual for this time of the year. |
| nuclear | ˈnuːkliɚ | PS | These people think that nuclear energy is not very safe. |
| improve | ɪmˈpruːv | PS | He has really improved his English grades. |
| university | ˌjunɪˈvɝsət̬i | SS | They require a university education for this job position. |
| universal | junɪˈvɝːsəl | SS | We should think about universal approaches to mental health problems. |
| supervision | ˌsuːpəˈvɪʒən | SS | Children are under supervision in this daycare center. |
| particular | pəˈtɪkjələr | UR | There is one particular thing you should know. |
| population | ˌpɑːpjəˈleɪʃən | UR | My home country's population is about 80 million. |
| instrument | ˈɪnstɹəmənt | UR | She plays a cool instrument in a school band. |
| volume | ˈvɑːljəm | UR | There is a large volume of books at the library. |
| irregular | ɪˈrɛg.jə.lɚ | UR | Please find two irregular verbs in this sentence. |
| monument | ˈmɑːnjəmənt | UR | Look at that big monument outside this building. |
| educate | ˈɛdʒəkeɪt | UR | Movies should also educate and not only entertain. |
| united | juˈnaɪtɪd | UU | The family was united to keep things peaceful. |
| schedule | ˈskɛdʒuːl | UU | We have a full schedule next week. |
| perfume | ˈpɝːfjuːm | UU | I like the new perfume she is wearing. |
| prostitute | ˈprɑːstətuːt | UU | They saw a prostitute outside the bar. |
| circulation | ˌsɝkjuˈleɪʃən | UU | He has bad circulation in his legs. |
| refugee | ˌrefjuˈdʒiː | UU | The book describes refugee camps very realistically. |

A complete list of stimuli items. Stimuli items were categorized as follows: PS (Primary Stressed), SS (Secondary Stressed), UU (Unstressed-unreduced), and UR (Unstressed-reduced). For analysis purposes, SS and UU were treated as a single unreduced (U) syllable.