languages

MDPI

Article

The Effect of Pitch Accent on the Perception of English Lexical Stress: Evidence from English and Mandarin Chinese Listeners

Fenqi Wang [1,*], Delin Deng [2], Kevin Tang [3,4] and Ratree Wayland [4]

[1] Department of Linguistics, Faculty of Arts and Social Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
[2] Department of Psychology and Human Development, Peabody College, Vanderbilt University, Nashville, TN 37235, USA; delin.deng@vanderbilt.edu
[3] Department of English and American Studies, Institute of English and American Studies, Faculty of Arts and Humanities, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany; kevin.tang@hhu.de
[4] Department of Linguistics, College of Liberal Arts and Sciences, University of Florida, Gainesville, FL 32603, USA; tang.kevin@ufl.edu (K.T.); ratree@ufl.edu (R.W.)
* Correspondence: fenqiw@sfu.ca

Abstract: The relative weighting of f0 and vowel reduction in English spoken word recognition at the sentence level were investigated in one two-alternative forced-choice word identification experiment. In the experiment, an H* pitch-accented or a deaccented word fragment (e.g., *AR-* in the word *archive*) was presented at the end of a carrier sentence for identification. The results of the experiment revealed differences in the cue weighting of English lexical stress perception between native and non-native listeners. For native English listeners, vowel quality was a more prominent cue than f0, while native Mandarin Chinese listeners employed both vowel quality and f0 in a comparable fashion. These results suggested that (a) vowel reduction is superior to f0 in signaling initial stress in the words and (b) f0 facilitates the recognition of word initial stress, which is modulated by first language.

Keywords: pitch accent; English lexical stress; cue weighting; speech perception; non-native listeners

Citation: Wang, Fenqi, Delin Deng, Kevin Tang, and Ratree Wayland. 2024. The Effect of Pitch Accent on the Perception of English Lexical Stress: Evidence from English and Mandarin Chinese Listeners. *Languages* 9: 87. https://doi.org/10.3390/languages9030087

Academic Editors: Lucrecia Rallo Fabra and Joan C. Mora

Received: 17 November 2023
Revised: 10 February 2024
Accepted: 17 February 2024
Published: 1 March 2024

1. Introduction

Most research on speech perception has focused on segmental contrasts rather than suprasegmental contrasts (e.g., Flege and Wayland 2019; Raphael 2021). The nature of suprasegmental processing in spoken word recognition has yet to be fully understood and adequately explained. It is well known that segmental information is essential to distinguish the target word from competitor words in spoken word recognition as available information unfolds online (Connine et al. 1994; Marslen-Wilson and Warren 1994; McQueen et al. 1994), whereas the role of prosodic information in spoken word recognition has generally received less attention. However, an increasingly large number of studies have revealed that in addition to segmental information, suprasegmental information also influences spoken word recognition (Cooper et al. 2002; Soto-Faraco et al. 2001; Van Donselaar et al. 2005).

Previous research has shown that lexical stress modulates lexical access such that word recognition is facilitated when the visual target matches the auditory prime in its stress pattern (Connell et al. 2018; Cooper et al. 2002; Reinisch and Weber 2012; Soto-Faraco et al. 2001; Van Donselaar et al. 2005). For example, using cross-modal priming experiments, Van Donselaar et al. (2005) showed that in Dutch (a language with lexical stress), a prior auditory prime with the first two syllables of the target word that were appropriately stressed (e.g., the auditory prime *okTO-* preceding the visual target *oktober*, which has penultimate stress) facilitated the recognition of the corresponding visual target word, whereas the participants found it more difficult to recognize the target word if the prior auditory prime was inappropriately stressed (e.g., *oktober* preceded by *OKto-*).

However, since multiple acoustic cues can signal lexical stress (see Section 1.2 below), the dynamic relationship between these cues and the influence of these cues on spoken

word recognition remains underinvestigated (Chrabaszcz et al. 2014; Zhang and Francis 2010). Additionally, the effect of the sentence-level pitch accent on the processing of lexical stress has been relatively unexplored (e.g., Liu 2019). Specifically, it is unclear how sentence-level pitch accents affect listeners' processing of stress in spoken word recognition and how it interacts with other cues to lexical stress such as vowel quality and fundamental frequency (f0). In addition, unlike English, Mandarin Chinese, as a tonal language, uses f0 as a cue to distinguish meanings between words. Research has demonstrated that the role of suprasegmental cues (e.g., f0) in distinguishing word meanings in a native language (L1) impacts the way learners of a second language (L2) apply these cues in the L2 (e.g., Qin et al. 2017; Wang et al. 1999; Wayland and Guion 2004). If so, we are curious about whether native Mandarin Chinese listeners who speak American English can still utilize f0 to discern English lexical stress despite the presence of sentence-level pitch accents and how their use of f0 and vowel quality differs from that of native English listeners.

1.1. Stress and Pitch Accent in English

Lexical stress refers to the increased prominence of a syllable within a word, which can result in lexical contrast (e.g., in English, *REcord* as noun vs. *reCORD* as verb). Previous studies on English lexical stress have identified fundamental frequency, intensity, duration, and vowel quality as acoustic correlates. These cues are utilized by English listeners for stress perception (Beckman 1986; Bolinger 1989; Campbell and Beckman 1997; Fry 1955, 1958, 1965; Lieberman 1965; Sluijter and Van Heuven 1996a, 1996b). However, f0 also serves as a cue to pitch accents within a sentence intonation (Beckman and Ayers 1997; Beckman and Hirschberg 1994; Ladd 2008). These pitch accents, which align with stressed syllables, can modify or even override the surface f0 cue of stress in these syllables (Fry 1958).

A pitch accent is added to a word to increase its prominence in the sentence, with the discourse structure dictating whether words receive a pitch accent and which pitch accent they have (Ladd 2008). Acoustically, pitch-accented words show an expanded pitch range, longer duration, and higher intensity compared to deaccented and post-accented words (Bolinger 1989; Gussenhoven 1994; Ladd 2008; Selkirk 1995). There are five types of pitch accents in American English described in the ToBI (tone and break indices) transcription: H*, L*, L* + H, L + H*, and H + !H* (Beckman and Ayers 1997), where H denotes a high-pitch target, L a low-pitch target, and * alignment to the stressed syllable. However, only H* (presentational pitch accent) and L + H* (contrastive pitch accent) are relevant for this study. H* is a high single-tone pitch accent with a gradual rise into the peak of the tone from the word onset (Beckman and Hirschberg 1994). L + H* is a bitonal low tone and a high tone on the accented (i.e., stressed) syllable; it contains an L target before leading up to the H* target (Beckman and Hirschberg 1994). Typically, the H* pitch accent is used to present new information, while the L + H* pitch accent is used for contrastive references. For example, in context (a), Speaker B is providing new information as an answer to the question *What did Mary say?*; in context (b), Speaker B is trying to make it clear that it is Mary, not Jane, who said *archive*.

a. H* on *Mary* and *archive*

Speaker A: What did Mary say?

Speaker B: Mary said "archive".

H* H*

b. L+H* on *Mary*

Speaker A: Jane said "archive".

Speaker B: No, Mary said "archive".

L+H*

Importantly, unlike the H* accent, the L + H* accent triggers deaccenting on the following words within the same intonational phrase. Thus, for Speaker B's reply, in context (a), *Mary* receives a presentational H* pitch accent and *archive* receives an H* pitch accent, while in context (b), only *Mary* receives an L + H* pitch accent and *archive* is deaccented. Acoustically, the stressed syllable of *archive* in context (a) is marked by a higher f0 relative to the unstressed syllable, whereas the f0 realized on the stressed syllable of the deaccented *archive* in context (b) does not exhibit this distinction (Beckman and Ayers 1997; Beckman and Hirschberg 1994; Ladd 2008). In other words, the F0 cue is present in H*-accented words to indicate stress, but this cue is absent in deaccented words.

To investigate the effects of different intonation contours (i.e., falling: declarative; rising: yes/no question) on native English and Mandarin Chinese listeners' perception of stress position, Liu (2019) conducted a forced-choice word identification task with 12 noun–verb pairs differing only in their stress pattern (e.g., *PERmit* vs. *perMIT*) in the carrier sentence (i.e., *This word is* ___./?). The findings indicated that Mandarin speakers had difficulty accurately perceiving stress positions when high tones were not aligned with the stressed syllable. In comparison, native English speakers also struggled to perceive stress accurately, though their misperception was reduced when the stress was on the initial syllable. It can be concluded that listeners from different L1 backgrounds may be affected differently by English sentential prosody in terms of cue usage. Moreover, native Mandarin Chinese listeners might be influenced by tone in their perception of English lexical stress.

1.2. Weighting of Acoustic Cues to English Lexical Stress

According to the cue-weighting theory, speech perception is a multidimensional process in which listeners may attend to multiple acoustic cues simultaneously to perceive sound contrasts, but the weight of each cue depends on its informativeness in signaling the contrasts in the language (Francis et al. 2008; Francis and Nusbaum 2002; Guion and Pederson 2007; Holt and Lotto 2006; Iverson et al. 2003). Previous findings conflict on the weight of segmental and suprasegmental cues to lexical stress in English, and the interaction among these cues remains to be clarified with more research.

As already mentioned above, multiple acoustic cues have been proposed to signal lexical stress in English: vowel quality, vowel duration, pitch (f0), and intensity (Beckman 1986; Fry 1955, 1958, 1965; Lieberman 1965; Sluijter and Van Heuven 1996a, 1996b). Compared with unstressed syllables, stressed syllables have been shown to have greater intensity and longer duration; additionally, all stressed syllables contain full vowels, whereas unstressed syllables tend to contain reduced vowels. However, because words may or may not be accented at the sentence level, f0, therefore, may not consistently be available as a cue to signal lexical stress (see Section 1.1 above), which motivates this study to investigate the effect of pitch accent on English lexical stress perception.

Many previous studies have investigated how suprasegmental cues (i.e., f0, duration, and intensity) influence lexical stress perception (Beckman 1986; Fry 1955; Lieberman 1965). For example, Fry (1955) manipulated the vowel duration ratio and intensity ratio between the two syllables in words with different stress placements according to word class (e.g., *OBject* vs. *obJECT*) to examine how vowel duration and intensity cues influence listeners' judgments of stress. The results showed that although the duration ratio and the intensity ratio are both cues to stress, the duration ratio emerged as a more effective cue than the intensity ratio. To further investigate the acoustic correlates of English lexical stress, Fry (1958) conducted three perceptual experiments to probe the effect of change in duration, intensity, and f0 on stress judgment with word pairs (e.g., *SUBject* vs. *subJECT*). The first experiment manipulated the duration ratio and intensity ratio as in Fry (1955), and its results showed that duration was a more effective cue to stress than intensity. This outcome aligns with Fry's (1955) findings. The second experiment manipulated the duration ratio and the f0 ratio between the two vowels in the words to see how these two cues may interact with each other. The findings revealed that the presence of a step change in f0 (i.e., which syllable in the word had a higher f0 than the other due to the step

change in f0) significantly influenced stress perception rather than the magnitude of the f0 change. The same effect of the change in the duration ratio was also observed as in the first experiment. The third experiment included an f0 change (i.e., linear and curvilinear) within one syllable to simulate the effect of intonation on the realization of stress in words. The results demonstrated that "sentence intonation is an overriding factor in determining the perception of stress and that in this sense the fundamental frequency cue may outweigh the duration cue" (p. 151).

Like suprasegmental cues, segmental cues also contribute to the perception of lexical stress in English. Specifically, native English listeners rely more on segmental cues than suprasegmental cues in stress perception because English lexical stress is consistently signaled by segmental information (full vs. reduced vowels) (Cutler 1986). This hypothesis was supported by the findings that stress perception is more affected by changes in segmental cues than by changes in suprasegmental cues (Cutler 1986; Cutler and Clifton 1984; Fear et al. 1995). To investigate how vowel quality interacts with other acoustic cues in stress perception, Zhang and Francis (2010) conducted three experiments investigating the perception of English lexical stress by native English and Mandarin Chinese listeners. They used the production of *DEsert* (noun) and *deSERT* (verb) as the tokens. In each experiment, two acoustic cues realized on the first syllable *de-* were manipulated in seven steps from stressed to unstressed values, while the value of the second syllable *-sert* was held constant. Combining the findings of the three experiments, it was concluded that the four acoustic cues were used by native English listeners to perceive English lexical stress, but vowel quality was a stronger cue to stress than other cues. Compared to native English listeners, native Mandarin Chinese listeners were more influenced by pitch contour conditions when processing vowel quality and F0. Specifically, they tended to use vowel quality and F0 as a combinatorial cue to stress in the natural pitch contour condition but as separate cues in the flat pitch contour condition. Furthermore, when native Mandarin Chinese listeners processed vowel quality with duration, they tended to behave like native English listeners in terms of cue weight in both pitch contour conditions. The results of both native English and Mandarin Chinese listeners suggested that there is a potential hierarchy of cues in English stress perception, but the study did not address the relative weighting of suprasegmental cues to lexical stress in English since this study did not compare suprasegmental cues to one another.

1.3. Lexical Stress in Spoken Word Recognition

Research on lexical processing has demonstrated that multiple competing word candidates are activated during the retrieval of a target word in spoken word recognition (Connine et al. 1994, 1997; Goldinger et al. 1989; Marslen-Wilson 1990; Marslen-Wilson and Warren 1994; McQueen et al. 1994). To distinguish between the target word and competing word candidates, rapid integration of all available information is required (Marslen-Wilson and Warren 1994; McQueen et al. 1999). Early research on the use of lexical stress in spoken word recognition had argued that suprasegmental cues to lexical stress might not constrain lexical access in English because segmental cues such as vowel quality are highly reliable for English listeners to identify lexical stress; English stressed syllables always contain non-reduced vowels, whereas unstressed syllables tend to contain reduced vowels (Halle and Vergnaud 1987; Hammond 1995).

To investigate the effect of lexical stress in spoken word recognition, Small, Simon, and Goldberg (Small et al. 1988) conducted a phoneme monitoring task with target phonemes following disyllabic homographs (e.g., *CONtract* and *conTRACT*) and non-homographs (e.g., *NAPkin*, regular words with only one legal stress placement). The task was to examine participants' response speed to the target phoneme preceded by correctly or incorrectly stressed homographs and non-homographs (e.g., *Mary was a recent CONvert/conVERT/NAPkin (f)rom Catholicism*; /f/ is the target phoneme). The results showed that the participants' response speed was slower when the target phonemes were preceded by incorrectly stressed non-homograph words than when the target phonemes were preceded by correctly stressed non-

homograph words, while the participants' response speeds to detect the target phonemes after correctly or incorrectly stressed homograph words were similar. The authors explained that the absence of a stress effect for homograph words was probably due to the participants' failure to pay enough attention to the lexical stress of the test words. This study suggested that prosodic information may be helpful in spoken word recognition since there was an effect of stress for non-homograph words.

However, the methods used in previous research may have made it difficult to tap into the effect of suprasegmental information on spoken word recognition. One concern about the earlier studies is that the participants were asked to respond after hearing the whole word. It may be the case that, in English, suprasegmental information becomes less useful for spoken word recognition once all the necessary segmental information has been heard, because there are very few segmentally near-identical but suprasegmentally distinct words in English. The significance of suprasegmental information for spoken word recognition can be quantified using a measure called functional load, which measures the extent to which a language utilizesa contrast (Hockett 1955; Martinet [1960] 1964). Surendran and Levow (2004) quantified the functional load of consonants, vowels, and stress/tones in English, Dutch, German, and Mandarin Chinese over two levels of phonological units: words and syllables. If the functional load of a contrast is higher than that of another contrast, then it encodes more information by the language. Comparing the functional load of different contrasts can reveal whether one type of contrast encodes more information than another, e.g., the amount of information encoded by stress can be higher or lower than those by other contrasts (consonants and vowels). The relative amount of information encoded by the same set of contrasts can vary depending on the linguistic unit being considered to carry information, e.g., relative to consonants and vowels, the amount of information encoded by stress in words might be different from that in syllables. They observed that, for English, the word-level functional load of stress is 490 times lower than that of vowels and 213 times lower than that of consonants, while the syllable-level functional load of stress is 5 times lower than that of vowels and 11.5 times lower than that of consonants. This functional load study suggests that English word structure makes stress relatively less important compared to syllable structure. Therefore, it would be better to tap into the effect of suprasegmental information on word recognition with words whose first syllables are segmentally identical but suprasegmentally distinct (e.g., *DIStance* vs. *disTINCT*). If suprasegmental information is used at an early stage in spoken word recognition, native listeners may respond faster to words whose first syllable matches the input segmentally and suprasegmentally compared with words whose first syllable only matches the input segmentally.

To further investigate the use of segmental and suprasegmental cues to stress in spoken word recognition, Connell et al. (2018) conducted a visual-world eye-tracking experiment with native English listeners and Chinese learners of English. They found that English listeners can use segmental and suprasegmental cues together to recognize English words, while Mandarin listeners were unable to use stress to access words when the segmental and suprasegmental cues were different between the first syllable of the target and competitor words. The authors ascribed the inability of Mandarin listeners to use suprasegmental cues in the vowel-reduction condition to the fact that reduced vowels in Standard Mandarin cannot appear in word-initial position, but this still does not fully explain Mandarin listeners' performance in the vowel-reduction condition.

Although previous studies have discussed how different cues to English lexical stress affect spoken word recognition, it is unclear whether (and if so, how) segmental cues to lexical stress (vowel quality) interact with suprasegmental cues (f0) in different sentence-level pitch accent contexts in English spoken word recognition given that f0 has been shown to also be a cue to intonational pitch accents (Beckman and Ayers 1997; Beckman and Hirschberg 1994; Ladd 2008). Specifically, given the greater weight of vowel quality cues to English lexical stress compared to f0 cues (Chrabaszcz et al. 2014; Zhang and Francis 2010), it is unclear whether f0 cues have a stronger effect on the perception of lexical stress when vowel quality cues are absent. This question was investigated with a forced-choice

word identification experiment with native English listeners. Their perception of English lexical stress was examined in different sentential contexts where the target word was accented or deaccented.

The following questions guided the study:

1. Are H*-accented di- and trisyllabic words with initial stress more accurately identified than deaccented disyllabic and trisyllabic words with initial stress?
2. Is f0 as effective as vowel quality in signaling initial stress in H*-accented and deaccented di- and trisyllabic words?
3. How do native Mandarin listeners perceive English lexical stress in H*-accented and deaccented di- and trisyllabic words in comparison to native English listeners?

For question 1, it was predicted that the H* pitch accent would have an enhancing effect on the use of lexical stress in spoken word recognition when the target and the competitor word differ solely in suprasegmental cues to stress compared to when they differ in both segmental and suprasegmental cues. For question 2, it was predicted that f0 would be less effective than vowel quality in signaling lexical stress and that, compared to Mandarin Chinese listeners, English listeners would perform better in identifying the initially stressed target word when the corresponding vowel of its competitor is reduced than when it is unreduced regardless of whether the target word is accented or not. For question 3, we predicted that if there is a facilitating effect of pitch accent, native Mandarin listeners would show a shorter response time and higher accuracy when the target word is accented, and the weight of vowel quality to stress perception would be diminished to some extent.

2. Methods

2.1. Participants

This experiment tested thirty-nine native English listeners and thirty-eight native Mandarin Chinese listeners who speak American English. The responses of four native English listeners and nine non-native English listeners were excluded due to low accuracy (i.e., below 78%, which is 1 standard deviation below the mean) in the filler trials. These participants completed a language background survey to confirm English or Mandarin Chinese as their first and dominant language and to report their biographical information such as age, gender, and residence status. None of the participants included in the analyses reported speech-, hearing-, or language-related disorders, and they were monetarily compensated for their participation.

2.2. Stimuli

Participants heard a word fragment (e.g., the first and stressed syllable *AR-* of the word *ARchive*) at the end of a carrier sentence (*Mary said* ___) and had to choose one of two words on the computer screen (e.g., either *archive* or *arcade*). Two different versions of the carrier sentence were used (1 and 2 below). In the first version (1), *Mary* was produced with a presentational H* pitch accent, and the word fragment was produced with an H* pitch accent. In the second version, *Mary* was produced with a contrastive L + H* pitch accent, and the target word fragment was deaccented.

In other words, both carrier sentences ended with the first syllable of the target word (e.g., the first syllable "*AR*" in the word "*archive*"). In addition, two experimental conditions were implemented: the non-vowel-reduction and the vowel-reduction conditions. In the

non-vowel-reduction condition, the first vowel of the competitor word was produced without a vowel reduction (e.g., target—*archive* [ˈɑɹ. kaɪv]—vs. competitor—*arCADE* [ɑɹ. ˈkejd]). On the other hand, the first vowel was reduced in the vowel-reduction condition (e.g., target—*CONcept* [ˈkɑn. sɛpt]—vs. competitor—*conCERN* [kən.ˈsɝn]).

Since the availability of f0 cues to lexical stress is dependent on pitch accenting at the sentence level, this study manipulated whether the target word was accented or deaccented by using carrier sentences that elicited or did not elicit a pitch accent on the target word. As shown in Table 1, "*Mary*" in carrier sentence (a) has a presentational pitch accent (H*), while in carrier sentence (b), it has a contrastive pitch accent (L + H*). The acoustic measurements of the word *Mary* and *said* in the two carrier sentences are given in Table 2. The corresponding spectrogram of each carrier sentence is presented in Figure 1. The manipulation of the pitch accent on the subject of the sentence has consequences for the realization of the target word in carrier sentence (a), and the target word is H*-accented, so the f0 realized on the stressed syllable is higher than that on the unstressed syllable, making it easier to identify stress, while in carrier sentence (b), the target word is deaccented, so the difference in f0 between the stressed and the unstressed syllable is less pronounced, thus making stress identification challenging.

Thirty pairs of disyllabic and trisyllabic critical target and competitor words were used as the experimental words, equally distributed in vowel-reduction and non-vowel-reduction conditions (see Appendix A). Fifty-one words in the critical trials were selected from Connell et al. (2018), and the remaining words in the critical trials were selected from the CELEX lexical database (Baayen et al. 1996). The first syllable of the target words was stressed with a full vowel in both the vowel-reduction and non-vowel-reduction conditions, but the first syllable of the competitor words was unstressed without vowel reduction in the non-vowel-reduction condition and with vowel reduction in the vowel-reduction condition. The number of letters and the number of syllables were matched between the words in a pair within each condition (see Appendix B), without significant differences shown in paired-samples *t*-tests (number of letters: $t(29) = 0.135$, $p = 0.894$; number of syllables: $t(29) = 1.278$, $p = 0.211$).

Table 1. Sample test items across conditions.

Carrier Sentence	No Vowel Reduction	Vowel Reduction
H* H* a. *Mary said* ___. (Accented target)	ARchive (vs. arCADE)	CONcept (vs. conCERN)
L+H* Deaccented b. *Mary said* ___. (Deaccented target)	ARchive (vs. arCADE)	CONcept (vs. conCERN)

Table 2. Acoustic characteristics of the words in two carrier sentences.

Carrier Sentence (a)	Max f0 (Hz)	Min f0 (Hz)	Mean f0 (Hz)	Mean Intensity (dB)	Duration (ms)
Mary	231	207	222	72	327
said	221	202	208	68	214
Carrier Sentence (b)					
Mary	212	175	194	72	398
said	182	160	171	61	215

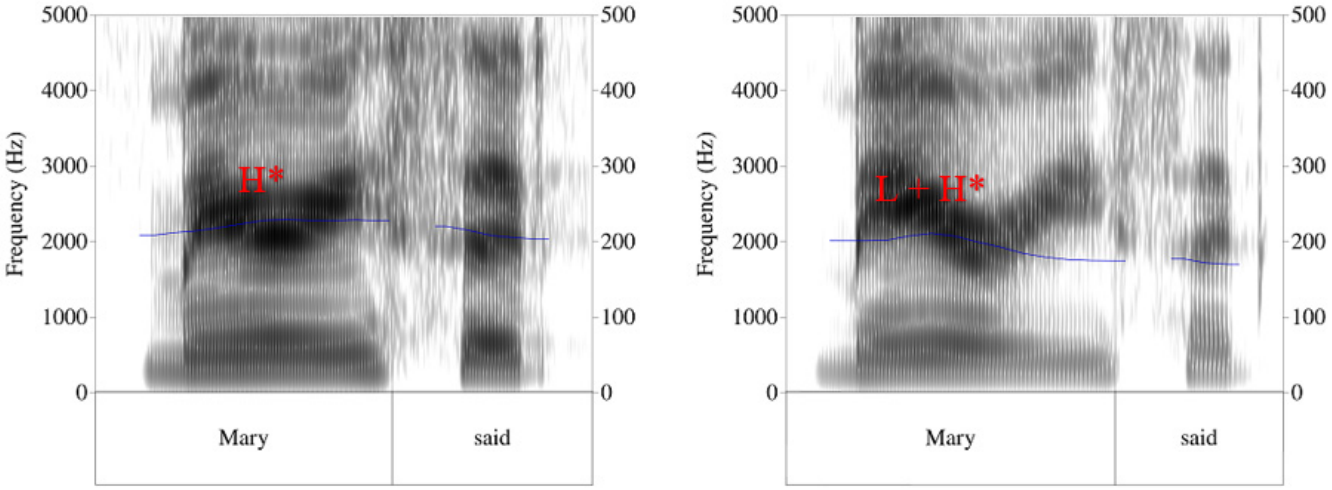

Figure 1. Spectrograms of carrier sentences with annotation. The blue lines on the spectrograms are the pitch contour. The (**left**) spectrogram is of carrier sentence with H* on *Mary*, and the (**right**) spectrogram is of carrier sentence with L + H* on *Mary*.

In addition, to control the differences between the target word and the competitor word in a pair, we also examined the word frequency and prevalence of these words. The word frequency was derived from the SUBTLEX-US corpus (Brysbaert and New 2009), which was expressed as a standardized log-transformed Zipf score (Brysbaert et al. 2018; van Heuven et al. 2014). Word prevalence indicates the number of people who know the word, which was obtained from an online study with over 220,000 participants (Brysbaert et al. 2019). The paired-sample *t* test showed that there was no significant difference between the target and competitor words in terms of word frequency ($t(29) = 0.019$, $p = 0.985$) and prevalence ($t(29) = -0.582$, $p = 0.565$).

Sixty-four filler trials were also included in the experiment, interspersed with the critical trials. The filler trials were used to counterbalance the stress status of the word fragments (i.e., stressed vs. unstressed) in the carrier sentences in the entire experiment. In the filler trials, 16 of the target words were stressed on the first syllable, and the remaining 48 target words were stressed on the second syllables. The first syllable of the target words in the filler trials were segmentally different from that of their competitor words (e.g., *HAbit* vs. *hoTEL*), except for the first segment in the first syllable.

One phonetically trained female native speaker of General American English was instructed to record the auditory stimuli. The recording was conducted using a Lenovo Ideapad Flex 4 laptop with a built-in microphone at a sampling rate of 44,100 kHz and a 16-bit amplitude resolution in a sound-attenuated environment. The speaker produced the same target words in the two carrier sentences, with each sentence being repeated five times. To keep the difference in the duration of the segmented syllables minimal, one recording of each carrier sentence was selected for the experiment. The first syllable of the target words was extracted from their original production at a segmentation point near the offset of the syllable that was sufficiently early to prevent coarticulatory spectral information from signaling the segmental content of the subsequent syllable. For example, the first syllable of *antonym* will be cut before the presence of spectral information of the next segment/t/. The extraction was carried out in Praat (Boersma and Weenink 2021) and checked by a trained phonetician. The extracted fragment of the target word was concatenated with its original carrier sentence in Praat. To determine the quality of the segmentation for the following experiment, two native English speakers and two Mandarin speakers who were not participants in the test were invited to verify the absence of coarticulatory information in the segmented syllables. They heard the segmented syllable (e.g., the first syllable *mer-* in the word *merchandise*) and then were given two-word choices (e.g., *merchandise* vs. *merciful*, one with the correct segmental continuation and one with the

incorrect segmental continuation, and both with the same stress) to see if they could predict the correct segmental continuation based on the segmented syllable they hear. If the cut-off point was sufficiently early, they should not have been able to predict the correct segmental continuation. Since three out of four participants achieved an accuracy score below 50%, which is the chance level, in this pilot study, the segmentation of the stimuli was deemed successful in excluding coarticulatory effects of nearby segments. The extracted syllables from the target words were spliced with their original carrier sentence (the same carrier sentence for all the stimuli) to which the target word belonged, which were performed using the *Concatenate* function in Praat. The duration, mean f0, and mean intensity were extracted from the first syllables of the target words using the ProsodyPro script in Praat (Version 5.7.8.7) (Xu 2013). To validate the stimuli for the perception experiment, simple linear regressions were performed to see whether the first syllable of the target words differed between the accented and the deaccented conditions in terms of duration, mean f0, and mean intensity. In each regression model, the fixed factor was the pitch accent context, and the response variable would be one of the acoustic measures (i.e., duration, mean f0, and mean intensity). The results indicated that there was only a significant main effect of the pitch accent context for mean f0 and mean intensity but not for duration (mean f0: Est. = −60.541, SE = 6.903, $t = -8.77$, $p < 0.001$; mean intensity: Est. = −9.240, SE = 0.761, $t = -12.14$, $p < 0.001$). To reduce the effect of intensity, the intensity of each stimulus was normalized to 70 dB based on the root-mean-square (RMS) amplitude using Chad Vicenik's Praat script[1].

2.3. Procedure

All participants completed a forced-choice word identification task on the experimental platform *FindingFive* (FindingFive Team 2019). Each participant was instructed to complete the experiment using their headphones in a quiet room. The participants heard auditory sentences ending with a word fragment, and two-word choices subsequently appeared on the computer screen. The participants were instructed to select the word they thought the auditory fragment they heard belonged to. To respond, they were required to put their index fingers on the keyboard and press the key corresponding to their response ("F" key = left word, "J" key = right word). The accuracy of each trial was recorded. A practice session of five target–competitor pairs from the filler trials with segmentally and suprasegmentally different first syllables (e.g., *seLECT* vs. *CANvass*) was provided to the participants before the beginning of the experiment with feedback on the accuracy of their responses. The participants would proceed to the experiment regardless of their practice accuracy. No feedback was provided during the experiment. There were two blocks in the experiment, and each block contained 62 trials that were randomized within the block. The participants were evenly assigned to either group A, with blocks in sequential order, or group B, starting with the second block and then the first block, in the experiment. The complete forced-choice word identification task took approximately 20 min.

2.4. Data Analysis

The dependent variable of Experiment 1 was the accuracy of the response. The within-subject independent variables were vowel reduction (non-vowel-reduction vs. vowel-reduction) and pitch accent (accented vs. deaccented). In the non-vowel-reduction condition, the vowel in the first syllable of the competitor word was full, while in the vowel-reduction condition, the vowel in the first syllable of the competitor word was reduced. For the accented condition, the extracted syllable came from the target word produced with the carrier sentence (a), while in the deaccented condition, the extracted syllable came from the target word produced with the carrier sentence (b). These two variables were the fixed factors in the logistic mixed-effects regression model.

For the random effect structure, we also included random intercepts by the participant and item in the models. The logistic mixed-effects regression model was performed using the glmer() function from the lme4 package in R (Bates et al. 2015; R Core Team 2022). The

"bobyqa" optimizer was used, and the maximum number of function evaluations for the optimizer (maxfun) was set to 10,000. The full model is provided as follows in the syntax of R:

```
glmer(Accuracy ~ Pitch Accent * Vowel Reduction + (1 | Participant) + (1 | Item), data, family = binomial("logit"),
      control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 10,000)))
```

3. Results

3.1. Native English Listeners

Figure 2 shows the native English listeners' identification accuracy rate, and Table 3 presents the summary of the coefficients of the fixed effects in the logistic mixed-effects regression model on the responses of the native English listeners.

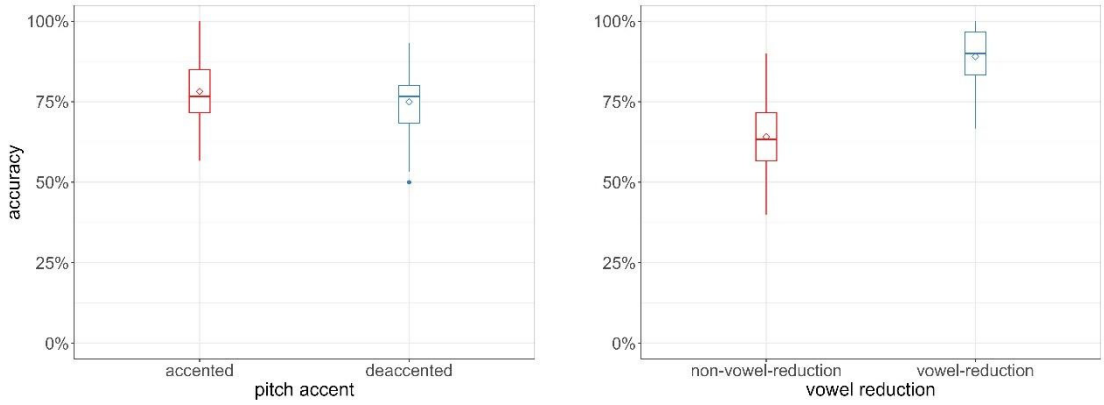

Figure 2. The accuracy rates by pitch accent (**left**) and the accuracy rates by vowel reduction (**right**) by native English listeners. The box spans from the first quartile to the third quartile. The line inside the box represents the median. The diamond dot inside the box represents the mean.

Table 3. Model parameters for the logistic mixed-effects model on accuracy of native English listeners.

Predictors	***β***	***SE***	***z***	***p***
(Intercept)	1.468	0.125	11.762	**<0.001**
Pitch accent (deaccented)	−0.241	0.170	−1.414	0.157
Vowel reduction (vowel-reduction)	1.661	0.174	9.553	**<0.001**
Pitch accent (deaccented): vowel red (vowel-reduction)	−0.082	0.340	−0.241	0.809
Random Effects				
σ^2		3.29		
$\tau_{00\ stimuli}$		0.21		
$\tau_{00\ participant}$		0.27		
ICC		0.13		
Marginal R^2/Conditional R^2		0.158/0.264		

Note: *p*-values < 0.05 are bolded and their corresponding variables are statistically significant.

The R-squared value of 0.264 suggested that the model explained 26.4% of the variance in the responses. As shown in Table 3, we found a significant main effect of vowel reduction (β = 1.661, *SE* = 0.174, z = 9.553, $p < 0.001$), suggesting that the native English listeners' identification accuracy rate was significantly higher when the initial vowel of the competitor word was reduced (M = 89.0% ± 9.3%) than when it was not reduced (M = 64.1% ± 12.3%). However, there was no significant main effect of pitch accent (β = −0.241, *SE* = 0.170, z = −1.414, p = 0.157). The interaction between pitch accent and vowel reduction was also not significant (β = −0.082, *SE* = 0.340, z = −0.241, p = 0.809).

3.2. Native Mandarin Chinese Listeners

Figure 3 shows the native Mandarin Chinese listeners' identification accuracy rate, and Table 4 presents the summary of the coefficients of the fixed effects in the logistic mixed-effects regression model on the responses of the native Mandarin Chinese listeners.

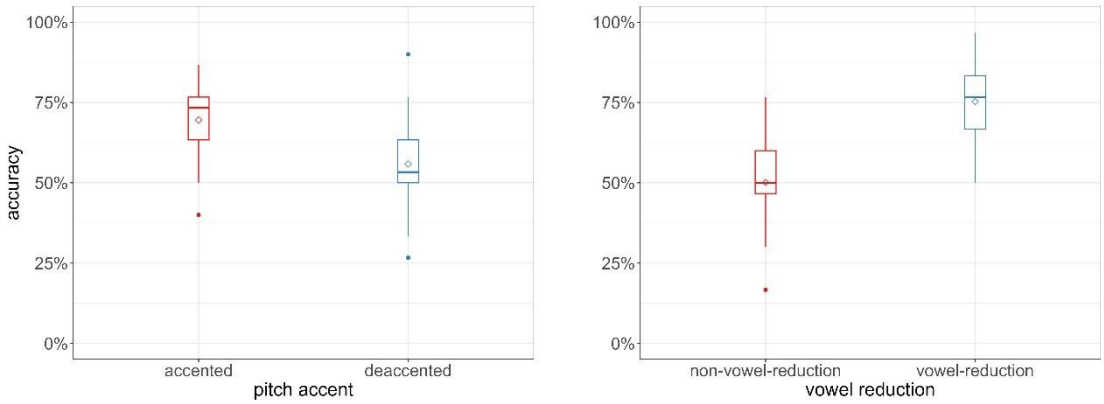

Figure 3. The accuracy rates by pitch accent (**left**) and the accuracy rates by vowel reduction (**right**) by native Mandarin Chinese listeners. The box spans from the first quartile to the third quartile. The line inside the box represents the median. The diamond dot inside the box represents the mean.

Table 4. Model parameters for the logistic mixed-effects model on accuracy of native Mandarin Chinese listeners.

Predictors	***β***	***SE***	***z***	***p***
(Intercept)	0.665	0.137	4.850	**<0.001**
Pitch accent (deaccented)	−0.728	0.230	−3.170	**0.002**
Vowel reduction (vowel-reduction)	1.339	0.231	5.801	**<0.001**
Pitch accent (deaccented): vowel reduction (vowel-reduction)	−0.404	0.459	−0.880	0.379
Random Effects				
σ^2		3.29		
$\tau_{00\ stimuli}$		0.59		
$\tau_{00\ participant}$		0.16		
ICC		0.19		
Marginal R^2/Conditional R^2		0.128/0.289		

Note: *p*-values < 0.05 are bolded and their corresponding variables are statistically significant.

The model's R-squared value was 0.289, indicating that 28.9% of the variance in the responses was explained by the model's fixed and random effects. As shown in Table 4, we found a significant main effect of vowel reduction (β = 1.339, SE = 0.231, z = 5.801, $p < 0.001$), suggesting that the non-native English listeners' identification accuracy rate was significantly higher when the initial vowel of the competitor word was reduced (M = 75.3% ± 12.0%) than when it was not reduced (M = 50.1% ± 13.6%). In addition, there was a significant main effect of pitch accent (β = −0.728, SE = 0.230, z = −3.170, $p < 0.05$), suggesting that the non-native English listeners' identification accuracy rate was significantly higher when the target word was accented (M = 69.5% ± 10.9%) than when the target word was deaccented (M = 55.9% ± 13.8%). The interaction between pitch accent and vowel reduction was not significant (β = −0.404, SE = 0.459, z = −0.880, p = 0.379).

4. Discussion

The results provided clear evidence that different cues were used differently by the native English and Mandarin Chinese listeners in the perception of English lexical stress. The native English listeners primarily relied on vowel quality as a cue for English lexical

stress, which is consistent with the results of Chrabaszcz et al. (2014). On the other hand, the native Mandarin Chinese listeners employed both vowel quality and f0 in an equivalent manner for the perception of English lexical stress. The higher accuracy of word identification response in the vowel-reduction condition than in the non-vowel-reduction condition suggested that both the native English and Mandarin Chinese listeners relied on the segmental difference in the first syllable of the two words to identify the target word. For the native Mandarin Chinese listeners only, their accuracy was higher in the accented condition than in the deaccented condition, suggesting a similar degree of reliance on f0 relative to vowel quality to identify lexical stress in spoken word recognition in both conditions.

Our findings with respect to native Mandarin Chinese listeners partially corroborate the conclusions of Zhang and Francis (2010) that vowel quality and f0 are both utilized by native Mandarin Chinese listeners to identify English lexical stress. Zhang and Francis (2010) found that vowel quality and f0 were used as a combinatorial cue to stress, whereas our study revealed that native Mandarin Chinese listeners employed these two cues independently, as evidenced by the lack of a significant interaction between pitch accent and vowel reduction. As opposed to native English listeners, native Mandarin Chinese listeners utilize f0 as an extra cue to identify English lexical stress. This could be attributed to the extensive use of f0 as a cue to tone in Mandarin. Moreover, native Mandarin Chinese listeners may rely on multiple cues depending on how accessible they are.

Native English listeners do not seem to be affected by pitch accent, which could be attributed to two possible causes. First, f0 is not a stable cue to lexical stress at the sentence level due to the fact that either a low or a high f0 may be associated with a stressed syllable. As Liu (2019) demonstrated, English disyllabic nouns with initial stress were poorly identified by native English listeners when produced with an L* pitch accent in a sentence ending with a rising intonation contour (L* H-H%). In this case, misalignment between pitch height and stressed syllable (stressed syllable aligned with a low pitch) has a strong and negative impact on spoken word recognition even among native English listeners. Thus, while f0 can function as a cue to lexical stress, its inconsistent association with stressed and unstressed syllables at the sentence level renders it a less reliable cue, in comparison to vowel quality and duration. Second, it might have been difficult for the listeners to perceive stress on the truncated syllables due to their short durations, even though they were extracted from the target word with stress on the first syllable. The brevity of the truncated syllable made it difficult for the target word to be distinguished from the competitor word. In other words, the listeners may not have been sensitive to the f0 change in the accented syllable induced by a sentence-level pitch accent due to its relatively short duration. A follow-up study could include target truncated syllables with enhanced f0 or lengthened duration.

Why did the participants select the less familiar word, as opposed to the more familiar word, as a strategy? Initially, we thought it may have a relationship with the duration of the truncated syllable. In a gating task, Tyler (1984) found that participants had a preference for high-frequency words only up to 150 ms, and afterwards, more low-frequency words were elicited. This provides converging evidence for our assumption, since most of the truncated syllables in our study had a duration of over 150 ms. We speculate that this might have something to do with how listeners allocate more perceptual effort during the processing of the first syllable of a less probable word. In a cross-linguistic study by King and Wedel (2020), it was found that segment composition within a word is optimized to provide listeners greater disambiguating information as they identify words in the speech stream. Specifically, less probable words tend to be composed of segments that are of higher informativity (informativity is defined as the average unpredictability) towards the beginning of the words. Perceptual studies have shown that listeners are more accurate at perceiving segments that are more informative (e.g., Bennett et al. 2018). These findings suggest that listeners are perceptually tuned to selectively attend to those phonetic dimensions that are informative (Davidson et al. 2007; Holt and Lotto 2006; McGuire 2007).

Together, these findings on how informative segments are distributed in the lexicon and how listeners are perceptually tuned to attend to high-information segments would suggest that listeners are accustomed to allocating more perceptual effort to the first syllable of a less probable word. Our listeners might have associated their perceptual effort toward the truncated word fragments with how prevalent they expected the target words to be. Given that our experiment required listeners to use only a word fragment to identify a word, we would expect listeners to generally allocate a high level of perceptual effort. The listeners could therefore develop a strategy to select the less probable word, which typically requires a high level of perceptual effort. To further account for the lexical processing mechanism behind this strategy, future studies are needed to explore the extent of word prevalence as a word processing strategy.

The findings of our study provide supporting evidence to a multisystemic model of L2 rhythm acquisition, in which various rhythm-related linguistic–systemic features are acquired under the L1 effect (Li and Post 2014). In our study, we found that both native English and Mandarin Chinese listeners can use vowel quality as a cue to English lexical stress, suggesting that native Chinese listeners have acquired certain prosodic features in their L2 English. However, compared to native English listeners, native Mandarin Chinese listeners also rely more on f0 in the perception of English lexical stress, suggesting a potential transfer effect of their L1 prosodic features. In this case, native Mandarin Chinese listeners' acquisition of L2 English lexical stress could fit into the multisystemic model of L2 rhythm acquisition such that they acquire both the prosodic features of English and the implementation of the prosodic differences between L1 Mandarin Chinese and L2 English. Broadly, if we treat L2 rhythm acquisition as a dynamic process (De Bot et al. 2007), we need to consider the interconnectedness between the subsystems to better account for L2 rhythm acquisition.

5. Conclusions

The findings of this study demonstrated that native English and Mandarin Chinese listeners vary in their use of vowel quality and f0 in the perception of English lexical stress. Native English speakers rely more heavily on vowel quality than f0 to identify lexical stress in spoken word recognition, whereas native Mandarin Chinese listeners rely equally on both vowel quality and f0. Compared to vowel quality, f0, the main acoustic correlate of pitch accent, is more susceptible to change in sentences with different intonational patterns, rendering it a less stable and thus less reliable cue for word-level stress, but its cue weight may be adjusted by native Mandarin Chinese listeners due to their first language's prosodic features.

Author Contributions: Conceptualization: F.W.; Data curation: F.W.; Formal Analysis: F.W., D.D., K.T., R.W.; Investigation: F.W., D.D., K.T., R.W.; Methodology: F.W., K.T., R.W.; Project administration: F.W.; Software: F.W.; Supervision: K.T., R.W.; Validation: F.W., D.D.; Visualization: F.W.; Writing—original draft: F.W.; Writing—review & editing: F.W., D.D., K.T., R.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board at the University of Florida (Protocol code: IRB202001699; Date of Approval: 16 September 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available upon request.

Acknowledgments: Thank you to Annie Tremblay for her brainstorming about the experimental design. Thank you to Meagan Durana for her proofreading.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Target and competitor words in experimental trials.

Non-Vowel-Reduction Condition		Vowel-Reduction Condition	
Target	**Competitor**	**Target**	**Competitor**
antonym	antenna	advocate	advisor
archive	arcade	column	cologne
campus	campaign	commerce	command
carnival	carnation	continent	container
distance	distaste	concept	concern
instrument	instructor	concert	conceit
intellect	intestine	confluence	confusion
introvert	intruder	promise	promotion
monster	monsoon	motive	material
musical	museum	congress	congratulate
mister	mistake	parrot	parade
particle	partition	polisher	policeman
booking	bouquet	programmer	procurer
ambulance	ambition	providence	provider
district	destroy	purchase	pursue

Table A2. Target and competitor words in filler trials.

Target	Competitor	Target	Competitor
motel	mystery	activity	absolute
Manhattan	monarch	antique	alternate
parole	Paris	biologist	blackmail
dismissal	diplomat	canteen	capture
suggest	summer	cartel	carefree
amnesia	altitude	diameter	deviate
inspection	illustrate	incise	iceberg
Mercedes	mediate	invasion	ignorance
trustee	truthful	machine	motion
receipt	relative	pursuit	protest
reward	rocket	review	rubric
gorilla	governor	routine	ribbon
consent	clumsy	deception	discount
reactor	relevant	percussion	prospect
direction	dialect	distinct	diamond
retention	registry	distress	delegate

Table A3. Target and competitor words in filler trials.

Target	Competitor	Target	Competitor
habit	hotel	secrete	season
cancel	collapse	bizarre	bargain
butter	balloon	between	biscuit
tunnel	tattoo	neglect	nocturne
captain	canal	polite	panel
atmosphere	approach	promote	principle
debut	dilute	conceal	canvass
pavement	patrol	freebie	forever
funeral	forget	ferment	fellow
passport	possess	guitar	garden
solider	suppose	lament	legend
option	obscure	reserve	rescue

Table A3. *Cont.*

Target	Competitor	Target	Competitor
college	cocoon	career	carbon
empire	emission	peculiar	porridge
formula	forbid	success	summon
palace	pecan	respond	ruthless

Appendix B

Table A4. Word frequency, word prevalence, number of letters, and number of syllables of words in experimental trials.

	Word Frequency *M* (*SD*)	Word Prevalence *M* (*SD*)	No. of Letters *M* (*SD*)	No. of Syllables *M* (*SD*)
non-vowel-reduction				
target	3.63 (0.97)	2.33 (0.18)	7.73 (1.16)	2.53 (0.52)
competitor	3.56 (0.73)	2.27 (0.22)	7.73 (1.16)	2.47 (0.52)
vowel-reduction				
target	3.61 (0.76)	2.29 (0.33)	7.87 (1.41)	2.40 (0.51)
competitor	3.70 (0.76)	2.27 (0.28)	7.93 (1.53)	2.73 (0.70)

Note

[1] Available at: http://phonetics.linguistics.ucla.edu/facilities/acoustic/IntensityScaler.txt (accessed on 16 September 2020).

References

Baayen, R. Harald, Richard Piepenbrock, and Leon Gulikers. 1996. *The CELEX Lexical Database (cd-rom)*. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]

Beckman, Mary E. 1986. Intonational structure in English and Japanese. *Phonology Yearbook* 3: 255–309. [CrossRef]

Beckman, Mary E., and Gayle Ayers. 1997. Guidelines for ToBI labelling. *The OSU Research Foundation* 3: 30.

Beckman, Mary E., and Julia Hirschberg. 1994. *The ToBI Annotation Conventions*. Columbus: Ohio State University.

Bennett, Ryan, Kevin Tang, and Juan Ajsivinac Sian. 2018. Statistical and acoustic effects on the perception of stop consonants in Kaqchikel (Mayan). *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 9: 9. [CrossRef]

Boersma, Paul, and David Weenink. 2021. Praat: Doing Phonetics by Computer. Available online: http://www.fon.hum.uva.nl/praat/ (accessed on 12 June 2020).

Bolinger, Dwight. 1989. *Intonation and Its Uses: Melody in Grammar and Discourse*. Redwood City: Stanford University Press.

Brysbaert, Marc, and Boris New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41: 977–90. [CrossRef]

Brysbaert, Marc, Pawel Mandera, and Emmanuel Keuleers. 2018. The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science* 27: 45–50. [CrossRef]

Brysbaert, Marc, Pawel Mandera, Samantha F. McCormick, and Emmanuel Keuleers. 2019. Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods* 51: 467–79. [CrossRef] [PubMed]

Campbell, Nick, and Mary Beckman. 1997. Stress, prominence, and spectral tilt. Paper presented at ESCA Tutorial and Research Workshop on Intonation: Theory, Models and Applications, Athens, Greece, September 18–20.

Chrabaszcz, Anna, Matthew Winn, Candise Y. Lin, and William J. Idsardi. 2014. Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech, Language, and Hearing Research* 57: 1468–79. [CrossRef]

Connell, Katrina, Simone Hüls, Maria Teresa Martínez-García, Zhen Qin, Seulgi Shin, Hanbo Yan, and Annie Tremblay. 2018. English Learners' Use of Segmental and Suprasegmental Cues to Stress in Lexical Access: An Eye-Tracking Study. *Language Learning* 68: 635–68. [CrossRef]

Connine, Cynthia M., Dawn G. Blasko, and Jian Wang. 1994. Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception & Psychophysics* 56: 624–36.

Connine, Cynthia. M., Debra Titone, Thomas Deelman, and Dawn Blasko. 1997. Similarity mapping in spoken word recognition. *Journal of Memory and Language* 37: 463–80. [CrossRef]

Cooper, Nicole, Anne Cutler, and Roger Wales. 2002. Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech* 45: 207–28. [CrossRef] [PubMed]
Cutler, Anne. 1986. Forbear is a Homophone: Lexical Prosody Does Not Constrain Lexical Access. *Language and Speech* 29: 201–20. [CrossRef]
Cutler, Anne, and Charles Clifton. 1984. The use of prosodic information in word recognition. In *Attention and Performance X: Control of Language Processes*. Edited by H. Bouma and D. G. Bouwhuis. London: Erlbaum, pp. 183–96.
Davidson, Lisa, Jason Shaw, and Tuuli Adams. 2007. The effect of word learning on the perception of non-native consonant sequences. *The Journal of the Acoustical Society of America* 122: 3697–709. [CrossRef] [PubMed]
De Bot, Kees, Wander Lowie, and Marjolijn Verspoor. 2007. A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition* 10: 7–21. [CrossRef]
Fear, Beverley D., Anne Cutler, and Sally Butterfield. 1995. The strong/weak syllable distinction in English. *The Journal of the Acoustical Society of America* 97: 1893–904. [CrossRef] [PubMed]
FindingFive Team. 2019. *FindingFive: A Web Platform for Creating, Running, and Managing Your Studies in One Place*. Cherry Hill: FindingFive Corporation (Nonprofit). Available online: https://www.findingfive.com (accessed on 8 September 2020).
Flege, James E., and Ratree Wayland. 2019. The role of input in native Spanish Late learners' production and perception of English phonetic segments. *Journal of Second Language Studies* 2: 1–44. [CrossRef]
Francis, Alexander L., and Howard C. Nusbaum. 2002. Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance* 28: 349. [CrossRef]
Francis, Alexander L., Valter Ciocca, Lian Ma, and Kimberly Fenn. 2008. Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics* 36: 268–94. [CrossRef]
Fry, Dennis Butler. 1955. Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America* 27: 765–68. [CrossRef]
Fry, Dennis Butler. 1958. Experiments in the perception of stress. *Language and Speech* 1: 126–52. [CrossRef]
Fry, Dennis Butler. 1965. The dependence of stress judgments on vowel formant structure. In *Phonetic Sciences*. Basel: Karger Publishers, pp. 306–11.
Goldinger, Stephen D., Paul A. Luce, and David B. Pisoni. 1989. Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language* 28: 501–18. [CrossRef]
Guion, Susan G., and Eric Pederson. 2007. Investigating the role of attention in phonetic learning. *Language Experience in Second Language Speech Learning* 17: 57–77.
Gussenhoven, Carlos. 1994. Focus and sentence accents in English. *Focus and Natural Language Processing* 3: 83–92.
Halle, Morris, and Jean-Roger Vergnaud. 1987. Stress and the cycle. *Linguistic Inquiry* 18: 45–84.
Hammond, Robert M. 1995. Foreign accent and phonetic interference: The application of linguistic research to the teaching of second language pronunciation. In *Second Language Acquisition Theory and Pedagogy*. London: Taylor & Francis Group, pp. 293–304.
Hockett, Charles Francis. 1955. *A Manual of Phonology*. No. 11. Baltimore: Waverly Press.
Holt, Lori L., and Andrew J. Lotto. 2006. Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America* 119: 3059–71. [CrossRef]
Iverson, Paul, Patricia K. Kuhl, Reiko Akahane-Yamada, Eugen Diesch, Yoh'ich Tohkura, Andreas Kettermann, and Claudia Siebert. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87: B47–B57. [CrossRef]
King, Adam, and Andrew Wedel. 2020. Greater early disambiguating information for less-probable words: The lexicon is shaped by incremental processing. *Open Mind* 4: 1–12. [CrossRef] [PubMed]
Ladd, D. Robert. 2008. *Intonational Phonology*. Cambridge: Cambridge University Press.
Li, Aike, and Brechtje Post. 2014. L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German learners of English. *Studies in Second Language Acquisition* 36: 223–55. [CrossRef]
Lieberman, Philip. 1965. On the acoustic basis of the perception of intonation by linguists. *Word* 21: 40–54. [CrossRef]
Liu, Yaobin. 2019. The influence of pitch contour on Mandarin speakers' perception of English stress. *University of Pennsylvania Working Papers in Linguistics* 25: 20.
Marslen-Wilson, William. 1990. Activation, competition, and frequency in lexical access. In *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. Edited by Gerry T. M. Altmann. Cambridge: The MIT Press, pp. 148–72.
Marslen-Wilson, William, and Paul Warren. 1994. Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review* 101: 653. [CrossRef] [PubMed]
Martinet, André. 1964. *Elements of General Linguistics*. Translated by Elisabeth Palmer. London: Faber and Faber. First published 1960.
McGuire, Grant L. 2007. Phonetic Category Learning. Doctoral dissertation, The Ohio State University, Columbus, OH, USA.
McQueen, James M., Dennis Norris, and Anne Cutler. 1994. Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20: 621. [CrossRef]
McQueen, James M., Dennis Norris, and Anne Cutler. 1999. Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance* 25: 1363. [CrossRef]
Qin, Zhen, Yu-Fu Chien, and Annie Tremblay. 2017. Processing of word-level stress by Mandarin-speaking second language learners of English. *Applied Psycholinguistics* 38: 541–70. [CrossRef]

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 22 October 2020).
Raphael, Lawrence J. 2021. Acoustic cues to the perception of segmental phonemes. In *The Handbook of Speech Perception*. Hoboken: Wiley-Blackwell, pp. 603–31.
Reinisch, Eva, and Andrea Weber. 2012. Adapting to suprasegmental lexical stress errors in foreign-accented speech. *The Journal of the Acoustical Society of America* 132: 1165–76. [CrossRef] [PubMed]
Selkirk, Elisabeth. 1995. Sentence prosody: Intonation, stress, and phrasing. In *The Handbook of Phonological Theory*. Hoboken: Wiley-Blackwell, vol. 1, pp. 550–69.
Sluijter, Agaath M. C., and Vincent J. Van Heuven. 1996a. Acoustic correlates of linguistic stress and accent in Dutch and American English. Paper presented at Fourth International Conference on Spoken Language Processing, ICSLP'96, Philadelphia, PA, USA, October 3–6; vol. 2, pp. 630–33.
Sluijter, Agaath M. C., and Vincent J. Van Heuven. 1996b. Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America* 100: 2471–85. [CrossRef]
Small, Larry H., Stephen D. Simon, and Jill S. Goldberg. 1988. Lexical stress and lexical access: Homographs versus nonhomographs. *Perception & Psychophysics* 44: 272–80.
Soto-Faraco, Salvador, Núria Sebastián-Gallés, and Anne Cutler. 2001. Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language* 45: 412–32. [CrossRef]
Surendran, Dinoj, and Gina-Anne Levow. 2004. The functional load of tone in Mandarin is as high as that of vowels. Paper presented at International Conference on Speech Prosody 2004, Nara, Japan, March 23–26.
Tyler, Lorraine K. 1984. The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics* 36: 417–27.
Van Donselaar, Wilma, Mariëtte Koster, and Anne Cutler. 2005. Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology Section A* 58: 251–73. [CrossRef] [PubMed]
van Heuven, Walter J. B., Pawel Mandera, Emmanuel Keuleers, and Marc Brysbaert. 2014. Subtlex-UK: A New and Improved Word Frequency Database for British English. *Quarterly Journal of Experimental Psychology* 67: 1176–90. [CrossRef] [PubMed]
Wang, Yue, Michelle M. Spence, Allard Jongman, and Joan A. Sereno. 1999. Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America* 106: 3649–58. [CrossRef] [PubMed]
Wayland, Ratree P., and Susan G. Guion. 2004. Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning* 54: 681–712. [CrossRef]
Xu, Yi. 2013. ProsodyPro—A Tool for Large-scale Systematic Prosody Analysis. Paper presented at Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France, August 30; pp. 7–10.
Zhang, Yanhong, and Alexander Francis. 2010. The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics* 38: 260–71. [CrossRef]