Vietnamese Native Speakers' Cues to the Perception of Stress

Stage II Qualifying Paper

Giang Ha Le

PI: Professor Chilin Shih, Professor Tania Ionin
Faculty Advisor: Professor Chilin Shih
Department of Linguistics
University of Illinois at Urbana - Champaign
Fall 2020

Abstract

L1 transfer affects the process of L2 acquisition in a significant way, both in perception and production, as learners have a tendency to apply phonological patterns of their native language to the target language. This study investigates the extent to which Vietnamese native speakers rely on F0 as a primary cue to perceive stress in English nonce words by manipulating the pitch contour around the stressed syllable by creating different environments where such pitch contours are realized, and subsequently measuring the differences in performance of stress location matching as a result. While the acoustic correlates of stress in English are F0, duration, intensity, and vowel quality (Fry, 1955; Libermann, 1960) [**?**], the acoustic correlates of tone in Vietnamese are F0, duration, and voice quality (Pham, 2000; Nguyen & Edmonton, 1997). Despite some overlapping of acoustic correlates, English lexical stress prediction cannot be predicated on pitch alone. For example, in a rising tonal contour context such as that of a yes/no question (L*H-H%), English stressed syllable actually receives a low pitch accent (Pierrehumbert, 1980).

The independent variables of this study are stress location in a nonce word, the number of syllables in the stimuli, and the type of intonation context where a statement context corresponds to a falling intonation pattern and a yes/no question context corresponds to a rising intonation pattern. The dependent variable of this study is the number or percentage of correct responses the participants give to a perceptual matching task. To avoid lexical retrieval and memorization effect, the nonce words were selected based on a search in a pronunciation corpus of American English. The nonce words have the same syllable shape as a real English word, follow English phonotactics, and are controlled for factors such as tendency of vowel reduction. Besides the nonce word items, a set of filler items was included in the test instrument, ranging from tokens that are minimal segmental contrast pairs to tokens that differ by syllable length. The experiment was repeated for a control group of L1 American English speakers. The tokens set was recorded by a native American English speaker, randomized during the actual experiment in blocks, and distributed to the participants in different test lists following the Latin square design. The participants listened to the stimuli with varying stress locations three times and then listened to the stimuli with either a statement or yes/no question intonation. They were asked to identify the sound that they heard previously which matches the sound they have just heard. Similar to (Ou, 2010)'s findings, the prediction for this study was that L1 Vietnamese L2 English speakers would show a significant difference in perceiving stress compared to the control group when the word has a yes/no intonation contour, because of Vietnamese speakers' tendency to rely on F0 as an acoustic cue for tone perception.

A mixed repeated measures ANOVA with a between subject factor was conducted over the collected data. It was found that in the bisyllabic words category, there is a statistically significant difference in stress matching accuracy between the control and the experimental group. Both sentence types and stress location have main effects on the stress matching accuracy, and there is an interaction between the L1 factor and sentence type, as well as between L1 and stress location. Followed-up independent samples t-tests with Bonferroni correction show that the source of the interaction is in the question condition and the word-initial stress condition across the two groups. This is fully in agreement with the prediction that we would see a difference in the stress matching accuracy between the L1 English speakers and L1 Vietnamese speakers in word-initial stress condition with question intonation. No significant difference was found in the stress matching accuracy of trisyllabic words. An analysis of the reaction time revealed that L1 Vietnamese speakers were more likely to change their answers for the perceptual task. The age of arrival factor was also analyzed and although there was a negative correlation between age of arrival and better performance at the stress matching task, this relationship was not statistically significant.
Keywords: *stress, prosody, intonation, second language acquisition, cross-language speech perception*

1 Introduction

In acquiring a second language, a learner experiences transfer effect from their first language in a variety of dimensions: phonology, morphosyntax, semantics, among others. Transfer effect is manifested in a difference in performance by the second language learners compared to the group of native speakers. For example, studies have found how discourse context used in the first language could affect the learners' preference towards using referential pronouns in the second language [**?**]. Specifically, (Li and Yang, 2016) found, using a back-translation task, that Chinese speakers prefer to omit pronouns in English because pro-drop is a common feature of Chinese and that Chinese speakers rely on context to infer topic continuity rather than maintain constant usage of pronouns [**?**]. Many other studies have also showed that L1 transfer effect is robust and evident in learners' performance.

This study seeks to investigate the extent to which L1 transfer effect influences how sensitive Vietnamese native speakers are to the position of lexical stress in American English. This study was motivated by the similarities between acoustic correlates between stress and tone in American English and Vietnamese, whereby the fundamental frequency (F0) and duration are acoustic cues that both L1 speakers of stress and tone languages use to identify their language's suprasegmental patterns. On the other hand, the way that L1 American English speakers and L1 Vietnamese speakers use F0 as cues for stress is predicted to be different, as L1 American English speakers are more familiar with the intonation contours of different sentence types in English than L1 Vietnamese L2 American English speakers. Notably, F0 of a word-final stressed syllable in a yes/no question context in English is lower than that of a word-final unstressed syllable. Because L1 Vietnamese speakers might equate higher F0 to lexical stress (Nguyen and Ingram, 2005), lower F0 on stressed syllable might make stress perception more difficult for the L1 tone language L2 stress language speakers than for the L1 American English control group in the yes/no question context compared to the statement context.

This paper reports findings of L1 Vietnamese L2 English speakers' stress perception performance using a forced choice perception task and a transcription task. The paper is divided into six main sections: (1) the Background and Literature Review section outlines concepts on which the study was based such as stress, tone, categorical perception, and summarizes results of a few related studies done on this topic, (2) the Research Question and Hypothesis section proposes two research questions of the study and their corresponding hypotheses, (3) the Methodology section describes the participants, procedure, and materials used in the study, (4) the Results section briefly presents the findings of the pilot study conducted prior to this full study and reported the statistics and inferential tests conducted on the collected data, (5) the Discussion section interprets the results given the research question and hypothesis, and (6) the Conclusion concludes the paper.

2 Background and Literature Review

2.1 Prosodic characteristics of American English and Vietnamese

2.1.1 American English as a stress-timed language

As this study rests upon understanding which acoustic correlates feature in stress perception, the obvious starting point is to define what stress is and which acoustic correlates are significant to detect stress. Stress is defined as an abstract relation of prominence, whereby prominence is realized by various acoustic cues and could belong to different levels (Kenstowicz, 1994). A word in English such as 'Alabama' has three levels of prominence, a primary stress on the third syllable, secondary stress on the first syllable, and the other two syllables are unstressed with reduced vowels. Articulatorily, a stressed syllable is produced with more effort than an unstressed syllable. Acoustically and perceptually, the correlates of stress in English include intensity or amplitude (loudness), length (duration), F0 (pitch), and vowel formants (vowel quality) such that stressed syllables are typically 'acoustically louder, longer and higher in pitch than other surrounding unstressed syllables' (Fry, 1955; Lehiste & Peterson, 1959; Libermann, 1960) (Spectral tilt or the amount of high-frequency energy in relation to the low-frequency energy is another acoustic correlate that was found to be associated with stress (Sluijter & Van Heuven, 1996), but as this factor has not been investigated in depth (Reetz and Jongman, 2009), the current census is for stress to have four main acoustic correlates). It has also been noted that, instead of an increase in F0, sometimes it is the change in any direction of F0 that is a correlate of stress. This point will become relevant and important as we discuss the intonation contour of stressed syllables in different sentence types in American English. It has also been reported that a difference of 5 Hz is sufficient to indicate

stress difference (Fry, 1955).

Among these four main acoustic correlates, intensity was found to be reliable acoustically but often considered a weak perceptual cue. Duration was found to be a reliable correlate and vowel quality was a rather poor acoustic cue (Fry, 1955; Sluijter & Van Heuven, 1996). Additionally, from a cross-linguistic perspective, different languages with stress may place different emphasis on acoustic correlates. For English, duration might be reliable and an important correlate of stress while loudness might be the most important correlate of stress in Russian.

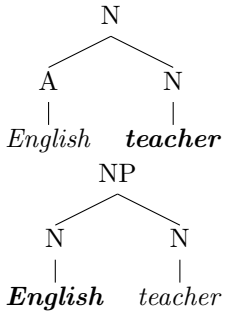
The below figure shows the phonetic measurements of F0, intensity, and duration of a nonce word *ferton*, uttered in isolation by an American English native speaker using two American English lexical stress patterns, word-initial and word-final. According to Table 1, it is not always the case that all three phonetic measurements of a stressed syllable are larger than those of an unstressed syllable, but nonetheless two phonetic measurements of the stressed syllable were larger in this case. For the word-initial stress case, F0 and intensity measurements of the first syllable were larger, and for the word-final stress case, the intensity and duration measurements were larger. It is clear from this instance that acoustic correlates of stress do not surface at the same degree or all at the same time.

	FERton		ferTON	
Syllable	FER	ton	fer	TON
Average F0 (Hz)	193.04	142.19	168.74	152.21
Intensity (dB)	74.98	64.93	65.47	70.18
Duration (ms)	296	330	213	438

Table 1: F0, duration, and intensity measurements of stressed and unstressed syllables in a standalone word context

Regarding stress patterns in the lexicon, English is mostly paradigmatic and sometimes lexical. The paradigmatic character of English distribution of stress can be seen in the way stress can be determined by morphological shape or part of speech, for example, the syllable before *-tion* is commonly stressed and many words' part of speech can only be distinguished by stress, as is in the case of *REcord* (noun) and *reCORD* (verb). Stress location is also dependent on the phonological shape of a word, for example, heavy syllables tend to attract stress more than light syllables do. Because stress is perceived as a relative property, in that it is perceived relative to other prominence level in other syllables, stressed syllables are often seen to be alternating in long English words in characteristic strong-weak rhythms. Furthermore, words of different etymologies or coming from different strata follow different stress rules, making lexical stress in English seem less regular compared to other languages with fixed stress patterns (Polish (stress on the penultimate syllable)) or Finnish (stress on the initial syllable)).

Beyond the words, stress position may be shifted if a stressed syllable were to be adjacent to another stressed syllable in an utterance, in a phenomenon termed *stress shift*. Moreover, stress and intonation in English play an important role in the constituent structures of phrases and compounds, speaker's attitude, and information structure of discourse. For example, the phrase *English teacher* receives stress on 'English' to signal that the phrase is a noun-noun compound meaning 'teacher of English' (compound stress rule) and the same phrase receives stress on 'teacher' to signal that the phrase is a modifier-head adjective-noun phrase meaning 'teacher who is English' (nuclear stress rule) (Liberman, 1975). In this case, stress alone determines the grammatical structure of the phrase, as illustrated below.



Interestingly, 'American history teacher' has two different interpretations but the same primary stress assignment (Liberman and Prince, 1977). [American [history teacher]] (history teacher who is American) has the stress pattern 2 - 1 - 3 while [[American history] teacher] (teacher of American history) has the stress pattern 3 - 1 - 2. Because the two parsings are almost the same in stress pattern, distinguishing

meanings between these two parsings requires a pause along the phrase boundary. Here, it seems that when stress alone is not enough of a cue for meaning distinction, other prosodic features need to be employed to facilitate communication. Back to 'an English teacher', it was found that the same stress pattern using the nuclear stress rule (with the 'teacher who is English' meaning) spoken in different intonations furthermore gives rise to different interpretations. (Liberman, 1975) identified four possible intonation contours for the phrase *an English teacher*, namely, a declarative intonation, a yes/no question intonation, an incredulity intonation, and finally an assertion intonation expressing obviousness. The intonation contours are represented as changes in the F0 contour throughout the utterance. These intonations reflect attitudes of the speaker and can be used to convey pragmatic intents. Contrastive stress also plays a crucial role in highlighting new information, for example, the utterance 'can I borrow ***that*** book' with a nuclear stress on 'that' implies the speaker wants to borrow 'that' book, as opposed to 'this' book.

Various phonological theories have been proposed to account for stress and intonation in American English. *The Sound Pattern of English* (Chomsky and Halle, 1968) described the rules of phrasal prominence and treated stress similarly to other phonemes, using binary features for its representation. However, as stress exhibits long-distance effects on the structure of phrases, such that a change in stress position within an intonation phrase could render a different interpretation of the phrase, it was clear that stress should be modeled differently from other segmental phonemes (Kenstowicz, 1994). Studying stress in whole utterances along with intonation and phrasing has enlightened the interaction between lexical, sentential stress and other prosodic features. (Liberman, 1975) started his proposal of a metrical stress theory by examining the stress pattern and tune or intonation of the vocative chants, outlining tune-text association principles that would link different tonal elements (high, high-mid, low-mid, and low) in a tune with the metrical structure of stress (strong and weak). Metrical stress theory (Liberman, 1975; Liberman and Prince, 1977) models stress with grid levels and allows stress to be independent from the phonemic strings. Constituents such as the foot and the word are represented as tiers in the grid, and stress is marked metrically for prominence. Below is the metrical grid for 'Apalachicoda' and its trochaic feet.

2					*	
1	*		*		*	
0	*	*	*	*	*	*
	A	pa	la	chi	co	la

Table 2: Metrical grid for *Apalachicola*

(Pierrehumbert, 1980) built upon suprasegmental phonology (Leben, 1973), autosegmental phonology (Goldsmith, 1976), and metrical theory (Liberman, 1975; Liberman and Prince, 1977) to propose a finite state grammar to model intonation in English. Her theory simplified the tonal elements in an intonation contour to two binary values, H and L. The pitch accent aligns with the most prominent syllable in a phrase and is therefore stressed. The pitch accent can be H* or L*, suggesting that a stressed syllable is not always pronounced with a higher pitch. Phrasal tones and boundary tones are tones at an intermediate phrase boundary and intonational phrase (IP) boundary, respectively. There are six types of pitch accents (H*, L*, H+L*, H*+L, L+H*, L*+H), two types of phrasal tones (H-, L-), and two types of boundary tones (L%, H%) in American English. One IP can have more than one intermediate phrase, therefore it can have more than one phrasal tone or boundary tone. If an IP has more than one pitch accent, the last pitch accent is termed the nuclear pitch accent. (Pierrehumbert, 1980)'s theory uses a finite state grammar to specify possible intonation contours in American English and the theory can be readily applied to computational models that use the finite state machine. The major intonation contours accounted for by the grammar are the statement intonation (H* L- L%), question intonation (H* L- H%), calling intonation (H*+L- H- L%), incredulous intonation (L*+H- L- H%), and asking for confirmation intonation (L* H- H%). Pierrehumbert's model gave rise to the TOBI prosody transcription convention (Silverman et al., 1992), which makes up of symbols for the pitch events such as pitch accents, phrase accents, boundary tones and two tiers, the tone and break index tiers.

2.1.2 Vietnamese as a tone language

While stress is defined in terms of relative prominence, a language with tone is defined as 'one in which an indication of pitch enters into the lexical realization of at least some morphemes' (Yip, 2002)

[?]. Acoustic correlates of tones are F0 indicating pitch movement and pitch height, length (duration), amplitude (intensity), and voice quality (Nguyen & Edmondson, 1997; Pham, 2000; Vu Thanh Phuong, 1981). Other tone correlates include pitch range, and beginning and ending points of pitch movement. Vietnamese is classified as a tone language where a syllable could carry different pitch patterns contrastive in meanings. In the standard Northern variety, a syllable could theoretically bear six or eight (Kirby, 2011) tones. The additional tones are checked tones that only occur in closed syllables ending in voiceless stops. Traditional analyses consider six tones to be phonemic in Northern Vietnamese. The following figure (Figure 1) shows the pitch patterns of six Northern Vietnamese tones, as spoken by a female native speaker on the syllable *ma*. The tones, from left to right, are mid level (ML), mid rising (MR), low falling with breathiness (LF), mid falling rising (MFR), mid rising with creakiness (MRC), and mid falling with creakiness (MFC). Voice quality changes such as breathiness and creakiness are distinctive features of Vietnamese tones.

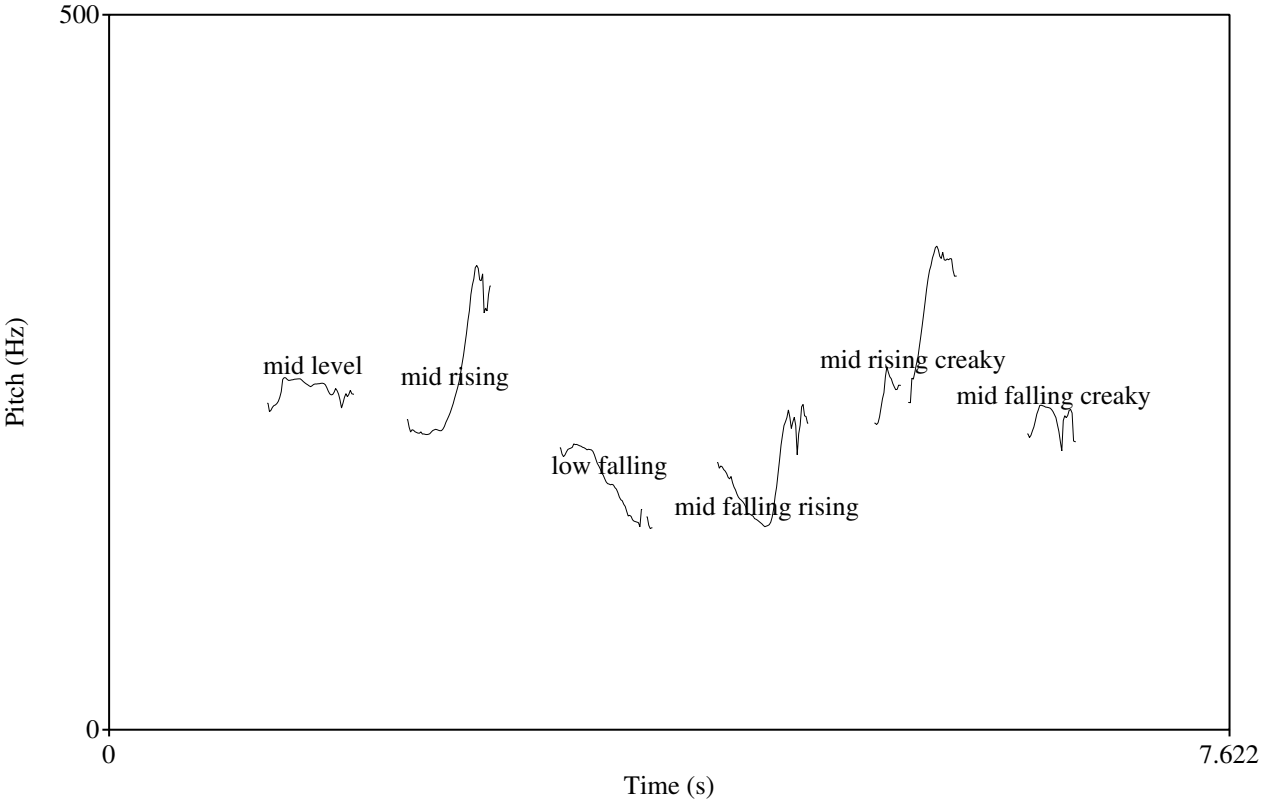


Figure 1: Pitch contours of six Northern Vietnamese tones

In the Central and Southern Vietnamese varieties, the MFR and MRC tones have reportedly been merged into one tone (Emeneau, 1951) [?], which could be a natural sequence of these tones being similar in contour coupled with the fact that syllables with the MRC tone occur much less often in the language to begin with. Furthermore, the distribution of tones is constrained by the syllable shape for closed syllables: in closed syllables with voiceless stop codas, only the MR and MFC tones are licensed. The following table shows this constraint applied on closed syllables with voiceless stop codas, compared to open syllables and closed syllables with nasal codas. Because only nasals, voiceless stops, and approximants can appear as codas, Table 3 presents all possible consonantal coda classes and their associated tone distributions.

	Open syllable	Closed syllable with nasal coda	Closed syllable with voiceless stop coda
ML	✓ma	✓man	X
MR	✓má	✓mán	✓mát
LF	✓mà	✓màn	X
MFR	✓ma?	✓ma?n	X
MRC	✓mã	✓mãn	X
MFC	✓ma.	✓ma.n	✓ma.t

Table 3: Tone distribution in Vietnamese

Having surveyed briefly the characteristics of stress in American English and tone in Vietnamese, let us turn to the next key concept that also serves as one of the central pieces of this paper's theoretical foundations: speech perception.

2.2 Speech Perception

Speech perception is the process of mapping of sounds into linguistic representations in the brain (Zsiga, 2013). This process begins with *hearing* where the speech signals are received by the ear and various information such as amplitude, frequency, timing, and so on of the signals is transmitted down the auditory pathway. Speech perception is thought to occur in the auditory cortex, where extracted auditory cues such as the F0, formant transitions, amplitude variations, and so on are matched to a stored linguistic unit [**?**]. The basilar membrane in the Cochea can perform Fourier-like analysis of the speech signals, specifically, the thin end of this membrane responds to high frequency components in the speech signal and the thick end of the membrane responds to low frequency components in the speech signal. Perception of sounds produced by different individuals is made possible by normalization of raw values so that perception is tuned to relative values, rather than absolute ones. This mechanism would enable us to map sounds produced by male and female speakers as belonging to the same category, despite the discrepancy in absolute terms. A notable feature of speech perception in humans is termed *categorical perception*, which means that humans tend to be good at distinguishing sounds that fall into different phonemic categories in their language as these sounds are contrastive and poor at distinguishing sounds that fall into the same category, even though acoustically these sounds might have different representations. (Liberman, 1957) tested participants' responses to synthesized stimuli of various voiced stops followed by a near-open front unrounded vowel and found that discrimination of the voiced stops along the place of articulation was robust, but discrimination of sounds within the same category was not. *Categorical perception* was also confirmed cross-linguistically (Abramson and Lisker,1970; Abramson and Lisker, 1973) by studies that looked at different languages' voice onset time (VOT) contrasts and the correlation with perception of the same sounds. The researchers found that the category boundary of VOT differs from language to language. For example, speakers of English have category boundary of 30 ms VOT for 'ba' and 'pa', thus tokens with VOT of more than 30 ms were perceived to be 'pa' and tokens with VOT of less than 30 ms were perceived to be 'ba'. For speakers of Spanish whose /b/ sound has a negative VOT, which indicates pre-voicing, and /p/ sound is voiceless unaspirated, the category boundary for 'ba' and 'pa' was found to be much lower, at 15 ms. For Thai speakers, categorical perception of voicing can be divided into three regions, /b/ would be perceived for VOT less than -20 ms, as Thai has pre-voicing /b/, /p/ would be perceived for VOT between -20 and +40 ms, and aspirated /p/ for VOT above +40 ms [**?**] [**?**]. Another well-known evidence supporting the *categorical perception* notion is Japanese speakers' insensitivity to the contrast between liquid sounds /r/ and /l/, as these sounds are allophones of the same phoneme in Japanese (Goto, 1971; Miyawaki et al., 1975) [**?**].

Speech perceptions studies often try to shed light onto the puzzle that humans are able to map continuous speech stream into discrete segments such as the phonemes. There seems to be a disconnection between the acoustics of speech and how the same information is represented abstractly in phonology. For instance, (Kenstowicz, 1994) remarked that it is difficult to segment waveforms into separate phonemes while discarding information from adjacent phonemes, due to coarticulation. Despite overlapping information in the acoustic signals, humans are able to perceive segments in discrete categories, as some studies have shown through evidence of categorical perception. To explain this curious dilemma, the *Motor Theory* (Liberman et al., 1967) posits that perception is more closely related to articulatory gestures than to acoustics. The argument for this arises from the fact that the same /d/ phoneme could

be perceived out of completely different acoustic signals, specifically, /d/ could be perceived from two different F2 formant transition patterns in /di/ and /du/, where the transition goes down in the former and up in the latter.

In the domain of prosody, native speakers of tone languages were found to have categorical perception of tones whereas naive listeners do not (Shen and Froud, 2016). Previous studies have also shown that native speakers of a stress language may have a smaller perceptual space of tones than that of native speakers of a tone language as they took longer to judge if two tones were different, suggesting that on average the tones were less distinct for the former group (Huang, 2001) [**?**]. Notably, several studies on stress perception by speakers of different L1 stress languages showed that speakers of L1 fixed-stress languages such as Arabic, Turkish, and French performed worse than speakers of unpredictable stress languages such as Spanish and speakers of L1 without stress such as Japanese, Korean, Chinese in stress perception tasks in English. Speakers of L1 with fixed stress were thought to lack a metalinguistic representation of contrastive stress in their L1 phonological representation and therefore were impaired in stress perception tasks, which is congruent with the *categorical perception* notion (Dupoux, Sebastian-Galles, Navarrete, and Peperkamp, 2008) [**?**]. Chinese speakers showed greater discrimination peakedness compared to English speakers in intonation perception, explained by the interaction of F0 in lexical tones and pitch contours of intonation in Chinese that is absent in English (Liu and Rodriguez, 2012). Interaction of lexical tone on prosody was shown via another study which found that Chinese speakers have a harder time identifying a rising intonation associated with questions when the final word in the sentence has a rising tone, compared to a final falling tone (Yuan, 2011). Similarly, (Ma et al., 2011) found that perception of intonation in questions and statements of Cantonese depends on the F0 of the final syllable. Besides F0, amplitude of the final word was also found to be a secondary cue to question or statement intonation perception (Morrow, 2013).

From these previous studies and conclusions, it could be expected that perception of stress by native speakers of a tone language would be significantly influenced by the L1's prosodic characteristics. The next section discusses the prosodic characteristics of stress and tone languages in order to provide more background towards forming the research questions.

2.2.1 Towards forming the research questions

What exactly are prosodic characteristics' differences between a stress versus a tone language? It turns out that the demarcation between a stress and a tone language is not as clear as what one would expect. English, though classified as a stress-timed language, does employ lexical pitch contrast akin to a tone language's technique in a limited number of cases, such as in the case of 'uh-huh', which can mean 'yes' or 'no', depending on the pitch patterns (Yip, 2002) [**?**].

Evaluating from the acoustic correlates of stress and tone described above suggests that acoustic correlates for stress and tones overlap in pitch and duration, thus it might be the case that learners of English whose native language is a tone language would rely on F0 and syllable length as primary cues to identify stress. On the other hand, studies have argued that English speakers have a tendency to perceive lexical stress using segmental cues, such as vowel reduction, rather than suprasegmental. Therefore, even though pitch might be the most important correlate of stress in English, according to the literature reviewed in **??**, and also used by Vietnamese speakers as cue in perception, the way pitch is correlated to stress position and understood by these two groups is not the same. For example, Ladefoged has demonstrated in his classic textbook for Phonetics *A Course in Phonetics* the different intonation patterns of the same word 'Amelia', where the second stressed syllable receives a high pitch in a statement sentence and a low pitch in a yes/no question, because the intonation for a statement is H* L L% and the intonation for a yes/no question is L* H H% (Ladefoged, 1993) [**?**]. This phenomenon was measured empirically in Ou (2010) and replicated here in Table 4 and Table 5 [**?**].

	FERcept		ferCEPT	
Syllable	FER	cept	fer	CEPT
F0 average (Hz)	281	141	147	233
Intensity (dB)	79	69	70	74
Duration (ms)	118	98	55	133

Table 4: Phonetic measures of stressed and unstressed syllables of nonce words in the falling intonation

	FERcept		ferCEPT	
Syllable	FER	cept	fer	CEPT
F0 average (Hz)	151	256	153	177
Intensity (dB)	65	73	63	69
Duration (ms)	116	119	63	69

Table 5: Phonetic measures of stressed and unstressed syllables of nonce words in the rising intonation

Ou (2010) explained that in the falling intonation of a statement context, the phonetic measures pattern as expected, in other words, all three phonetic measures of the stressed syllable were higher than those of the unstressed syllable. In the rising intonation of a yes/no question context, however, this pattern no longer holds. In fact, all three measures of the stressed syllable in the word-initial position were lower than those of the unstressed syllable. Ou (2010) explained that in the rising intonation condition, 'when the second syllable is stressed, it has a low rising pitch contour, when the second syllable is unstressed, it has a high rising pitch contour' [**?**]. That is because the word with initial stress would have a rising contour that starts earlier than the word with final stress. Ou (2010) manipulated the context where the stressed syllable occurs, and predicted that speakers of a tone language would have difficulty in perceiving the stress location in a low rising tonal contour context such as that of a yes/no question (L* H H%), as English stressed syllable actually receives a low tone then [**?**]. This study motivated the first research question and the pilot study reported in this paper.

Nguyen and Ingram (2005) [**?**] confirmed that Vietnamese speakers were able to differentiate stress contrasts in English as well as the native speaker group did in production, because Vietnamese speakers are used to using pitch and intensity as cue in tonal production. This study lent support for the hypothesis that speakers of languages whose acoustic cues match those of the target language would employ those cues in perception and/or production of the L1. Factors such as duration and vowel reduction, non-native elements of prosody perception, were not readily employed when the learners tried to identify stress in English, nor were they successfully replicated in production (Nguyen and Ingram, 2005) [**?**]. From the perception side, Nguyen (2017) specifically looked at tonal assignment as a result of stress perception. The study concluded that an English syllable could be perceived to carry a certain Vietnamese tone, depending on the syllable structure and relative F0 levels. This study's method centered around asking participants to transcribe target syllables that they have heard into the closest equivalent in Vietnamese using Vietnamese orthography. This method would inevitably be able to elicit the participants' mapping of English word stress with Vietnamese tones because Vietnamese orthography mandates tone assignment for each written syllable. The target syllables are bisyllabic real words such as *present - verb* and *present - noun*, and compounds such as *silver fish - fish made of silver* and *silver fish - not a golden fish* and *silverfish - a type of insect*. The result of this study showed that Vietnamese native speakers used relative F0 as cue for stress to tone mapping (Nguyen, 2017) [**?**]. Interestingly, the prediction that English syllables ending in obstruents will be perceived as having the rising tone (MR) if stressed and having the drop tone (MFC) if unstressed was only partially born out. Syllables ending in obstruents that are unstressed were still perceived to carry the rising tone (MR) by many speakers. The study also analyzed different responses of speakers due to dialectal differences with respect to tone assignment of stressed syllables ending in a sonorant, where Southern speakers preferred to assign the MR tone and speakers of the Central and Northern dialect preferred to assign the ML tone. In a similar vein, the topic of tonal assignment to stress in foreign words borrowed into Vietnamese could be inferred from a report on phonological adaptation of French loanwords in Vietnamese by Scholvin and Meinschaefer (2018) [**?**], whose data seemed to show that final syllables in a bisyllabic loanword are often assigned a level or rising tone, rather than a low tone, which seems to be congruent with the fact that French words are stressed on the last syllables and Vietnamese level and rising tones have a tendency to be associated with stressed syllables.

Not many studies have been done to examine how Vietnamese learners acquire English word stress, especially from the perception side. (Nguyen, 2005) reported findings on the transfer of tonal correlates in Vietnamese speakers' production of English words and found that while Vietnamese learners' production of English word stress mimicked F0 and intensity contrasts between stressed and unstressed syllables, the production of beginners fell short of replicating vowel duration and quality contrasts accurately, namely, the beginning learners had issues with reducing vowels to schwas in unstressed syllables.

3 Research Questions and Hypotheses

Based on the prior literature, particularly motivated by the results found in Ou (2010) and Nguyen (2017), this study proposes the following three research questions.

1. Research question 1: Do L1 Vietnamese L2 English speakers have more difficulty than American English native speakers in identifying stress location in yes/no question context where the stressed syllable does not necessarily have higher F0 than the other syllable(s)?

2. Research question 2: Would difficulty in identifying the stress location in the yes/no question context manifest in a delayed reaction to the stimuli in the L2 English group?

3. Research question 3: Does early arrival bestow an advantage on L2 speakers? Would L2 speakers who moved to the United States at a young age perform better at the stress matching task?

Given previous studies such as Ou (2010) and Nguyen (2017), the predictions to the aforementioned research questions are as follows.

1. Prediction 1: Vietnamese native speakers would have difficulty perceiving lexical stress location in contexts of a yes/no question when the stress is not word-final because of L1's prominence of pitch as a cue for tone perception and of pitch's unreliability as indicator of stress location in this case for the L1 Vietnamese speakers.

2. Prediction 2: Yes, difficulty in identifying the stress location in the yes/no question context would manifest in a delayed reaction to the stimuli and slower response from the L2 English participants.

3. Prediction 3: Early arrival might create an advantage for L2 learners and early arrivers would perform better than later arrivers at the stress matching task.

4 Methodology

4.1 Participants

The target participants for this study are adult L1 Vietnamese L2 English speakers with normal hearing ability. The control group is made up of adult native American English speakers. The recruitment process was done both via Amazon Mechanical Turk and offline and all study procedures were conducted with approval of the Institutional Review Board at the University of Illinois at Urbana Champaign. Forty-six participants participated in the study, representing two groups: (1) native American English speakers (n=26, age range = 18-69, average age = 37.88, median age = 34.5) and (2) native Vietnamese speakers who currently reside in the United States (n = 20, age range = 19-39, average age = 30.45). All participants reported no hearing difficulties. Each participant signed an informed consent form prior to taking part in the study and all received a compensation of 10 dollars for the time spent on the survey. The participants submitted responses to the survey at a distance from the researcher [1]. In order to verify that the participants are indeed native speakers of the respective languages, cloze tests in English and Vietnamese were included as a screening procedure. Performance on the fillers of the main test was also used as part of criteria to screen Amazon Turk workers. If the percentage score of the cloze test and fillers was less than 80 percent, the worker's response was discarded.The English cloze test consists of 50 fill in the blank questions and the Vietnamese cloze test consists of 14 fill in the blank questions. American English native speakers spent on average 16 minutes on the English cloze test while the Vietnamese speakers spent on average 25 minutes on the English cloze test. All participants in the experimental group scored at least 90 percent on the Vietnamese cloze test, in other words, no one got more than one wrong for this screening test. Results of the English cloze test from the two groups will be discussed in the Discussion section.

Both groups answered a questionnaire about their language background, hearing ability, and demographics before the study began.

All Vietnamese speakers were born in Vietnam (in Hai Phong (n = 1), Ho Chi Minh City (n = 10), Ha Noi (n = 8), Dong Nai (n = 1), or Hue (n = 1)). Nine participants reported to be native speakers of the northern dialect, 10 participants reported to be native speakers of the southern dialect, and 1 participant reported to be native speaker of the central dialect. On average, Vietnamese speakers began

[1] because of restrictions on in-person research due to the ongoing coronavirus pandemic

learning English from 8.1 years old (median = 8), with one starting the earliest from 4 years old and one starting the latest from 15 years old. The length of formal instruction as well as experience in using English for the Vietnamese native speakers is about 22 years on average, with a range between 14 to 32 years. On average, they moved to the United States at 20.75 years of age (median = 18), the youngest at 11 and the oldest at 37. The shortest duration of residence in the United States is 2.08 years and the longest duration of residence is 19.08 years. On average, the experimental group has been living in the United States for 9.73 years (median = 11). They reported knowledge of other languages other than English, including Spanish, French, Mandarin Chinese (n = 6), French (n = 5), Spanish (n = 1), Japanese (n = 1), Portuguese (n = 1), and Korean (n = 1).

The American English native speakers were all born in the United States (California (n = 8), Illinois (n = 3), Ohio (n = 2), Florida (n = 2), Georgia (n = 1), Michigan (n = 3), North Carolina (n = 2), New York (n = 2), New Hampshire (n = 1), Texas (n = 1), Washington State (n = 1). They reported knowledge of other languages other than English, including Mandarin Chinese (n = 3), Spanish (n = 11), French (n = 2), Japanese (n = 3), Bisaya (n = 1), Cantonese (n = 2). 11 participants reported to be monolinguals.

4.2 Materials

To avoid lexical retrieval and memorization effect, nonce words were created as stimuli for this experiment. In cases where it was difficult for a brand new nonce word to be created, a real word was used if it was deemed to be rare or is a proper name or a loanword. In most cases, the stimuli ended up to be proper names. In order to choose the nonce words, an interactive search in the CMU Pronouncing Dictionary for US English corpus provided in the NLTK toolkit was performed. The target stimuli included bisyllabic and trisyllabic words, thus the first set of queried results consists of bisyllabic words and the second set of queried results consists of trisyllabic words. The two sets consist of words with ambiguous stress patterns, in other words, the first set returns bisyllabic words where both word-initial and word-final stress are possible (Appendix A.1) and the second set returns trisyllabic words where word-initial, word-medial, and word-final stress are possible (Appendix A.2). For the majority of cases, the stimuli were directly taken from the returned set of words from the query, with some modification to ensure that vowel reduction is unlikely during the recording of these words with different stress patterns. Words that are also proper names were also used directly without modification, for example *taebak*. Because the search for trisyllabic words with ambiguous stress positions returned only three results, trisyllabic words that may be proper names were used and the different stress positions applied on them. All of the stimuli were double checked by a native speaker to make sure that no real common words appeared among the stimuli, and that the different stress positions within each stimulus are possible in English.

Test sentences were created to satisfy 4 test conditions for bisyllabic words, due to the crossing of 2 factors: the stress position (2 levels: word-initial and word-final) with the intonation context type (2 levels: statement sentence and yes/no question sentence), as shown in Table 6 and 6 test conditions for trisyllabic words, due to the crossing of 2 factors: the stress position (3 levels: word-initial, word-medial, and word-final) with the intonation context type (2 levels: statement sentence and yes/no question sentence), as shown in Table 7. The sentences were used as an aid for the recording procedure, in order to elicit a natural statement and question intonation from the English speaker. The sentences are similar in structure and the stimuli appear in all test sentences as nouns or proper names, in order to control for part of speech and bias related to part of speech due to English paradigmatic characteristics of stress. Based on the hypothesis posited in Section Research Questions and Hypotheses, L1 Vietnamese L2 English speakers are predicted to have more difficulty and lower performance than the control group in identifying stress location of the conditions marked with the double asterisks. The tables below show the stimuli created for the recording step. These stimuli were later extracted from the originally recorded sentences so that only the words remain for the participants to evaluate.

	Stress initial	Stress final
Word in a statement sentence	He is a GAUbert.	He is a gauBERT.
Word in a yes/no question sentence	**Is he a GAUbert?	Is he a gauBERT?

Table 6: 2 X 2 design for bisyllabic words. A tokens set example

	Stress initial	Stress middle	Stress final
Word in a statement sentence	That is a CAsey-beer.	That is a caSEY-beer.	That is a casey-BEER.
Word in a yes/no question sentence	**Is that a CAsey-beer?	**Is that a caSEY-beer?	Is that a casey-BEER?

Table 7: 2 X 3 design for trisyllabic words. A tokens set example

Seventeen bisyllabic words were chosen for recording, resulting in 102 sound files (17 x 6, 2 stimuli uttered in isolation in two stress patterns, 2x2 additional stimuli based on the crossed conditions) and 19 trisyllabic words were chosen for recording, resulting in 171 sound files (19 x 9, 3 stimuli uttered in isolation in three stress patterns, 3x2 additional stimuli based on the crossed conditions) (see Appendix A.3). 19 word pairs were created as fillers. The fillers were mainly word pairs that differ in other dimensions such as segmental contrasts, vowel length, word length, and so on. A mix of real and nonce words could be found in the fillers set. A sentence containing one of the words in the pair was created for recording. The fillers were also used to screen workers in Amazon Mechanical Turk. Table 8 below shows a few examples of fillers and the word pairs' differences. 16 real words with stress location differences were recorded for a transcription task. [2]

Filler	Type of difference
close /z/ and close /s/	voicing
repair and repaired	presence of the past morpheme
tirade and tirades	presence of the plural morpheme
fim and feam	vowel length
wander and wonder	vowel quality

Table 8: Filler examples

The recording was done by a graduate student of Linguistics who is an American English native speaker, using Praat at sampling frequency 44,100 Hz. Reviewing of the recording shows that it was difficult for the speaker to control for vowel reduction in the recording of trisyllabic words. Consequently, many trisyllabic words that were recorded ended up not being used in the study. In total, 16 bisyllabic words and 6 trisyllabic words were chosen for the study. The fillers set was supplemented by a previously recorded fillers set that was used for the pilot study, totalling 38 fillers (see appendix A.5 for all the fillers).

All words were manually labelled and extracted from the original recording using Praat for further acoustic analysis, totalling 264 sound files. High pass filtering was applied when applicable to remove noise from the recording. The sound clips were concatenated for each stimulus, with a 1-second silence period in between two words or in between three words and copied to create a new sound clip with the sounds repeated three times, and each repetition was separated by a 2-second silence period. Stimuli were embedded in a Qualtrics survey in autoplay mode, so that participants would not have control over when the sounds are played to them. The example below demonstrates how GAUbert and ROseemund appeared in the test instrument. The order of the first round of stimuli does not correspond to the stress order in the words; in other words, sometimes, the word with syllable-initial stress was played first, sometimes the other way around. The order of the stimuli's presentation was randomized.

Audio playing: GAUbert - pause - gauBERT - pause - GAUbert - pause - gauBERT - pause - GAUbert - pause - gauBERT

Audio playing: GAUbert (question intonation).

Question: The sentence being played contains one of the sounds you've heard previously. Which one is it? Answer choices: FIRST SOUND, SECOND SOUND

[2] These data will be used for another project and not discussed in this paper

Audio playing: ROseemund - pause - roSEEmund - pause - roseeMUND - pause - ROseemund - pause - roSEEmund - pause - roseeMUND - pause - ROseemund - pause - roSEEmund - pause - roseeMUND

Audio playing: ROseemund (statement intonation).

Question: The sentence being played contains one of the sounds you've heard previously. Which one is it? Answer choices: FIRST SOUND, SECOND SOUND, THIRD SOUND

The test instrument was built from 4 blocks, where each block contains 5 or 6 stimuli (4 stimuli from the bisyllabic words, corresponding to 4 conditions and either one or two trisyllabic words) and 10 or 9 fillers (see Appendix A.6). Each test block therefore contains 15 stimuli. The stimuli were randomized within each block. Six test lists were created using a Latin square design, so that each participant only sees one stimulus from each tokens set. The test lists were distributed relatively proportionally among the participants. Each test list collected from between 3 to 6 responses.

After finishing the speech perception task, L1 Vietnamese L2 English participants were asked to transcribe 16 real words using Vietnamese orthography, specifying in details the tone marks and diacritics[3].

4.3 Procedure

The participants answered the test online via a web link issued by Qualtrics[4]. Six test lists were available for the L1 English speakers and six other test lists were available for the L1 Vietnamese speakers. Besides the background questionnaire and the screening cloze test mentioned in the previous section, the test instrument includes two practice questions, 60 forced choice speech perception questions, and 16 transcription questions that only the experimental group completed[5].

In the main test, the participants were instructed to listen to the stimuli in groups of two or three and take note of the difference(s) among the items. The participants completed two practice questions before the actual test began. The audio of the stimuli was repeated three times, for example, 'GAUbert - pause - gauBERT - pause - GAUbert - pause - gauBERT - pause - GAUbert - pause - gauBERT'. The participants did not have control over the audio. The audio played automatically when they moved through the survey page by page. After the audio finished playing, the participants would click 'Next' to go to the forced choice perception task. In this task, the participants listened to a word containing one of the stimuli previously uttered as a single word, but uttered with a statement or a question intonation. For example, 'GAUbert (question intonation)'. This stimulus was also repeated three times. The participants then answered if the word just uttered was the first or second word (or third word if the question concerns a trisyllabic word) of the words that they have just heard previously. Once the participants answered and moved onto the next question, they were not able to return and edit their answer. The time participants spent on each question, the reaction time, and the number of clicks in response to the stimuli were recorded.

The transcription task consisted of an audio of a word, repeated three times to the participants. Only Vietnamese speakers completed this task. The speakers were instructed to transcribe the words they heard using Vietnamese orthography, specifying tone markers and diacritics.

4.4 Acoustic statistics of the stimuli

All stimuli used in the test were natural recordings. Upon obtaining the recordings of the stimuli, some acoustic statistics were extracted from the speech samples (see Table 4.4). As this study concerns the impact on F0 on stress perception, the average F0 values of all stressed syllables were calculated for each stimulus. The procedure was done using the Praat software. The stressed syllables were manually segmented and labeled in a TextGrid, and a script was written to extract these stressed syllables as well as query out F0 values from three equally spaced time points. The table below presents the F0 values of stressed syllables in the bisyllabic stimuli and the stressed syllable's duration. Where the stressed syllable was too short, it was not always possible to obtain three F0 values from the speech signal. The word-intial stress was coded as 10, the word-final stress was coded as 01, _d was a code for the statement intonation, and q_ was a code for the question intonation. For many of these cases, the F0 on the stressed syllable in the word initial position, question intonation was lowered than the F0 in the stressed syllable

[3]These data will be used for another project and not discussed in this paper

[4]The full tests are located at https://illinoislas.qualtrics.com/jfe/form/SV_3lztPSumrMhCbOd (experimental group test) https://illinoislas.qualtrics.com/jfe/form/SV_54kNUrkuoC3IrqJ (control group test). These are only two test lists out of twelve test lists that were constructed

[5]These data will be used for another project and not discussed in this paper

in the word-initial position, statement intonation. However, it is not always the case, we might need to look at the relative change within a word to see how the F0 in the word-initial position compares between questions and statements.

Stimuli	Duration	F0 at A	F0 at B	F0 at C	Average F0
10_gaubert_d	0.207	196.750	164.05	156.690	176.720
10_gaubert_q	0.209	155.250	150.930	159.070	155.083
01_gaubert_d	0.244	172.550	133.340	–undefined–	152.945
01_gaubert_q	0.234	161.630	184.980	–undefined–	173.305
10_pokgrom_d	0.182	–undefined–	219.460	–undefined–	219.460
10_pokgrom_q	0.277	–undefined–	212.760	–undefined–	212.760
01_pokgrom_d	0.562	191.25	147.970	–undefined–	147.970
01_pokgrom_q	0.320	171.020	192.450	–undefined–	181.735
10_taebak_d	0.248	–undefined–	207.160	147.92	207.160
10_taebak_q	0.200	–undefined–	160.980	169.860	165.420
01_taebak_d	0.377	170.69	–undefined–	–undefined–	170.69
01_taebak_q	0.412	–undefined–	289.03	–undefined–	289.03
10_saevir_d	0.197	201.65	176.880	164.95	176.880
10_saevir_q	0.170	186.520	175.240	178.390	180.050
01_saevir_d	0.364	185.180	–undefined–	–undefined–	185.180
01_saevir_q	0.246	178.900	180.91	–undefined–	178.900
10_moray_d	0.277	173.730	165.100	166.05	169.415
10_moray_q	0.192	201.520	183.190	169.520	184.743
01_moray_q	0.303	170.890	206.190	289.800	222.293
01_moray_d	0.437	193.110	–undefined–	–undefined–	193.110
10_moshee_d	0.299	167.800	180.07	178.32	167.800
10_moshee_q	0.287	174.740	226.750	–undefined–	200.745
01_moshee_d	0.536	158.200	167.630	–undefined–	162.915
01_moshee_q	0.349	–undefined–	196.160	362.45	196.160
10_panshee_d	0.286	–undefined–	212.400	186.450	199.425
10_panshee_q	0.265	–undefined–	178.590	201.240	189.915
01_panshee_d	0.473	–undefined–	168.530	–undefined–	168.530
01_panshee_q	0.331	–undefined–	321.480	–undefined–	321.480
10_tanvo_d	0.328	–undefined–	201.870	165.93	201.870
10_tanvo_q	0.314	–undefined–	182.280	169.440	175.860
01_tanvo_d	0.377	201.40	153.980	–undefined–	153.980
01_tanvo_q	0.341	166.460	206.960	–undefined–	186.710
10_oblak_d	0.212	–undefined–	175.130	186.90	175.130
10_oblak_q	0.174	172.03	163.410	173.940	168.675
01_oblak_d	0.480	185.490	–undefined–	–undefined–	185.490
01_oblak_q	0.421	173.280	329.17	–undefined–	173.280
10_anna_d	0.294	–undefined–	197.430	158.41	197.430
10_anna_q	0.180	146.71	175.500	187.730	181.615
01_anna_d	0.313	191.370	–undefined–	182.470	186.920
01_anna_q	0.318	174.740	212.45	–undefined–	174.740
10_danbah_d	0.300	–undefined–	187.580	159.58	187.580
10_danbah_q	0.272	186.04	161.910	–undefined–	161.910
01_danbah_d	0.372	208.770	–undefined–	–undefined–	208.770
01_danbah_q	0.296	–undefined–	248.64	–undefined–	248.64

Stimuli	Duration	F0 at A	F0 at B	F0 at C	Average F0
10_bennet_d	0.213	175.140	–undefined–	188.810	181.975
10_bennet_q	0.218	190.330	210.04	187.790	189.060
01_bennet_d	0.283	182.690	175.760	–undefined–	179.225
01_bennet_q	0.326	171.580	231.100	–undefined–	201.340
10_miro_d	0.257	199.540	180.540	158.34	190.040
10_miro_q	0.254	184.330	184.350	218.470	195.717
01_miro_d	0.376	204.740	156.620	–undefined–	180.680
01_miro_q	0.367	174.120	189.200	305.17	181.660
10_zaabir_d	0.335	177.680	–undefined–	184.330	181.005
10_zaabir_q	0.314	–undefined–	162.37	188.840	188.840
01_zaabir_d	0.369	218.92	153.370	172.320	162.845
01_zaabir_q	0.379	–undefined–	196.400	–undefined–	196.400
10_kantar_d	0.351	–undefined–	187.430	–undefined–	187.430
10_kantar_q	0.322	–undefined–	171.280	173.35	171.280
01_kantar_d	0.440	–undefined–	163.130	172.130	167.630
01_kantar_q	0.396	–undefined–	183.590	318.560	251.075
10_peenan_d	0.308	–undefined–	243.630	194.860	219.245
10_peenan_q	0.234	–undefined–	201.850	206.360	204.105
01_peenan_d	0.396	204.47	171.300	167.410	169.355
01_peenan_q	0.356	177.810	264.220	295.960	245.997

5 Results

5.1 Pilot study result

Prior to conducting this full-scale study, a small pilot study was conducted with six participants. The control group was recruited via Amazon Mechanical Turk and the experimental group was recruited offline. The pilot experimental group's age range is 18 to 34: one participant is 18-24 and two participants are 25-34 of age. They have all started learning English between primary and middle school: at 8, 10, and 11 years old. None of them reported having lived in the United States [6]. They all know a third language (Japanese) with two participants reporting an intermediate level. Two participants reported standardized English test scores, in TOEFL (99), IELTS (7), and TOEIC (955). The control group's age range is 35 to 74, with two participants reporting 35 to 44 age range and the other participant reporting 65 to 74 age range. None of them has had knowledge of another language other than English. The participants responded to the test online via a web link issued by Surveymonkey[7]. Besides the background questionnaire, the test instrument includes one practice question, 80 forced choice speech perception questions, and 8 transcription questions that only the experimental group completed. The format of the forced choice speech perception questions and the transcription questions is similar to as the format in the full study, except that the stimulus for each question in the perception task included a full sentence, the audios were only played twice in the pilot and no reaction time or time spent on each question was recorded.

For the perception task, this pilot study measures the percentage of correct answers as the dependent variable. Table 9 below shows the raw score (out of 40) and percentage of correct answers in all test categories by the target and control group. The lowest scoring subject was a member in the experimental group, at 50% and the highest scoring subject was a member in the control group, at 90%. Not all participants in the control group performed at ceiling however, one member in the control group scored lower than two members in the experimental group and was only slightly better than the lowest scoring participant. The standard deviation statistic across 10 test categories/conditions shows that the highest scoring member in the experimental group has more variability in their accuracy across categories than the highest scoring member in the control group, whose standard deviation is the lowest among six participants. The lowest scoring member in the control group has the most variability in their accuracy

[6] As this was a pilot study, we weren't adamant about finding L2 speakers who reside in the United States. However, this was a requirement in the full study

[7] The full tests are located at https://www.surveymonkey.com/r/JQQCY92 (experimental group pilot test) https://www.surveymonkey.com/r/Q688VWJ (control group pilot test)

across categories, which could be interpreted that they were much more accurate in certain categories than others.

Subject	Raw Score	Percentage	Average score across categories	Std. across categories
VI 1	20	50	2	1.05
VI 2	29	72.5	2.9	1.37
VI 3	28	70	2.8	1.14
Control 1	29	72.5	2.9	0.99
Control 2	25	62.5	2.5	1.51
Control 3	36	90	3.6	0.70

Table 9: Raw score and percentage of correct answers in speech perception task

Table 10 presents a by-subject summary of raw scores for all categories of the syllabic words in the experimental group. Based on these statistics, it seems the word-initial stress was particularly more difficult for the experimental group than the other categories, as its mean score in percentage was lower than the score of the word-final stress category. For the condition word-initial stress in yes/no question, which was predicted to be more difficult for the experimental group than the control group, the mean score was also the lowest among all mean scores across categories. The condition word-final stress in declarative context seemed less challenging for the three participants than the other categories (mean score in percentage at 92%).

Category	VI 1	VI 2	VI 3	Mean score	Standard deviation
Bisyllabic: word-initial stress + declarative	2 (50%)	2 (50%)	2 (50%)	2 (50%)	0.00
Bisyllabic: word-initial stress + yes/no	1 (25%)	3 (75%)	1 (25%)	1.67 (42%)	1.15
Bisyllabic: word-final stress + declarative	3 (75%)	4 (100%)	4 (100%)	3.67 (92%)	0.58
Bisyllabic: word-final stress + yes/no	2 (50%)	2 (50%)	3 (75%)	2.33 (58%)	0.58

Table 10: By-subject summary of raw scores for all categories of the bisyllabic words in the experimental group

Table 11 presents a by-subject summary of raw scores for all categories of the syllabic words in the control group. The mean score seems to be consistent with our prediction that Vietnamese native speakers would have difficulty perceiving lexical stress location in contexts of a yes/no question when the stress is not word-final, as the control group's mean score was 92% while the experimental group's mean score was 42% for the word-initial stress + yes/no question category. Both groups performed best in perceiving word-final stress in a declarative context (experimental group: 92% and control group: 100%). It was interesting that the control group had low mean score for the word-initial stress in a declarative context, but this seemed largely due to the performance of one participant in particular.

Category	Control 1	Control 2	Control 3	Mean score	Standard deviation
Bisyllabic: word-initial stress + declarative	3 (75%)	0	4 (100%)	2.33 (58%)	2.08
Bisyllabic: word-initial stress + yes/no	4 (100%)	3 (75%)	4 (100%)	3.67 (92%)	0.58
Bisyllabic: word-final stress + declarative	4 (100%)	4 (100%)	4 (100%)	4.00 (100%)	0.00
Bisyllabic: word-final stress + yes/no	3 (75%)	2 (50%)	3 (75%)	2.67 (67%)	0.58

Table 11: By-subject summary of raw scores for all categories of the bisyllabic words in the control group

For the trisyllabic words, the L1 Vietnamese L2 English group performed well with the declarative context, as can be seen in the mean score at 92% for word-initial stress and word-medial stress items in declarative contexts shown in Table 12. However, the participants seemed to have difficulty with the word-final stress conditions (67% for the declarative context and 33% for the yes/no question context). The two conditions that were predicted to be challenging for these participants were word-initial stress + yes/no and word-medial stress + yes/no and this preliminary result shows that indeed the mean score was not very high for these conditions, at 58%. There was a lot of variation in the word-medial stress + yes/no condition (std = 2.08) because the highest performing participant got all items correct while the lowest performing participant did not get any items correct.

Category	VI 1	VI 2	VI 3	Mean score	Standard deviation
Trisyllabic: word-initial stress + declarative	3 (75%)	4 (100%)	4 (100%)	3.67 (92%)	0.58
Trisyllabic: word-initial stress + yes/no	2 (50%)	4 (100%)	1 (25%)	2.33 (58%)	1.53
Trisyllabic: word-medial stress + declarative	3 (75%)	4 (100%)	4 (100%)	3.67 (92%)	0.58
Trisyllabic: word-medial stress + yes/no	0	4 (100%)	3 (75%)	2.33 (58%)	2.08
Trisyllabic: word-final stress + declarative	3 (75%)	2 (50%)	3 (75%)	2.67 (67%)	0.58
Trisyllabic: word-final stress + yes/no	1 (25%)	0	3 (75%)	1.33 (33%)	1.53

Table 12: By-subject summary of raw scores for all categories of the trisyllabic words in the experimental group

Table 13 shows the result of raw scores for all categories of the trisyllabic words in the control group. The preliminary statistics shown here seem to be supporting the hypothesis proposed in this pilot, as the experimental group had lower mean scores than the control group when working with the word-initial and word-medial stress position in a yes/no question context.

Category	Control 1	Control 2	Control 3	Mean score	Standard deviation
Trisyllabic: word-initial stress + declarative	2 (50%)	0	4 (100%)	2.00 (50%)	2.00
Trisyllabic: word-initial stress + yes/no	3 (75%)	4 (100%)	4 (100%)	3.67 (92%)	0.58
Trisyllabic: word-medial stress + declarative	4 (100%)	3 (75%)	4 (100%)	3.67 (92%)	0.58
Trisyllabic: word-medial stress + yes/no	1 (25%)	3 (75%)	4 (100%)	2.67 (67%)	1.53
Trisyllabic: word-final stress + declarative	3 (75%)	4 (100%)	2 (50%)	3.00 (75%)	1.00
Trisyllabic: word-final stress + yes/no	2 (50%)	2 (50%)	3 (75%)	2.33 (58%)	0.58

Table 13: By-subject summary of raw scores for all categories of the trisyllabic words in the control group

The below Table 14 shows the tone assignment result collected from three participants in the transcription task. The transcription data collected at this stage are not very meaningful, because two participants did not use Vietnamese tone markers in their transcriptions, making it ambiguous whether they intended the tone assignment to be the ML tone, which typically is not marked orthographically by any tone markers, or whether they specifically did not assign any tone to the word. Participants seemed to have misunderstood this task, as one participant directly used English to transcribe the sound, and another participant used the stress mark to transcribe stress rather than use Vietnamese tones. Only one participant transcribed the words with tones and Vietnamese orthography and one transcription (of *massif*) matches this pilot study's prediction. The transcription of *benet* was unexpected, as the stressed syllable was assigned a low falling tone and the same syllable unstressed was assigned a mid level tone.

Word	V1	V2	V3
MASsif	-	-	MR
masSIF	-	-	MFC
MOray	-	ML	ML
moRAY	-	ML	LF
BEnet	ML	ML	LF
beNET	ML	ML	ML
OBlak	ML	ML	ML
obLAK	ML	ML	ML

Table 14: Tone assignment result

Overall, the preliminary descriptive statistics of the pilot shows some support for the hypothesis regarding stress perception of L1 Vietnamese L2 English speakers. The mean scores of accuracy in stress location identification for stimuli in yes/no questions were lower than those in statement statements, and lower than the same categories' scores of the control group, as can be seen in tables 10 to table 13. In

the case of trisyllabic words, the standard deviation of accuracy scores of the experimental group for the yes/no questions was higher than that for the statement statement, due to one participant's higher than average performance in those categories. The participant also had the highest level of proficiency among the group. The overall result showed that the control group did not outperform the experimental group by a great amount, and thus the pilot's finding was inconclusive. We also had the lowest performing member in the control group scoring lower than two participants in the experimental group. This might be due to rather lax screening procedures of the control group's participants conducted for the pilot study, which was subsequently improved in the later full-scale study.

5.2 Full study results

5.2.1 Cloze test and perception tasks

A summary of the performance of the two groups on the cloze test is presented in the box plot in figure 2. The L1 English group outperformed the L1 Vietnamese group and the difference in the score mean is statistically significant ($p < 0.05$), shown in the Welch two sample t-test result below.

Listing 1: Summary of L1 Vietnamese group's cloze test scores

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
24.00	35.00	40.00	38.84	44.00	47.00

Listing 2: Summary of L1 English group's cloze test scores

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
41.00	44.25	47.50	46.77	48.00	50.00

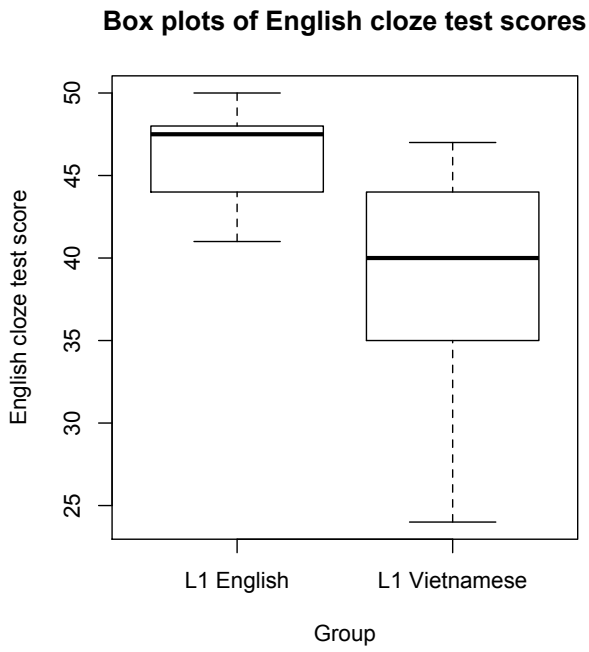


Figure 2: Box plot of two groups' performance on the English cloze test

Listing 3: Two-tailed t-test result comparing two groups' cloze score means

```
Welch Two Sample t-test

data:  en$CLOZE.SCORE and vi$CLOZE.SCORE
t = 5.2762, df = 22.391, p-value = 2.561e-05
```

```
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  4.814455 11.039796
sample estimates:
mean of x mean of y
 46.76923  38.84211
```

Regarding the main test, both groups' performance in the perception task (including the fillers) and stress perception task (only test stimuli are considered) is shown in the box plots 3 and 4. Again, the L1 English group's score distribution suggests that the L1 English speakers outperformed the L1 Vietnamese group, and the mean score differences were found to be statistically significant ($p < .05$).

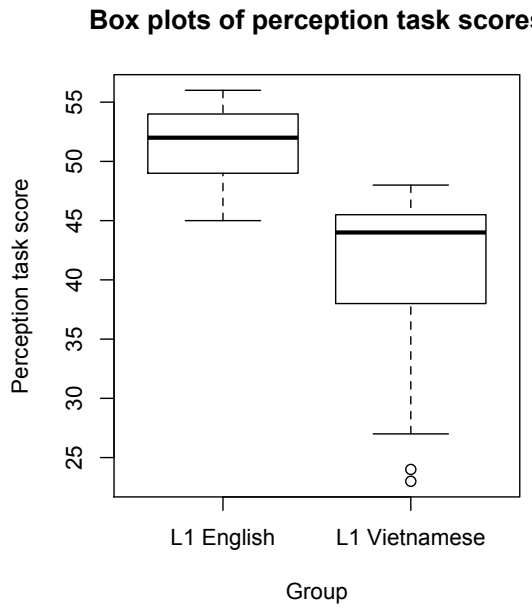


Figure 3: Box plot of two groups' performance on the main test

Listing 4: Two-tailed t-test result comparing two groups' perception score means

```
        Welch Two Sample t-test

data:  en$PERCEPTION.SCORE and vi$PERCEPTION.SCORE
t = 5.7848, df = 23.805, p-value = 5.979e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  6.712593 14.164331
sample estimates:
mean of x mean of y
 51.03846  40.60000
```

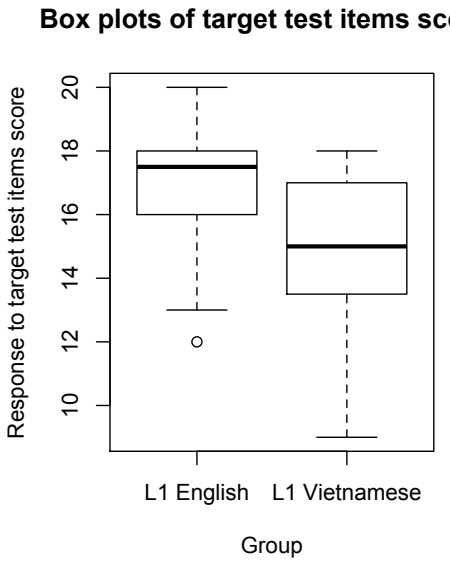


Figure 4: Box plot of two groups' performance on the stress perception test

Listing 5: Two-tailed t-test result comparing two groups' stress perception score means

```
        Welch Two Sample t-test

data:  en$STRESS.PERCEPTION.SCORE and vi$STRESS.PERCEPTION.SCORE
t = 3.4017, df = 33.484, p-value = 0.001749
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.010244 4.012833
sample estimates:
mean of x mean of y
 16.96154  14.45000
```

5.2.2 Identification accuracy

The identification accuracy for bisyllabic words was calculated as the percentage of correct response with respect to matching the word to the same stress pattern, out of the total number of stimuli. For each intonation type, this identification accuracy was calculated out of eight stimuli for bisyllabic words. The figures in 5 and 6 show that accuracy was generally lowered for the question intonation, and especially so for the L1 Vietnamese group. The same chart for the trisyllabic words was created and this pattern is quite evident here as well. The identification accuracy for trisyllabic words of each intonation type was calculated out of three stimuli that were presented to each participant. From 8, it can be seen that more than half the participants in the L1 Vietnamese group fails to correctly match trisyllabic words with the correct stress pattern (n = 12 with the identification accuracy being 0), while there was only one such instance in the L1 English group. Most of the L1 English speakers did moderately well, getting two out of three correct for the majority, but this group of stimuli was a challenge for the L1 English speakers too, for very few managed to score 100 percent. The identification accuracy of trisyllabic words in statement intonation among the L1 English speakers looks much better, with more than half of the participants scoring at least two out of three correct.

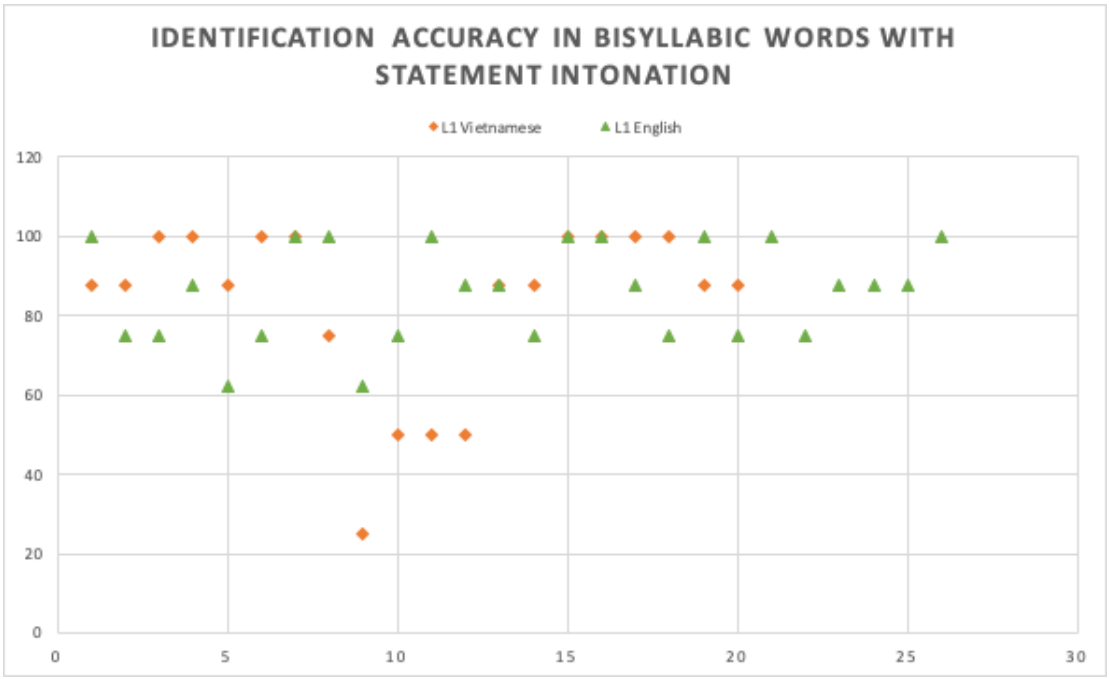


Figure 5: The identification accuracy for bisyllabic words in statement intonation for both groups

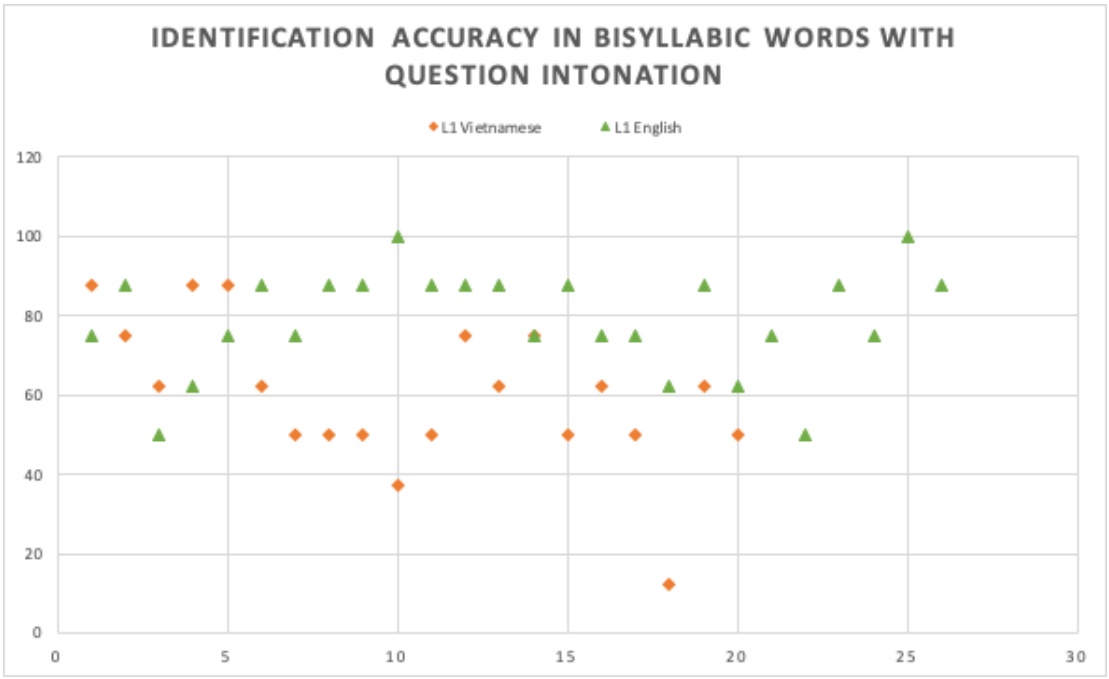


Figure 6: The identification accuracy for bisyllabic words in question intonation for both groups

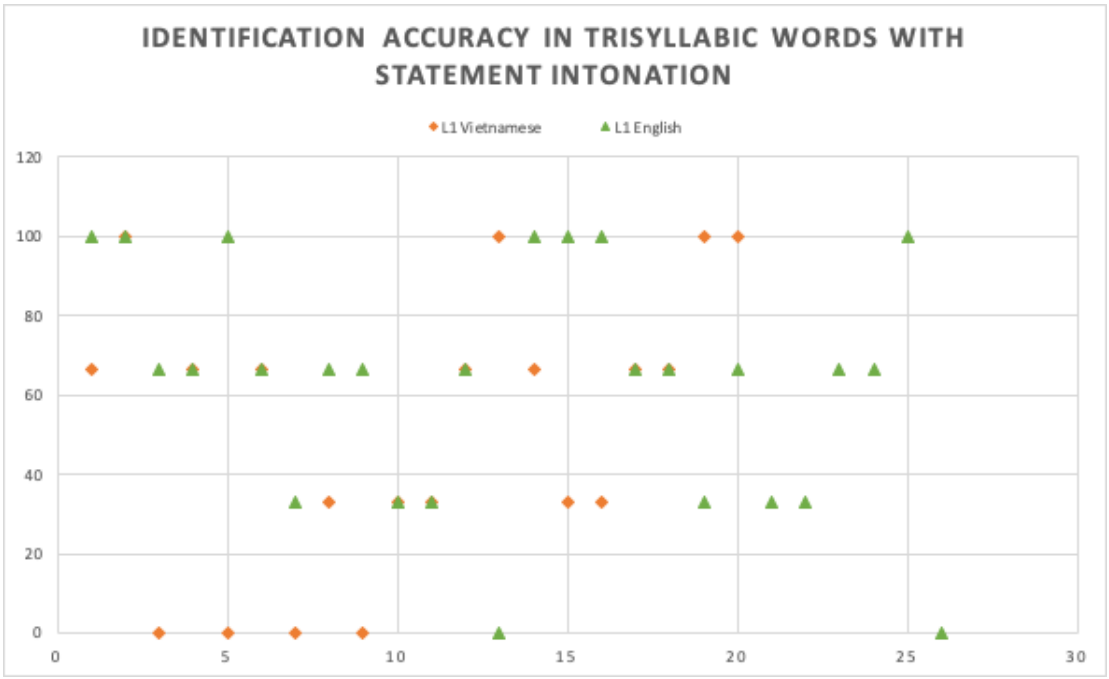


Figure 7: The identification accuracy for trisyllabic words in statement intonation for both groups

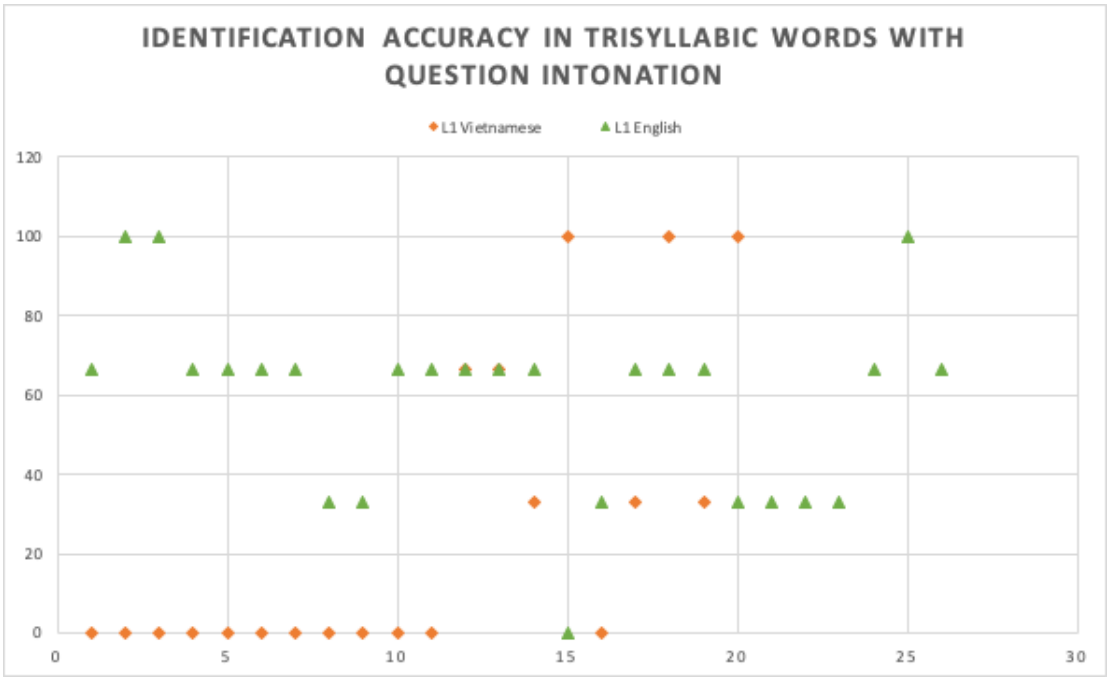


Figure 8: The identification accuracy for trisyllabic words in question intonation for both groups

5.2.3 Mixed effects analysis

5.2.3.1 Bisyllabic stimuli

In order to answer research question 1, this study examines the performance of participants, by measuring the number of correct responses as dependent variable when varying one between-subjects factor (L1) and two within-subjects factors (the intonation contour type with two levels: sentence or question intonation and the stress location in the word with two levels: initial or final for the bisyllabic words).

A general linear mixed ANOVA was run with the aforementioned factors and the results presented below. The sentence intonation level is labeled as SENTENCE, the question intonation level is labeled as QUESTION, the word-initial stress location is labeled as 10, the word-final stress location is labeled as 01, L1 English group is coded as 0, and L1 Vietnamese group is coded as 1 in the subsequent statistical summaries.

In Table 15, it can be seen that the experiment group's mean scores are lower than the control group's mean scores across conditions, except for the word-final stress, statement intonation condition, where the mean scores of the two groups are about the same.

Descriptive Statistics				
L1		Mean	Std. Deviation	N
STATEMENT.10	0	3.35	0.629	26
	1	3.10	1.294	20
	Total	3.24	0.970	46
STATEMENT.01	0	3.54	0.706	26
	1	3.55	0.759	20
	Total	3.54	0.721	46
QUESTION.10	0	3.27	0.827	26
	1	2.00	1.124	20
	Total	2.72	1.148	46
QUESTION.01	0	3.04	0.774	26
	1	2.80	1.005	20
	Total	2.93	0.879	46

Table 15: Descriptive statistics of stress perception scores in bisyllabic target stimuli

The within-subject effects summary in Table 16 shows that there is a significant main effect of intonation type, $F(1, 44) = 22.7$, $p < .05$ on the stress perception score. There is also a statistically significant main effect of the stress position, $F(1, 44) = 5.762$, $p < .05$. There is no evidence showing any interaction between the within subjects variables, sentence type and stress location, which is also supported by the interaction plot in 9. On the other hand, there is an interaction effect between the intonation type and the L1 factor, $F(1, 44) = 6.254$, $p < .05$ and similarly there is an interaction effect between the stress location and the L1 factor as well, $F(1, 44) = 6.517$, $p < .05$.

Tests of Within-Subjects Effects						
Measure:	MEASURE_1					
Source		Type III SoS	df	Mean Square	F	Sig.
Sentence_type	Sphericity Assumed	16.646	1	16.646	22.728	0.000
	Greenhouse-Geisser	16.646	1.000	16.646	22.728	0.000
	Huynh-Feldt	16.646	1.000	16.646	22.728	0.000
	Lower-bound	16.646	1.000	16.646	22.728	0.000
Sentence_type * L1	Sphericity Assumed	4.580	1	4.580	6.254	0.016
	Greenhouse-Geisser	4.580	1.000	4.580	6.254	0.016
	Huynh-Feldt	4.580	1.000	4.580	6.254	0.016
	Lower-bound	4.580	1.000	4.580	6.254	0.016
Error(Sentence_type)	Sphericity Assumed	32.224	44	0.732		
	Greenhouse-Geisser	32.224	44.000	0.732		
	Huynh-Feldt	32.224	44.000	0.732		
	Lower-bound	32.224	44.000	0.732		
Stress_location	Sphericity Assumed	4.148	1	4.148	5.762	0.021
	Greenhouse-Geisser	4.148	1.000	4.148	5.762	0.021
	Huynh-Feldt	4.148	1.000	4.148	5.762	0.021
	Lower-bound	4.148	1.000	4.148	5.762	0.021
Stress_location * L1	Sphericity Assumed	4.692	1	4.692	6.517	0.014
	Greenhouse-Geisser	4.692	1.000	4.692	6.517	0.014
	Huynh-Feldt	4.692	1.000	4.692	6.517	0.014
	Lower-bound	4.692	1.000	4.692	6.517	0.014
Error(Stress_location)	Sphericity Assumed	31.678	44	0.720		
	Greenhouse-Geisser	31.678	44.000	0.720		
	Huynh-Feldt	31.678	44.000	0.720		
	Lower-bound	31.678	44.000	0.720		
Sentence_type * Stress_location	Sphericity Assumed	0.015	1	0.015	0.020	0.887
	Greenhouse-Geisser	0.015	1.000	0.015	0.020	0.887
	Huynh-Feldt	0.015	1.000	0.015	0.020	0.887
	Lower-bound	0.015	1.000	0.015	0.020	0.887
Sentence_type * Stress_location * L1	Sphericity Assumed	1.689	1	1.689	2.271	0.139
	Greenhouse-Geisser	1.689	1.000	1.689	2.271	0.139
	Huynh-Feldt	1.689	1.000	1.689	2.271	0.139
	Lower-bound	1.689	1.000	1.689	2.271	0.139
Error(Sentence_type*Stress_location)	Sphericity Assumed	32.724	44	0.744		
	Greenhouse-Geisser	32.724	44.000	0.744		
	Huynh-Feldt	32.724	44.000	0.744		
	Lower-bound	32.724	44.000	0.744		

Table 16: Within-subjects effects

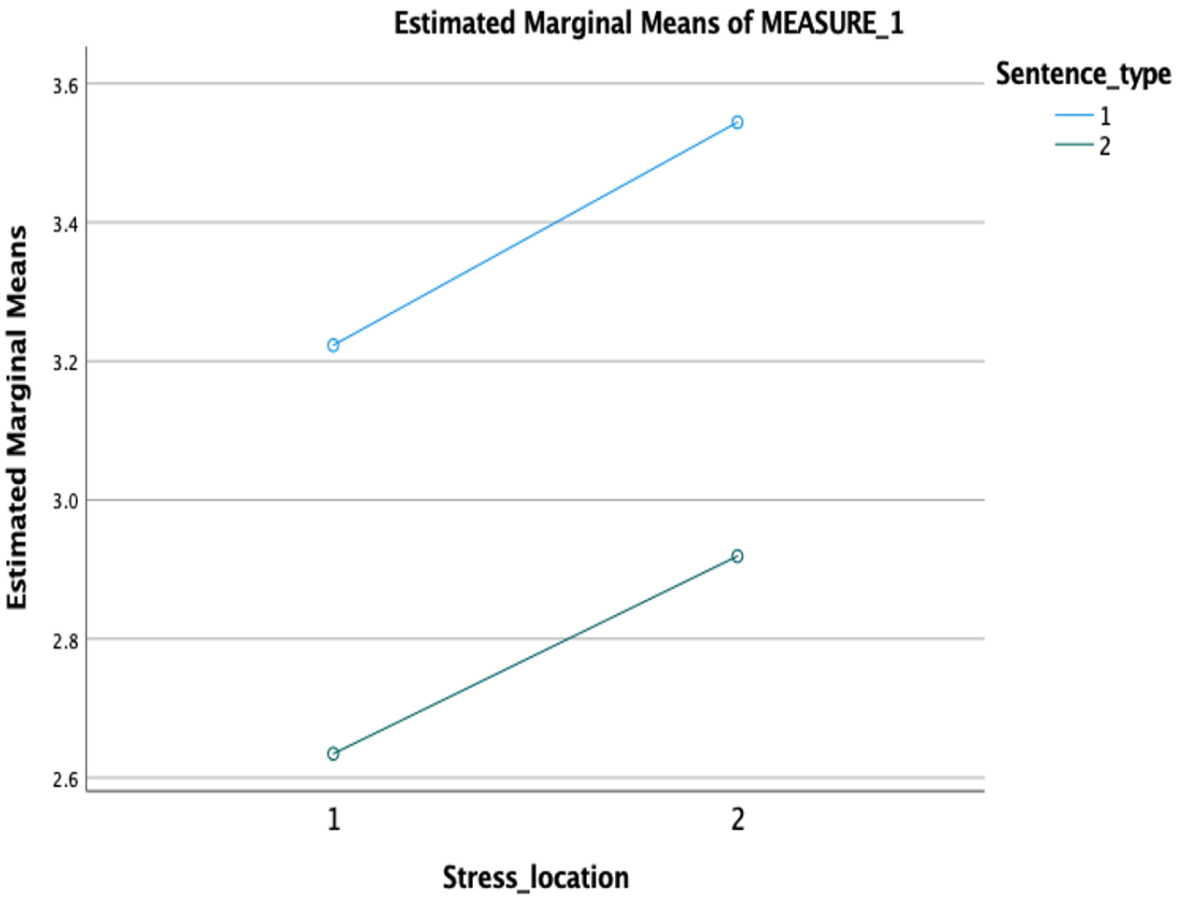


Figure 9: Interaction plot

The between-subjects effects summary below shows that there is a statistically significant difference between the stress perception score across the two groups, $F(1, 44) = 8.629$, $p < .05$, and going by the descriptive statistics, this means there is evidence supporting that the L1 English speakers' stress perception performance is better than the L1 Vietnamese speakers' performance, for bisyllabic words.

Tests of Between-Subjects Effects					
Measure:	MEASURE_1				
Transformed Variable:	Average				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	1716.122	1	1716.122	1726.043	0.000
L1	8.579	1	8.579	8.629	0.005
Error	43.747	44	0.994		

Table 17: Between-subjects effects

To examine the interaction between the L1 factor and intonation type/stress pattern factors, independent samples t-tests with a Bonferroni correction were run. The result summaries below reveal that the source of the interaction is in the question intonation and the word-initial stress location, crossing with the L1 factor. Therefore, there is evidence supporting that L2 speakers performed worse in question intonation and word-intial stress conditions, fully in line with the hypothesis set forth at the beginning of the paper. Detailed results of these independent samples t-tests are presented below.

According to the descriptive summary of stress perception scores in Table 19, the L1 English speakers scored on average 3 points higher than the L2 English speakers when presented with statement stimuli, and they scored 18 points on average more than the L2 English speakers when presented with *question*

stimuli. The t-test, shown in Table **??** shows statistically significant difference in the question intonation category after a Bonferroni correction (t(44) = -4.056, p <0.0125 or 0.05/4), but not in the statement category. The effect size of this difference is quite large, at 17.1892, shown in Table 20.

Group Statistics					
L1		N	Mean	Std. Deviation	Std. Error Mean
FACTOR.STATEMENT	1	20	83.125	21.9430	4.9066
	0	26	86.058	12.4132	2.4344
FACTOR.QUESTION	1	20	60.000	18.4069	4.1159
	0	26	78.846	13.1193	2.5729

Table 18: Stress perception scores of L1 English (=0) and L1 Vietnamese (=1) groups for words with different intonation types

Independent Samples Test					
	Levene's Test for Equality of Variances		t-test for Equality of Means		
	F	Sig.	t	df	Sig. (2-tailed)
FACTOR.STATEMENT	3.993	0.052	-0.574	44	0.569
			-0.535	28.207	0.597
FACTOR.QUESTION	1.490	0.229	-4.056	44	0.000
			-3.883	32.929	0.000

Table 19: Independent samples tests result

Independent Samples Effect Sizes					
		Standardizera	Point Estimate	95% Confidence Interval	
				Lower	Upper
FACTOR.STATEMENT	Cohen's d	17.1892	-0.171	-0.754	0.414
	Hedges' correction	17.4893	-0.168	-0.741	0.407
	Glass's delta	12.4132	-0.236	-0.821	0.353
FACTOR.QUESTION	Cohen's d	15.6237	-1.206	-1.835	-0.566
	Hedges' correction	15.8965	-1.186	-1.804	-0.556
	Glass's delta	13.1193	-1.437	-2.131	-0.722

Table 20: Independent samples effect sizes

According to the descriptive summary of stress perception scores by stress location in Table 22, the L1 English speakers scored on average about 21 points higher than the L2 English speakers when presented with word-intial stress stimuli, and they scored 3 points on average more than the L2 English speakers when presented with word-final stress stimuli. The t-test, shown in Table 22 shows statistically significant difference in the word-intial category after a Bonferroni correction (t(44) = -4.056, p <0.0125), but not in the word-final category. The effect size of this difference is quite large, at 16.8501, shown in Table 23.

Group Statistics					
L1		N	Mean	Std. Deviation	Std. Error Mean
FACTOR.10	1	20	63.750	21.0341	4.7034
	0	26	82.692	12.7852	2.5074
FACTOR.01	1	20	79.375	16.8561	3.7691
	0	26	82.212	15.0719	2.9559

Table 21: Stress perception scores of L1 English (=0) and L1 Vietnamese (=1) groups for words with different stress locations

Independent Samples Test					
	Levene's Test for Equality of Variances		t-test for Equality of Means		
	F	Sig.	t	df	Sig. (2-tailed)
FACTOR.10	7.733	0.008	-3.780	44	0.000
			-3.554	29.522	0.001
FACTOR.01	0.418	0.521	-0.601	44	0.551
			-0.592	38.492	0.557

Table 22: Independent samples tests result

Independent Samples Effect Sizes					
		Standardizera	Point Estimate	95% Confidence Interval	
				Lower	Upper
FACTOR.10	Cohen's d	16.8501	-1.124	-1.747	-0.490
	Hedges' correction	17.1443	-1.105	-1.717	-0.482
	Glass's delta	12.7852	-1.482	-2.183	-0.760
FACTOR.01	Cohen's d	15.8670	-0.179	-0.762	0.406
	Hedges' correction	16.1440	-0.176	-0.749	0.399
	Glass's delta	15.0719	-0.188	-0.772	0.399

Table 23: Independent samples effect sizes

The mixed effects ANOVA analysis on bisyllabic words suggests that there is a clear difference in stress perception performance between the two groups of participants, where the L1 English speakers performed better than the L1 Vietnamese speakers. Notably, the analysis gives support to the hypothesis that the difference between the two groups are due to the question intonation condition and the word-initial condition. The effect size of this result is large, suggesting that it is quite robust.

5.2.3.2 Trisyllabic stimuli

A mixed effects ANOVA analysis was performed on the trisyllabic stimuli, but no significant differences were found among the within subjects variables (see Table 25). Performance based on the mean scores between the two groups looks quite comparable, as seen in Table 24 and the between subject variable (L1) also doesn't show any significant effect (see Table 26).

Descriptive Statistics				
L1		Mean	Std. Deviation	N
STATEMENT.100	0	0.65	0.485	26
	1	0.40	0.503	20
	Total	0.54	0.504	46
STATEMENT.010	0	0.77	0.430	26
	1	0.45	0.510	20
	Total	0.63	0.488	46
STATEMENT.001	0	0.58	0.504	26
	1	0.65	0.489	20
	Total	0.61	0.493	46
QUESTION.100	0	0.50	0.510	26
	1	0.45	0.510	20
	Total	0.48	0.505	46
QUESTION.010	0	0.58	0.504	26
	1	0.40	0.503	20
	Total	0.50	0.506	46
QUESTION.001	0	0.69	0.471	26
	1	0.65	0.489	20
	Total	0.67	0.474	46

Table 24: Descriptive statistics of perception scores in trisyllabic target stimuli

Tests of Within-Subjects Effects						
Measure:	MEASURE_1					
Source		Type III SoS	df	Mean Square	F	Sig.
Sentence_type	Sphericity Assumed	0.100	1	0.100	0.568	0.455
	Greenhouse-Geisser	0.100	1.000	0.100	0.568	0.455
	Huynh-Feldt	0.100	1.000	0.100	0.568	0.455
	Lower-bound	0.100	1.000	0.100	0.568	0.455
Sentence_type * L1	Sphericity Assumed	0.100	1	0.100	0.568	0.455
	Greenhouse-Geisser	0.100	1.000	0.100	0.568	0.455
	Huynh-Feldt	0.100	1.000	0.100	0.568	0.455
	Lower-bound	0.100	1.000	0.100	0.568	0.455
Error(Sentence_type)	Sphericity Assumed	7.769	44	0.177		
	Greenhouse-Geisser	7.769	44.000	0.177		
	Huynh-Feldt	7.769	44.000	0.177		
	Lower-bound	7.769	44.000	0.177		
Stress_location	Sphericity Assumed	0.934	2	0.467	2.557	0.083
	Greenhouse-Geisser	0.934	1.889	0.494	2.557	0.087
	Huynh-Feldt	0.934	2.000	0.467	2.557	0.083
	Lower-bound	0.934	1.000	0.934	2.557	0.117
Stress_location * L1	Sphericity Assumed	0.804	2	0.402	2.200	0.117
	Greenhouse-Geisser	0.804	1.889	0.425	2.200	0.120
	Huynh-Feldt	0.804	2.000	0.402	2.200	0.117
	Lower-bound	0.804	1.000	0.804	2.200	0.145
Error(Stress_location)	Sphericity Assumed	16.073	88	0.183		
	Greenhouse-Geisser	16.073	83.127	0.193		
	Huynh-Feldt	16.073	88.000	0.183		
	Lower-bound	16.073	44.000	0.365		
Sentence_type * Stress_location	Sphericity Assumed	0.368	2	0.184	0.617	0.542
	Greenhouse-Geisser	0.368	1.905	0.193	0.617	0.534
	Huynh-Feldt	0.368	2.000	0.184	0.617	0.542
	Lower-bound	0.368	1.000	0.368	0.617	0.436
Sentence_type * Stress_location * L1	Sphericity Assumed	0.324	2	0.162	0.544	0.582
	Greenhouse-Geisser	0.324	1.905	0.170	0.544	0.574
	Huynh-Feldt	0.324	2.000	0.162	0.544	0.582
	Lower-bound	0.324	1.000	0.324	0.544	0.465
Error(Sentence_type*Stress_location)	Sphericity Assumed	26.219	88	0.298		
	Greenhouse-Geisser	26.219	83.820	0.313		
	Huynh-Feldt	26.219	88.000	0.298		
	Lower-bound	26.219	44.000	0.596		

Table 25: Within-subjects effects

Tests of Between-Subjects Effects					
Measure:	MEASURE_1				
Transformed Variable:	Average				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	86.332	1	86.332	275.877	0.000
L1	1.115	1	1.115	3.562	0.066
Error	13.769	44	0.313		

Table 26: Between-subjects effects

5.2.3.3 Reaction time in bisyllabic stimuli

In order to answer research question 2, 'Would difficulty in identifying the stress location in the yes/no question context manifest in a delayed reaction to the stimuli in the L2 English group?' , various reaction time measures were examined. The participants' reaction time was measured by activating the Question: Timing feature in Qualtrics. According to Qualtrics, this feature gives userrs four metrics:

1. First Click: Number of seconds from when the page loads to the first click
2. Last Click: Number of seconds from when the page loads to the click before the "Next" button is selected
3. Page Submit: Number of seconds from when the page loads to when the "Next" button is selected
4. Click Count: Number of a respondent clicks on a page

In this study, three metrics were specifically examined, First Click, Last Click, and Click Count. The first click should give an indication of how fast the participant makes a decision from the moment the audio is played. The click count should gives an indication about how many times the participant changes their mind in the course of answering the question. The last click should give an indication about the total time elapsed since the audio is played, or the total reaction time. The expectation is that L2 English speakers would take longer to answer questions, especially in the QUESTION.10 condition. However, this prediction did not bore out. What seems to be significant is L2 English speakers' likelihood for more a click count more than 1, in other words, they might be more likely to change their minds about the answers, compared to the L1 group. Looking at the descriptive data, the L1 English speakers typically took longer to answer questions than the L2 English speakers, however there was no evidence supporting that this difference was statistically significant. Mixed effects ANOVA models are presented below, as well as followed up post hoc tests, to illustrate the points just made.

Table 27 shows that L1 English speakers take longer on average than L1 Vietnamese speakers for the first click, except for the QUESTION.10 condition, when both groups take roughly the same time for the first click, on average. Table 28, however, does not give support for the difference between the two groups, the p-value is 0.347 >0.05 and we cannot reject the null hypothesis that there was no difference in the time to the first click based on L1 factor.

Descriptive Statistics				
L1		Mean	Std. Deviation	N
FIRSTCLICK.STATEMENT.10	0	11.16150	9.637444	26
	1	8.50850	4.605819	20
	Total	10.00802	7.894627	46
FIRSTCLICK.STATEMENT.01	0	10.05042	8.143267	26
	1	9.47900	6.168538	20
	Total	9.80198	7.279312	46
FIRSTCLICK.QUESTION.10	0	15.27388	13.417920	26
	1	14.31685	11.139615	20
	Total	14.85778	12.355024	46
FIRSTCLICK.QUESTION.01	0	13.37965	7.712741	26
	1	9.47255	3.395079	20
	Total	11.68091	6.461390	46

Table 27: First click in four conditions

Table 29 shows that there are main effects from two within subject variables, the sentence type and stress location factors. For the sentence type, $F(1, 44) = 10.677$, $p < .05$; and for the stress location factor, $F(1, 44) = 5.456$, $p < .05$. A followed up pairwise comparison of within subjects factors was conducted and two crossed-conditions were found to be significant.

Tests of Between-Subjects Effects					
Measure:	MEASURE_1				
Transformed Variable:	Average				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	23734.389	1	23734.389	115.916	0.000
L1	184.896	1	184.896	0.903	0.347
Error	9009.187	44	204.754		

Table 28: Tests of Between-Subjects Effects, Dependent variable: First Click

Tests of Within-Subjects Effects						
Measure:	MEASURE_1					
Source		Type III SoS	df	Mean Square	F	Sig.
Sentence_type	Sphericity Assumed	495.669	1	495.669	10.677	0.002
	Greenhouse-Geisser	495.669	1.000	495.669	10.677	0.002
	Huynh-Feldt	495.669	1.000	495.669	10.677	0.002
	Lower-bound	495.669	1.000	495.669	10.677	0.002
Sentence_type * L1	Sphericity Assumed	7.598	1	7.598	0.164	0.688
	Greenhouse-Geisser	7.598	1.000	7.598	0.164	0.688
	Huynh-Feldt	7.598	1.000	7.598	0.164	0.688
	Lower-bound	7.598	1.000	7.598	0.164	0.688
Error(Sentence_type)	Sphericity Assumed	2042.648	44	46.424		
	Greenhouse-Geisser	2042.648	44.000	46.424		
	Huynh-Feldt	2042.648	44.000	46.424		
	Lower-bound	2042.648	44.000	46.424		
Stress_location	Sphericity Assumed	133.736	1	133.736	5.456	0.024
	Greenhouse-Geisser	133.736	1.000	133.736	5.456	0.024
	Huynh-Feldt	133.736	1.000	133.736	5.456	0.024
	Lower-bound	133.736	1.000	133.736	5.456	0.024
Stress_location * L1	Sphericity Assumed	2.132	1	2.132	0.087	0.769
	Greenhouse-Geisser	2.132	1.000	2.132	0.087	0.769
	Huynh-Feldt	2.132	1.000	2.132	0.087	0.769
	Lower-bound	2.132	1.000	2.132	0.087	0.769
Error(Stress_location)	Sphericity Assumed	1078.572	44	24.513		
	Greenhouse-Geisser	1078.572	44.000	24.513		
	Huynh-Feldt	1078.572	44.000	24.513		
	Lower-bound	1078.572	44.000	24.513		
Sentence_type * Stress_location	Sphericity Assumed	123.028	1	123.028	3.514	0.067
	Greenhouse-Geisser	123.028	1.000	123.028	3.514	0.067
	Huynh-Feldt	123.028	1.000	123.028	3.514	0.067
	Lower-bound	123.028	1.000	123.028	3.514	0.067
Sentence_type * Stress_location * L1	Sphericity Assumed	71.549	1	71.549	2.044	0.160
	Greenhouse-Geisser	71.549	1.000	71.549	2.044	0.160
	Huynh-Feldt	71.549	1.000	71.549	2.044	0.160
	Lower-bound	71.549	1.000	71.549	2.044	0.160
Error(Sentence_type*Stress_location)	Sphericity Assumed	1540.354	44	35.008		
	Greenhouse-Geisser	1540.354	44.000	35.008		
	Huynh-Feldt	1540.354	44.000	35.008		
	Lower-bound	1540.354	44.000	35.008		

Table 29: Tests of Within-Subjects Effects, Dependent variable: First Click

The Table 30 shows that the first click's mean is higher for the word-initial condition and word-final condition compared to the statement condition, and the first click's mean for the question condition is higher than both stress conditions, across all subjects. This suggests that the statement condition is the easiest condition for the speech perception task. The Table 31 shows paired samples t-test results, and when the Bonferroni correction was applied, a significant effect was found for the statement vs. 10 and question vs. 01 pairs. Looking at the means, the result could be interpreted that all subjects had shorter first click time for the statement condition compared to the word-initial stress position, and all subjects had longer first click time for the question condition compared to the word-final stress position. Therefore, it seems that the statement condition is the most straightforward condition for all subjects, resulting in faster decision time, and the question and word-initial condition were relatively more challenging for all subjects. This is in line with our findings that the word-initial and question conditions would be challenging, but in this case, it was established that the L1 English group also experienced longer first click time for these conditions.

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	STATEMENT	19.81000	46	13.991768	2.062974
	10	24.86580	46	17.423821	2.569002
Pair 2	STATEMENT	19.81000	46	13.991768	2.062974
	01	21.48289	46	12.372051	1.824159
Pair 3	QUESTION	26.53870	46	17.434976	2.570647
	10	24.86580	46	17.423821	2.569002
Pair 4	QUESTION	26.53870	46	17.434976	2.570647
	01	21.48289	46	12.372051	1.824159

Table 30: Within-subjects pairwise comparison: Descriptive Statistics

Paired Samples Test							
		Paired Differences			t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean			
Pair 1	FC.STATEMENT - FC.10	-5.055804	9.897398	1.459292	-3.465	45	0.001
Pair 2	FC.STATEMENT - FC.01	-1.672891	6.418318	0.946329	-1.768	45	0.084
Pair 3	FC.QUESTION - FC.10	1.672891	6.418318	0.946329	1.768	45	0.084
Pair 4	FC.QUESTION - FC.01	5.055804	9.897398	1.459292	3.465	45	0.001

Table 31: Within-subjects pairwise comparison: t-tests for First Click

The descriptive stat summary of click counts in Table 32 suggests that there is a small difference between click counts between the L1 Vietnamese and L1 English groups. The L1 Vietnamese speakers tend to click a little more on average than the L1 English speakers, suggesting that they could be a bit more likely to change their minds. The between variables effect is borderlining significant, $F(1, 44) = 3.778$, $p = 0.058$, so it seems there is some difference between the two groups when it comes to click counts (see 33). The within-subjects effect showed that there is an interaction between the stress location and the L1 factor, leading to a followed-up independent samples t-test where the source of the interaction would be investigated.

Descriptive Statistics				
L1		Mean	Std. Deviation	N
CLICKCOUNT.STATEMENT.10	0	4.46	0.508	26
	1	4.50	0.607	20
	Total	4.48	0.547	46
CLICKCOUNT.STATEMENT.01	0	4.35	0.892	26
	1	4.85	0.988	20
	Total	4.57	0.958	46
CLICKCOUNT.QUESTION.10	0	4.58	0.809	26
	1	4.65	0.875	20
	Total	4.61	0.829	46
CLICKCOUNT.QUESTION.01	0	4.31	0.736	26
	1	4.75	1.251	20
	Total	4.50	1.006	46

Table 32: Click count in four conditions

Tests of Between-Subjects Effects					
Measure:	MEASURE_1				
Transformed Variable:	Average				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	3753.162	1	3753.162	4484.781	0.000
L1	3.162	1	3.162	3.778	0.058
Error	36.822	44	0.837		

Table 33: Tests of Between-Subjects Effects, Dependent variable: Click count

Tests of Within-Subjects Effects						
Measure:	MEASURE_1					
Source		Type III SoS	df	Mean Square	F	Sig.
Sentence_type	Sphericity Assumed	0.046	1	0.046	0.076	0.783
	Greenhouse-Geisser	0.046	1.000	0.046	0.076	0.783
	Huynh-Feldt	0.046	1.000	0.046	0.076	0.783
	Lower-bound	0.046	1.000	0.046	0.076	0.783
Sentence_type * L1	Sphericity Assumed	0.002	1	0.002	0.003	0.953
	Greenhouse-Geisser	0.002	1.000	0.002	0.003	0.953
	Huynh-Feldt	0.002	1.000	0.002	0.003	0.953
	Lower-bound	0.002	1.000	0.002	0.003	0.953
Error(Sentence_type)	Sphericity Assumed	26.199	44	0.595		
	Greenhouse-Geisser	26.199	44.000	0.595		
	Huynh-Feldt	26.199	44.000	0.595		
	Lower-bound	26.199	44.000	0.595		
Stress_location	Sphericity Assumed	0.012	1	0.012	0.024	0.878
	Greenhouse-Geisser	0.012	1.000	0.012	0.024	0.878
	Huynh-Feldt	0.012	1.000	0.012	0.024	0.878
	Lower-bound	0.012	1.000	0.012	0.024	0.878
Stress_location * L1	Sphericity Assumed	1.969	1	1.969	3.888	0.055
	Greenhouse-Geisser	1.969	1.000	1.969	3.888	0.055
	Huynh-Feldt	1.969	1.000	1.969	3.888	0.055
	Lower-bound	1.969	1.000	1.969	3.888	0.055
Error(Stress_location)	Sphericity Assumed	22.276	44	0.506		
	Greenhouse-Geisser	22.276	44.000	0.506		
	Huynh-Feldt	22.276	44.000	0.506		
	Lower-bound	22.276	44.000	0.506		
Sentence_type * Stress_location	Sphericity Assumed	0.461	1	0.461	0.497	0.484
	Greenhouse-Geisser	0.461	1.000	0.461	0.497	0.484
	Huynh-Feldt	0.461	1.000	0.461	0.497	0.484
	Lower-bound	0.461	1.000	0.461	0.497	0.484
Sentence_type * Stress_location * L1	Sphericity Assumed	0.026	1	0.026	0.028	0.867
	Greenhouse-Geisser	0.026	1.000	0.026	0.028	0.867
	Huynh-Feldt	0.026	1.000	0.026	0.028	0.867
	Lower-bound	0.026	1.000	0.026	0.028	0.867
Error(Sentence_type*Stress_location)	Sphericity Assumed	40.784	44	0.927		
	Greenhouse-Geisser	40.784	44.000	0.927		
	Huynh-Feldt	40.784	44.000	0.927		
	Lower-bound	40.784	44.000	0.927		

Table 34: Tests of Within-Subjects Effects, Dependent variable: Click Count

A followed-up independent samples t-test was conducted for L1 and the stress location, with two levels (word-intial stress and word-final stress), and the result was tabulated in Table 35 below. A significant difference in the click count variable was found for the word-final variable in Table 36, $t(44) = 2.459$, $p = 0.019$, suggesting that L1 Vietnamese speakers would be more likely to change their mind

or have higher mean click per question for the word-final stress category. This is an interesting effect that was not expected in the hypothesis of the paper.

Group Statistics					
L1		N	Mean	Std. Deviation	Std. Error Mean
CLICKCOUNT.10	1	20	9.15	1.182	0.264
	0	26	9.04	0.958	0.188
CLICKCOUNT.01	1	20	9.60	1.429	0.320
	0	26	8.65	1.093	0.214

Table 35: L1 and stress location descriptive statistics

Independent Samples Test						
	Levene's Test for Equality of Variances		t-test for Equality of Means			
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference
CLICKCOUNT.10	0.727	0.398	0.354	44	0.725	0.112
			0.344	36.063	0.733	0.112
CLICKCOUNT.01	3.895	0.055	2.546	44	0.014	0.946
			2.459	34.627	0.019	0.946

Table 36: L1 and stress location Independent Samples Test

Similarly to the first click's descriptive summary, the mean of the time to the last click of the L1 English group is longer than the L2 English group, especially in statement stimuli, see Table 37. However, this difference is not found to be statistically significant, as the Table 38 shows ($p = 0.366$). Among the within-subject effects, sentence type was found to be significant to the first click variable. A paired sample t-test for two levels (SENTENCE vs. QUESTION) was run for all participants, and it was found that the duration to the last click for statement intonation is significantly shorter than the duration to the last click for the question intonation ($t(45) = -3.870$, $p < .05$) (see Table 40. This is another evidence that the statement intonation is more straightforward for all participants in the stress perception task.

Descriptive Statistics				
L1		Mean	Std. Deviation	N
LASTCLICK.STATEMENT.10	0	12.51465	10.981134	26
	1	8.21865	4.149307	20
	Total	10.64683	8.882417	46
LASTCLICK.STATEMENT.01	0	10.24815	8.057308	26
	1	8.58475	3.764518	20
	Total	9.52493	6.537998	46
LASTCLICK.QUESTION.10	0	16.06965	13.146626	26
	1	13.84030	6.299639	20
	Total	15.10037	10.678172	46
LASTCLICK.QUESTION.01	0	13.65350	7.638222	26
	1	14.06360	11.116790	20
	Total	13.83180	9.199686	46

Table 37: Last click in four conditions

Tests of Between-Subjects Effects					
Measure:	MEASURE_1				
Transformed Variable:	Average				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	26696.715	1	26696.715	130.284	0.000
L1	171.000	1	171.000	0.835	0.366
Error	9016.148	44	204.912		

Table 38: Tests of Between-Subjects Effects, Dependent variable: Last click

Paired Samples Test							
		Paired Differences			t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean			
Pair 1	S v. Q	-8.760413043478260	15.3528306	2.26365138	-3.870	45	0.000

Table 40: Paired Samples Test for LASTCLICK.STATEMENT vs. LASTCLICK.QUESTION

Tests of Within-Subjects Effects						
Measure:	MEASURE_1					
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Sentence_type	Sphericity Assumed	921.853	1	921.853	15.581	0.000
	Greenhouse-Geisser	921.853	1.000	921.853	15.581	0.000
	Huynh-Feldt	921.853	1.000	921.853	15.581	0.000
	Lower-bound	921.853	1.000	921.853	15.581	0.000
Sentence_type * L1	Sphericity Assumed	48.442	1	48.442	0.819	0.370
	Greenhouse-Geisser	48.442	1.000	48.442	0.819	0.370
	Huynh-Feldt	48.442	1.000	48.442	0.819	0.370
	Lower-bound	48.442	1.000	48.442	0.819	0.370
Error(Sentence_type)	Sphericity Assumed	2603.289	44	59.166		
	Greenhouse-Geisser	2603.289	44.000	59.166		
	Huynh-Feldt	2603.289	44.000	59.166		
	Lower-bound	2603.289	44.000	59.166		
Stress_location	Sphericity Assumed	47.350	1	47.350	1.705	0.198
	Greenhouse-Geisser	47.350	1.000	47.350	1.705	0.198
	Huynh-Feldt	47.350	1.000	47.350	1.705	0.198
	Lower-bound	47.350	1.000	47.350	1.705	0.198
Stress_location * L1	Sphericity Assumed	78.550	1	78.550	2.828	0.100
	Greenhouse-Geisser	78.550	1.000	78.550	2.828	0.100
	Huynh-Feldt	78.550	1.000	78.550	2.828	0.100
	Lower-bound	78.550	1.000	78.550	2.828	0.100
Error(Stress_location)	Sphericity Assumed	1222.174	44	27.777		
	Greenhouse-Geisser	1222.174	44.000	27.777		
	Huynh-Feldt	1222.174	44.000	27.777		
	Lower-bound	1222.174	44.000	27.777		
Sentence_type * Stress_location	Sphericity Assumed	0.242	1	0.242	0.008	0.928
	Greenhouse-Geisser	0.242	1.000	0.242	0.008	0.928
	Huynh-Feldt	0.242	1.000	0.242	0.008	0.928
	Lower-bound	0.242	1.000	0.242	0.008	0.928
Sentence_type * Stress_location * L1	Sphericity Assumed	0.000	1	0.000	0.000	0.998
	Greenhouse-Geisser	0.000	1.000	0.000	0.000	0.998
	Huynh-Feldt	0.000	1.000	0.000	0.000	0.998
	Lower-bound	0.000	1.000	0.000	0.000	0.998
Error(Sentence_type*Stress_location)	Sphericity Assumed	1273.911	44	28.953		
	Greenhouse-Geisser	1273.911	44.000	28.953		
	Huynh-Feldt	1273.911	44.000	28.953		
	Lower-bound	1273.911	44.000	28.953		

Table 39: Tests of Within-Subjects Effects, Dependent variable: Last Click

5.2.3.4 Age of arrival effect

It was found that the age of arrival is negatively correlated to the perception score and cloze test score of L2 English speakers, as seen below. The slope of this variable is significant in a linear regression model. It would be a natural next step to ask how the age of arrival affects the stress perception accuracy in the L2 English speakers group, as in research question 3. 'Does early arrival bestow an advantage on L2 speakers? Would L2 speakers who moved to the United States at a young age perform better at

the stress matching task than people who moved to the United States as a later age?' The prediction would be that people with earlier age of arrival should perform better at stress perception task and that the coefficient on age of arrival factor would also be negative in a linear model. While there is a negative correlation between the age of arrival and the stress perception score, the finding suggests that relationship is not statistically significant (p = 0.313). This suggests that stress perception might be particularly difficult and resistant to language acquisition even at younger age of arrival.

Listing 6: Age of arrival as predictor of perception score

```
lm(formula = PERCEPTION.SCORE ~ AGE.ARRIVAL, data = vi_data)

Residuals:
    Min      1Q  Median      3Q     Max
-17.030  -1.249   1.476   3.824  11.127

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   52.4899     4.8364  10.853  2.5e-09 ***
AGE.ARRIVAL   -0.5730     0.2217  -2.584   0.0187 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.673 on 18 degrees of freedom
Multiple R-squared:  0.2706,    Adjusted R-squared:  0.2301
F-statistic:  6.68 on 1 and 18 DF,  p-value: 0.0187
```

Listing 7: Age of arrival as predictor of cloze test score

```
lm(formula = CLOZE.TEST ~ AGE.ARRIVAL, data = vi_data)

Residuals:
    Min      1Q  Median      3Q     Max
-29.573  -1.894   1.134   5.704  13.283

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   51.7337     6.9584   7.435 6.85e-07 ***
AGE.ARRIVAL   -0.7149     0.3190  -2.241   0.0379 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.6 on 18 degrees of freedom
Multiple R-squared:  0.2182,    Adjusted R-squared:  0.1747
F-statistic: 5.022 on 1 and 18 DF,  p-value: 0.03787
```

Listing 8: Age of arrival as predictor of stress perception score

```
lm(formula = vi$STRESS.PERCEPTION.SCORE ~ vi$AGE.OF.ARRIVAL,
    data = vi)

Residuals:
    Min      1Q  Median      3Q     Max
-5.5218 -1.2028  0.8613  2.1190  3.5317

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        16.43737    2.01160   8.171 1.81e-07 ***
vi$AGE.OF.ARRIVAL -0.09578    0.09222  -1.039    0.313
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.775 on 18 degrees of freedom
Multiple R-squared:  0.05654,    Adjusted R-squared:  0.004126
F-statistic: 1.079 on 1 and 18 DF,  p-value: 0.3127
```

6 Discussion

Findings about stress perception between the L1 English and L1 Vietnamese groups validated our hypothesis and prediction at the start of the paper. The result of a stress matching task showed significant differences in how speakers of different L1 approached the task of perceiving stress. Based on the previous literature and what stress and tone's acoustic correlates are, a probable explanation to this difference in performance is in how L1 English speakers use acoustic cues for stress differently compared to the L2 English learners' group. Due to what we know about acoustic correlates of stress for perception and acoustic correlates of tones for perception, namely the correlates of stress in English include intensity or amplitude (loudness), length (duration), F0 (pitch), and vowel formants (vowel quality) and the correlates of tone in Vietnamese include F0, length (duration), amplitude (intensity), and voice quality, we posited that the two correlates that are characteristic of tones in Vietnamese and that also feature in stress, F0 and duration, were likely to have been used as cues for L1 Vietnamese L2 English speakers in detecting stress. Interestingly, how F0 is used as a correlate in detecting stress was found to differ between the two groups and this is likely to give us the explanation of why we saw a difference in performance between the two groups. Previous studies suggested that L1 English speakers would listen to more cues other than F0, such as vowel quality, and changes in F0, and the entire intonation contour of an utterance at once to determine the stress position of a word. This experiment tried to remove the effect of vowel reduction by making sure that the recordings did not contain an obvious extent of vowel reduction. That leaves the L1 English speakers with the changes in F0 and the entire intonation contour, as well as loudness, duration, and intensity as cues in detecting stress. How these speakers rank these cues remain a question that not many studies have addressed, but what is interesting in L1 English speakers' approach towards stress identification, as pointed out by (Ou, 2010), is that the entire intonation contour would be under consideration, rather than only abrupt changes in F0. Timing in such intonation contours matters too, for instance, in bisyllabic words, when the second syllable is stressed, this syllable has a low rising pitch contour, when the second syllable is unstressed, it has a high rising pitch contour' [**?**], because the word with initial stress would have a rising contour that starts earlier than the word with final stress. Such cues are used by the L1 English speakers, but it is doubtful that the L2 English speakers were aware of such correlate, when listening to stressed syllable in different intonation contours.

The issue of using the same linguistic cues differently giving rise to different performances and representation about language is not a new one in linguistics. This was seen again and again, as in language pairs that share the same segmental phonetic unit in their sound inventory, but employ this unit differently. The phoneme - allophone contrast explained why speakers of English would have a difficult time producing or listening to minimal pairs in Vietnamese where the voiceless alveolar stop is contrastive by aspiration at word-initial position. This is also in line with the 'categorical perception' concept, people are good at distinguishing phonological units that are contrastive in their language, that are tied to a difference in meaning, while considering other non-contrastive features as noise and are much less sensitive to those differences. As Vietnamese speakers use F0 predominantly as a correlate for tones, F0 change is what the L1 Vietnamese group was listening for, while other cues such as the entire intonation contour and intensity were not picked up on.

The reaction time results yielded some interesting findings and more insights beyond what the research question and the hypothesis originally set up to explore. From a superficial look at the descriptive statistics, it seems that L1 English group tends to take longer than L2 English group to respond to stress perception questions, although no statistical significance was reached for this difference. Interestingly, a mixed effect ANOVA analysis showed that there was a difference between the two groups in the mean click count per question, with the L1 Vietnamese group more likely to have higher mean click count per question. The within-subjects effects analysis suggested that there is an interaction between the stress location factor and the L1 factor. A followed up independent samples t-test showed that this difference was due to the word-final stress position. In other words, L1 Vietnamese group was found to have statistically significant higher mean click count per question, in particular in respond to the stimuli with word-final stress. This finding is interesting and it suggests there might be a preference for word-initial stress in this group and that is why word-final stress had higher mean click count. This is interesting

and slightly unexpected, because after the discusion about the difficulty in perceiving stress in the word initial question condition, the expectation would be that the higher mean click count would be due to a word-initial stress factor, not a word-final stress factor. This finding opens new questions for further investigation, such as about the bias or preference for word-intial stress among the L1 Vietnamese group. Would there be such a preference, and if so, why? What are the native preferences in Vietnamese, and is there an existing preference for word-initial stress inherently in the L1?

The first click's within subjects effects analysis showed that both sentence type and stress location were significant factors in the difference in the first click behavior among the two groups. All subjects had shorter first click time for the statement condition compared to the word-initial stress position, and all subjects had longer first click time for the question condition compared to the word-final stress position. Therefore, it seems that the statement condition is the most straightforward condition for all subjects, resulting in faster decision time, and the question and word-initial condition were relatively more challenging for all subjects. This is in line with our findings that the word-initial and question conditions would be challenging, but in this case, it was established that the L1 English group also experienced longer first click time for these conditions. This finding is certainly out of the scope of what the hypothesis set out to discover. It seems that L1 English speakers also found that the word-initial and question conditions to be more challenging than the statement condition. It could be explained that more cues needed to be extracted from the speech signal for the subjects to make a matching decision, thus making the task more challenging.

An interesting development of this paper in the future would be to incorporate acoustic measurements of the stressed syllable into the statistical analysis, in order to get at a more fundamental relationship in the process of perceiving stress: what are the F0 values of the stressed syllables and how L2 speakers would respond to those values, or what are the duration values and how L2 speakers would respond to those values. Perhaps a pattern would emerge and we would see more clearly if there is a relationship between average F0 values and how L2 English speakers assign stress for each stimulus. The acoustic analysis presented in the previous section would be a good starting point for such a discussion.

Another development to the study that would provide more insights to how L1 Vietnamese speakers perceive stress is looking at stress to tone mappings of the participants. Previous studies that have looked at syllable shape and tone in the L1 interacting with stress perception and production include studies in Thai and Vietnamese. Jangjamras (2011) hypothesized that Thai native speakers would not have much difficulty perceiving English lexical stress because they would rely on pitch and vowel duration as perceptual cues, at the same time, they would have difficulty producing stress on syllables of the shape CVVO due to the restrictions of tone assignment on this type of syllables in the L1. The study used English nonce words as stimuli for both production and perception tasks. The production task's result showed that syllable shape affected how well Thai speakers were able to produce stress while the perception task's result showed that Thai speakers performed as well as American English native speakers on perception and both groups had more difficulty identifying final stress compared to initial stress, which was contradictory to the expectation that Thai speakers would perform better on final stress prediction because Thai exhibits fixed final stress in polysyllabic words.

The stress to tone assignment question would be motivated by a restriction of syllable shape and tone assignment in Vietnamese, whereby closed syllables ending in voiceless stop codas can only be assigned either a rising or a falling tone. If there is indeed a match in perception between stressed syllable with higher pitch or rising tone and unstressed syllable with lower pitch or falling tone, then the tone assignment of stressed and unstressed syllables in closed syllables could be expected to pattern with the perceived stress. Codas in Vietnamese can only be nasals and voiceless stop consonants. Therefore, loan words from English necessarily undergo adaptation of codas. This process primarily involves merging of a richer set of coda classes to a more limited set of coda classes available in Vietnamese phonotactics, where voiceless and voiced stops, fricatives, and affricates, in other words, obstruents codas, are all mapped to voiceless stops, nasals are mapped to nasals, liquids to voiceless stops or dropped, and glides are often dropped. Liquid codas such as /l/ and /r/ are likely to be dropped, unless orthographical bias would induce an adaptation similar to the French uvular fricative to Vietnamese voiceless velar stop (Kang, Pham, and Storme, 2014) [**?**]. Because of these patterns, the research question for a future study should make a distinction between closed syllables with obstruent codas where there will be a mapping to Vietnamese voiceless stops, and the rest, where the coda stays as nasal or dropped, neither of which obeys the tone assignment restriction.

The prediction for such an investigation would be as follows: L1 Vietnamese L2 English speakers would be predicted to assign the MR tone to the stressed syllable and MFC tone to the unstressed syllable of closed syllables with obstruent codas, and they are predicted to assign a rising or high tone

on the stressed syllable and a falling tone to the unstressed syllable for other syllable shapes, as shown by the predicted Vietnamese transcriptions in Table 41.

	Closed syllables with obstruent codas	Other syllable shapes
Stressed	MASsif - *mát síp* - MR tone	COMcave - *com / cóm cây* - ML or MR tone
Unstressed	masSIF *ma.t síp* - MFC tone	comCAVE - *còm cây* - MF tone

Table 41: Transcription task's expectation

Regarding this study's third research question, our analysis about the age of arrival and effect on stress perception were certainly intriguing. While the age of arrival had a significant effect on other test scores such as the cloze test score, the perception task score as a whole, which includes performance on the fillers, there was no evidence that the age of arrival had an impact on the stress perception task. This finding suggests that stress perception is a very difficult task for L2 English speakers to catch up with L1 English speakers, especially in a language pair where the same acoustic cue would be utilized differently across the two groups. It would be harder to unlearn a previous signal from the L1 and acquiring a new signal, and the L1 transfer effect have been shown to be very robust for the process of stress perception in the Vietnamese-English language pair.

7 Conclusion

This study started out with a goal to research into the difference in stress perception between the L1 English speakers and the L1 Vietnamese L2 English speakers. The findings agreed with the hypotheses and prediction this paper presented at the beginning, and the effects were very clear. The reaction times also strengthened the argument that test conditions such as question and word-initial stress location were more challenging than the statement condition. This study opens doors to a multitude of questions, some of which have been raised previously in the Discussion section. The study was motivated by the study about Taiwanese EFL learners' perception of stress (Ou, 2010), and the result confirms the hypothesis that L1 Vietnamese L2 English speakers found it more difficult to identify stress when the cue of pitch is manipulated by contexts, especially in the word-intial question test condition. A bigger question would be how robust this hypothesis is when extended to other tonal languages that use F0 as a perceptual cue in similar ways? What about speakers of pitch-accent languages? Would they exhibit the same difficulty compared to tonal native speakers? More generalizations and insights could be discovered as more experiments and data were to be collected from speakers of other tonal languages as we increase our understanding about how speakers utilize perceptual cues cross-linguistically.

Acknowledgement

I would like to express my appreciation for the helpful feedback and and suggestions from Professor Tania Ionin, Professor Chilin Shih, and Professor Jeff Green. Your comments and inputs helped improve this paper tremendously. I am thankful for UIUC Department of Linguistics for supporting this study with the Human Subject Funding Award. I would also like to thank the participants of this study for their time and valuable responses, without which analyses and findings would not have been possible.

A Appendix

A.1 Queried result - Bisyllabic words

['effect', 'foretaste', 'likud's', 'produce', 'siam', 'amman', 'reset', 'desert', 'digests', 'resets', 'katyn', 'concerts', 'firsthand', 'reject', 'aigner', 'beacham', 'decade', 'jacquet', 'emerged', 'mme', 'quemoy', 'seguin', 'costumes', 'object', 'rejects', 'compound', 'furlett', 'transport', 'canucks', 'excise', 'saddam's', 'verdon', 'eugene', 'transfers', 'rigueur', 'convict', 'moshe's', 'refill', 'moray', 'perrault', 'minot', 'elkind', 'catain', 'madame', 'rocard', 'impact', 'canuck', 'concert', 'transferred', 'erode', 'miro's', 'compounds',

‘foretastes’, ‘sayiid’, ‘contrasts’, ‘rupees’, ‘oblate’, ‘bourgeois’, ‘petard’, ‘pasha’, ‘clarisse’, ‘lavie’, ‘tamil’, ‘supine’, ‘complex’, ‘detour’, ‘ravel’, ‘discount’, ‘gaubert’s’, ‘subject’, ‘todays’, ‘concave’, ‘chemins’, ‘escort’, ‘alum’, ‘conflict’, ‘davao’, ‘protract’, ‘impacts’, ‘traverse’, ‘obit’, ‘begun’, ‘inclines’, ‘detail’, ‘yourselves’, ‘buffet’, ‘rebel’, ‘brasil’, ‘levin’, ‘concrete’, ‘fatah’, ‘sunscreen’, ‘caches’, ‘annexed’, ‘banshee’, ‘batiks’, ‘barnard’, ‘debuted’, ‘imprint’, ‘decades’, ‘golan’, ‘constructs’, ‘benet’, ‘ferment’, ‘isaak’, ‘sistine’, ‘markel’, ‘chauffeurs’, ‘erupt’, ‘romance’, ‘levin’s’, ‘ghafar’, ‘sharon’, ‘combine’, ‘allies’, ‘digest’, ‘project’, ‘converts’, ‘scurdell’, ‘deyton’, ‘mahmoud’, ‘incline’, ‘intrigued’, ‘kilcrease’, ‘emerge’, ‘maurice’, ‘research’, ‘capri’, ‘inlaws’, ‘abbe’, ‘import’, ‘gaubert’, ‘conduct’, ‘content’, ‘chauffeur’s’, ‘affix’, ‘ines’, ‘bangkok’, ‘tamils’, ‘escrow’, ‘messrs.’, ‘doiron’, ‘convert’, ‘kadar’, ‘narrates’, ‘tabak’, ‘defects’, ‘legit’, ‘yourself’, ‘contest’, ‘today’s’, ‘insides’, ‘present’, ‘pogrom’, ‘mahmud’, ‘akins’, ‘morass’, ‘sabir’, ‘transfer’, ‘costume’, ‘defect’, ‘effects’, ‘fertile’, ‘record’, ‘sunscreens’, ‘adults”, ‘detours’, ‘participants’, ‘impulse’, ‘communes’, ‘ally’, ‘michel’, ‘converse’, ‘sharon’s’, ‘mistry’, ‘deserts’, ‘perverts’, ‘saddam’, ‘morays’, ‘adult’, ‘compacts’, ‘convicts’, ‘overt’, ‘abend’, ‘debut’, ‘traversed’, ‘construct’, ‘combat’, ‘marsal’, ‘chavez’, ‘thibert’, ‘pervert’, ‘bernard’, ‘commune’, ‘likud’, ‘ijaz’, ‘ahlen’, ‘massif’, ‘sinclair’, ‘into’, ‘baton’, ‘jerrell’, ‘mendez’, ‘kanell’, ‘gamete’, ‘miro’, ‘inbound’, ‘chauffeur’, ‘natal’, ‘objects’, ‘presents’, ‘obits’, ‘recess’, ‘redress’, ‘dabah’, ‘adisq’, ‘compress’, ‘console’, ‘contests’, ‘cannot’, ‘intrigue’, ‘moshe’, ‘mathilde’, ‘transports’, ‘hashish’, ‘records’, ‘projects’, ‘curie’, ‘adults’, ‘chauffeured’, ‘syringe’, ‘olah’, ‘butane’, ‘hostile’, ‘kvamme’, ‘intrigues’, ‘oblak’, ‘details’, ‘islam’, ‘savir’, ‘veilleux’, ‘imports’, ‘capri’s’, ‘pogroms’, ‘minute’, ‘lahaie’, ‘excerpts’, ‘conflicts’, ‘akin’, ‘contrast’, ‘refills’, ‘compact’, ‘excerpt’, ‘buffets’, ‘carmel’, ‘decor’, ‘contents’, ‘benet’s’, ‘benzene’, ‘rebels’, ‘transform’, ‘regress’, ‘mahmood’]

A.2 Queried result - Trisyllabic words

[‘liliane’, ‘redifer’, ‘versatile’]

A.3 Stimuli - Bisyllabic words

A.3.1 Category 1: stress on the first syllable in a statement context

1. He is a BENnet
2. He is a GAUbert
3. He is a POKgrom
4. He is a TAEbak
5. It is a MIro
6. She is a SAEvir
7. She is a ZAAbir
8. That is a KANtar
9. That is a MOray
10. That is a MOshee
11. That is a PANshee
12. That is a TANvo
13. That is an OBlak
14. She is my ANna
15. This is a PEEnan
16. He is my DANbah

A.3.2 Category 2: stress on the first syllable in a yes/no question context

1. Is he a BENnet?
2. Is he a GAUbert?
3. Is he a POKgrom?
4. Is he a TAEbak?
5. Is it a MIro?
6. Is she a SAEvir?
7. Is she a ZAAbir?
8. Is that a KANtar?
9. Is that a MOray?
10. Is that a MOshee?
11. Is that a PANshee?
12. Is that a TANvo?
13. Is that an OBlak?
14. She is my ANna?
15. Is this a PEEnan?
16. Is he my DANbah?

A.3.3 Category 3: stress on the final syllable in a statement context

1. He is a benNET
2. He is a gauBERT
3. He is a pokGROM
4. He is a taeBAK
5. It is a miRO
6. She is a saeVIR
7. She is a zaaBIR
8. That is a kanTAR
9. That is a moRAY
10. That is a moSHEE
11. That is a panSHEE
12. That is a tanVO
13. That is an obLAK
14. She is my anNA
15. This is a peeNAN
16. He is my danBAH

A.3.4 Category 4: stress on the final syllable in a yes/no question context

1. Is he a benNET?
2. Is he a gauBERT?
3. Is he a pokGROM?
4. Is he a taeBAK?
5. Is it a miRO?
6. Is she a saeVIR?
7. Is she a zaaBIR?
8. Is that a kanTAR?
9. Is that a moRAY?
10. Is that a moSHEE?
11. Is that a panSHEE?
12. Is that a tanVO?
13. Is that an obLAK?
14. She is my anNA?
15. Is this a peeNAN?
16. Is he my danBAH?

A.4 Stimuli - Trisyllabic words

A.4.1 Category 1: stress on the first syllable in a statement context

1. That is a YOkotaa
2. That is my ROseemund
3. That is a PEnoquin
4. That is a CAseybeer
5. He is my ANfelar
6. That is a VItaely

A.4.2 Category 2: stress on the first syllable in a yes/no question context

1. Is that a YOkotaa?
2. Is that a VItaely?
3. Is that my ROseemund?
4. Is that a PEnoquin?
5. Is that a CAseybeer?
6. Is he my ANfelar?

A.4.3 Category 3: stress on the middle syllable in a statement context

1. That is a yoKOtaa
2. That is my roSEEmund
3. That is a peNOquin
4. That is a caSEYbeer
5. He is my anFElar
6. That is a viTAEly

A.4.4 Category 4: stress on the middle syllable in a yes/no question context

1. Is that a yoKOtaa?
2. Is that a viTAEly?
3. Is that my roSEEmund?
4. Is that a peNOquin?
5. Is that a caSEYbeer?
6. Is he my anFElar?

A.4.5 Category 5: stress on the final syllable in a statement context

1. That is a yokoTAA
2. That is my roseeMUND
3. That is a penoQUIN
4. That is a caseyBEER
5. He is my anfeLAR
6. That is a vitaeLY

A.4.6 Category 6: stress on the final syllable in a yes/no question context

1. Is that a yokoTAA?
2. Is that a vitaeLY?
3. Is that my roseeMUND?
4. Is that a penoQUIN?
5. Is that a caseyBEER?
6. Is he my anfeLAR?

A.5 Fillers

1. obit — opit
2. rocard — rocards
3. saron — sharon
4. sequin — sequins
5. concave — comcave
6. nahmuh — mahmuh
7. davao — dafao
8. chauffeur — chauffeured
9. capri — cabri
10. oma — uma
11. covert — overt
12. versatile — verzatile
13. buffet — buffets
14. izaak — isaak
15. alan — ahlen
16. kolan — golan
17. bath — bathe
18. detour — detours
19. akins — akin
20. sayiid — zayiid
21. speak — spoke
22. stalk — stalked
23. clock — cloak
24. tirade — tirades
25. spring — spurring
26. surmount — seamount
27. repair — repaired
28. wonder — wander
29. promoter — promoted
30. caught — cot
31. close (v) — close (adj)
32. live (v) — live (adj)
33. loose — lose
34. together — altogether
35. flowers — flower
36. polar — bolar
37. feam — fim
38. dispose — despise

A.6 Full test

A.6.1 Block 1

1. obit — opit
2. rocard — rocards
3. saron — sharon
4. sequin — sequins
5. tanVO (question)
6. concave — comcave
7. nahmuh — mahmuh
8. BENnet (question)
9. davao — dafao
10. chauffeur — chauffeured
11. ANfelar (statement)
12. TAEbak (statement)
13. capri — cabri
14. oma — uma
15. moRAY (statement)

A.6.2 Block 2

1. pokGROM (question)
2. covert — overt
3. gauBERT (statement)
4. ZAAbir (statement)
5. PEEnan (question)
6. versatile — verzatile
7. buffet — buffets
8. izaak — isaak
9. alan — ahlen
10. peNOquin (statement)
11. kolan — golan
12. bath — bathe
13. detour — detours
14. akins — akin
15. sayiid — zayiid

A.6.3 Block 3

1. miRO (statement)
2. speak — spoke
3. ANna (statement)
4. stalk — stalked
5. clock — cloak
6. ROseemund (question)
7. tirade — tirades
8. spring — spurring
9. surmount — seamount
10. repair — repaired
11. MOshee (question)
12. saeVIR (question)
13. wonder — wander
14. promoter — promoted
15. vitaeLY (statement)

A.6.4 Block 4

1. caught — cot
2. close (v) — close (adj)
3. danBAH (question)
4. OBlak (statement)
5. live (v) — live (adj)
6. loose — lose
7. together — altogether
8. flowers — flower
9. polar — bolar
10. caSEYbeer (question)
11. feam — fim
12. dispose — despise
13. yokoTAA (question)
14. panSHEE (statement)
15. KANtar (question)

A.7 English cloze test

20 days volunteering ____ Tien Son Field Hospital

Working at the field hospital, Linh understood that this is the place where Covid-19 patients are treated, that she would be exposed to F0, and the risk of ____ would be higher than that in contact ____ work. In early August, Vo Quang Linh, a 3rd year student ____ in General Nursing at Da Nang University of Medical Technology and Pharmacy was transferred ____ Tien Son field hospital. Before ____, he belonged to a group of volunteers doing contact ____ in Thanh Khe district, but due to few infected people and a large number of ____, he was transferred to a more ____ area.

Besides Linh, about 200 students from other universities also ____ to support. Tien Son field hospital was ____ on August 1. It has two basements and four floating floors, ____ an area of more than 10,000 square meters. This is the second field hospital that was established, after the ____ facility at Hoa Vang District Medical Center, and it has a ____ of 200 beds. Depending on the ______ of the epidemic, the field hospital can ____ the maximum capacity to over 700 beds. In the same group ____ Linh is Vu Thi Van, her close friend, who is training at Tien Son. Previously, Van's group also did contact ____, but the group was put on ____, pending a new transfer order. Waiting for more than a week, Van became _____ as she watched her friends in Hai Chau and Hoa Vang toiled _____. Their work never ____ while she was just sitting ____ home. She wanted to ____ but was not given the _____. The first day in Tien Son, Linh, Van and the group of volunteers took ____ to observe to understand the ____ to care for the patient. Not knowing the specific work, no one could hide their ____ because this is where a Covid-19 patient ____ be treated. "Many longtime physicians are at ____ of infection, and I do not have much experience", Van thought. During the training session, ____ were divided ____ groups. Each team has one doctor, one nurse and two volunteers to take ____ of one to four hospital beds in a _____. The main job is to monitor the patient, under the ____ of medical doctors; deliver medicine with the nurses, ration food, supplies for the sick; help patients ____ through secure buildings; and make sure there is no cross-____ in the hospital. Doctors also ____ some hypothetical ____ such as fire, power outage, or ____ of the patients so that students can practice their skills. At the end of the session, the groups will give ____ to each other to improve each day. Van said that the new infection ____ is a challenge but also an ____ for medical students to gain more ____. Nearly 20 days after ____ the assignment, Linh and Van and a group of volunteers are improving their skills ____ treating patients with ____ diseases. Although the work is repetitive, it is not ____, and everyone is ____ to contribute and do work that ____ the community.

A.8 Vietnamese cloze test

Á quân Đường lên đỉnh Olympia 2020 Vũ Quốc Anh đang khiến cộng _____ mạng "phát sốt" với cử chỉ dung dị của mình trên sàn đấu đêm chung kết.

Trước khi bước vào phần thi Về đích, Vũ Quốc Anh bỗng xin phép trường _____ cho vài giây.

Em cúi xuống rồi bước ra với chiếc quần xắn cao để _____ đôi tất màu xanh.

Những tưởng chỉ muốn chỉnh lại trang _____ cho lịch sự, Quốc Anh lại khiến mọi người rất bất ngờ.

Đôi tất màu xanh của Quốc Anh _____ cho cả trường quay phải lắng đọng.

Quốc Anh sau đó đã bật _____ khi gửi lời cảm ơn đến gia đình, thầy cô, bạn bè ở vùng đất đỏ Tây Nguyên.

"Từ nhỏ em đã xem Olympia và đó đã trở thành một phần cuộc sống của em, Olympia đã thay đổi mã gen của em rồi", Quốc Anh sụt _____chia sẻ.

Được biết, trong ngày thi chung kết, chỉ có mỗi mẹ Quốc Anh bay từ Đắk Lắk ra Hà Nội để cổ _____ con trai, riêng cha em thì ở nhà tiếp sức từ xa.

Có lẽ vì thế, chàng á _____ muốn dành cho cha mình một lời cảm ơn thật đặc biệt trước khi bước vào vòng thi đầy gây cấn.

Hành động này của Quốc Anh khiến khán giả trường quay và cộng đồng mạng vô cùng xúc _____.

Vốn lạnh lùng và quyết _____ qua từng phần thi, chàng á quân lại gây bất ngờ với cử chỉ dung dị, đong đầy tình cảm dành cho gia đình.

Một người dùng Facebook xúc động chia sẻ: "Mình đã khóc khi nhìn thấy hành động xắn quần để lộ đôi tất của ba tặng cho Quốc Anh, mình rất tiếc khi chiếc vòng Nguyệt _____ đó lại thuộc vào tay người khác chứ không phải bạn ấy."

Quốc Anh cũng đã rất bản lĩnh và quyết tâm để dành quyền trả lời câu hỏi, sự nỗ lực ấy dù không được đền đáp xứng đáng, nhưng đã để lại trong lòng mình cũng như tất cả mọi người một sự ngưỡng mộ và khâm _____ đáng nể.

Mong rằng tương lai của bạn sẽ chinh phục được nhiều _____ núi cao hơn".

Figure 10: Cloze test for Vietnamese native speakers

A.9 Python script to create stimuli

```
#!/usr/bin/python3

# Script to process sound files.

import fnmatch
import glob
import os
from collections import defaultdict
from pydub import AudioSegment

# Create a dictionary of each item: a list of its filenames.
def create_files_dict(dir):
    path = os.getcwd()
    ext = ".wav"
    files = fnmatch.filter(os.listdir(path), '*' + ext)
    filesdict = defaultdict(list)
    for file in files:
        if file.split("_")[1].isdigit():
            key = file.split("_")[2]
```

```
        else:
            key = file.split("_")[1]
        filesdict[key].append(file)
    return filesdict

# Create the test (different in different test lists)
def createTestItems(filesdict):
    # Create test items
    sil = AudioSegment.silent(duration=1500)
    for val in filesdict.values():
        for v in val:
            if "_t_" in v or "_d_" in v or "_q_" in v:
                testsound = AudioSegment.from_wav(v)
                toReturn = testsound + sil +
                testsound + sil + testsound
                # Note: remember to change the test list's code below:
                toReturn.export("TESTA1-" + v, format="wav")
            else:
                pass

# Create stimuli (these should be the same across different test lists)
def createStimuli(filesdict):
    shortsil = AudioSegment.silent(duration=1000)
    longsil = AudioSegment.silent(duration=2000)
    for k in filesdict.keys():
        # Two-syllable words. Presentation order: 10 and then 01 (or 12)
        if (k == "taebak" or k == "moray" or
         k == "zaabir" or k == "gaubert" or
         k == "anna" or k == "miro" or
         k == "oblak" or k == "panshee" or
         k == "davao" or k == "chauffeured" or
         k == "sequins" or k == "capri" or
         k == "rocard" or k == "izaak" or
         k == "akin" or k == "detour" or k == "covert" or
         k == "versatile" or k == "surmount" or
         k == "wander" or k == "cloak" or
         k == "tirade" or k == "promoted" or
         k == "cot" or k == "feam" or k == "flowers" or
         k == "together" or k == "live"):
            for fn in filesdict.get(k):
                if "_s_1" in fn:
                    s1 = AudioSegment.from_wav(fn)
                elif "_s_2" in fn:
                    s2 = AudioSegment.from_wav(fn)
            sound = s1 + shortsil + s2 + longsil + s1
             + shortsil + s2 + longsil + s1 + shortsil + s2
            sound.export("STIMULI-" + k + ".wav", format="wav")
        # Two-syllable words. Presentation order: 01 and then 10 (or 21)
        elif (k == "bennet" or k == "tanvo" or
         k == "peenan" or k == "pokgrom" or
          k == "moshee" or k == "saevir" or
           k == "kantar" or k == "danbah" or
            k == "oma" or k == "mahmud" or
             k == "concave" or k == "obit" or
              k == "sharon" or k == "sayiid" or
               k == "buffet" or k == "alan" or
                k == "kolan" or k == "bath" or
                 k == "speak" or k == "spring" or
```

```
78                  k == "repaired" or k == "stalked" or
79                   k == "close" or k == "dispose" or
80                    k == "polar" or k == "loose"):
81              for fn in filesdict.get(k):
82                  if "_s_1" in fn:
83                      s1 = AudioSegment.from_wav(fn)
84                  elif "_s_2" in fn:
85                      s2 = AudioSegment.from_wav(fn)
86              sound = s2 + shortsil + s1 + longsil +
87               s2 + shortsil + s1 + longsil +
88                s2 + shortsil + s1
89              sound.export("STIMULI-" + k + ".wav", format="wav")
90          # Three syllable word. Presentation order: 123
91          elif k == "anfelar":
92              for fn in filesdict.get(k):
93                  if "_s_1" in fn:
94                      s1 = AudioSegment.from_wav(fn)
95                  elif "_s_2" in fn:
96                      s2 = AudioSegment.from_wav(fn)
97                  elif "_s_3" in fn:
98                      s3 = AudioSegment.from_wav(fn)
99              sound = s1 + shortsil + s2 + shortsil + s3 +
100              longsil + s1 + shortsil + s2 + shortsil + s3 +
101               longsil + s1 + shortsil + s2 + shortsil + s3
102             sound.export("STIMULI-" + k + ".wav", format="wav")
103         # Three syllable word. Presentation order: 132
104         elif k == "penoquin":
105             for fn in filesdict.get(k):
106                 if "_s_1" in fn:
107                     s1 = AudioSegment.from_wav(fn)
108                 elif "_s_2" in fn:
109                     s2 = AudioSegment.from_wav(fn)
110                 elif "_s_3" in fn:
111                     s3 = AudioSegment.from_wav(fn)
112             sound = s1 + shortsil + s3 + shortsil + s2 + longsil +
113              s1 + shortsil + s3 + shortsil + s2 + longsil +
114               s1 + shortsil + s3 + shortsil + s2
115             sound.export("STIMULI-" + k + ".wav", format="wav")
116         # Three syllable word. Presentation order: 213
117         elif k == "vitaely":
118             for fn in filesdict.get(k):
119                 if "_s_1" in fn:
120                     s1 = AudioSegment.from_wav(fn)
121                 elif "_s_2" in fn:
122                     s2 = AudioSegment.from_wav(fn)
123                 elif "_s_3" in fn:
124                     s3 = AudioSegment.from_wav(fn)
125             sound = s2 + shortsil + s1 + shortsil + s3 + longsil +
126              s2 + shortsil + s1 + shortsil + s3 + longsil +
127               s2 + shortsil + s1 + shortsil + s3
128             sound.export("STIMULI-" + k + ".wav", format="wav")
129         # Three syllable word. Presentation order: 213
130         elif k == "vitaely":
131             for fn in filesdict.get(k):
132                 if "_s_1" in fn:
133                     s1 = AudioSegment.from_wav(fn)
134                 elif "_s_2" in fn:
135                     s2 = AudioSegment.from_wav(fn)
```

```
                elif "_s_3" in fn:
                    s3 = AudioSegment.from_wav(fn)
            sound = s2 + shortsil + s1 + shortsil + s3 + longsil +
             s2 + shortsil + s1 + shortsil + s3 + longsil +
              s2 + shortsil + s1 + shortsil + s3
            sound.export("STIMULI-" + k + ".wav", format="wav")
        # Three syllable word. Presentation order: 231
        elif k == "roseemund":
            for fn in filesdict.get(k):
                if "_s_1" in fn:
                    s1 = AudioSegment.from_wav(fn)
                elif "_s_2" in fn:
                    s2 = AudioSegment.from_wav(fn)
                elif "_s_3" in fn:
                    s3 = AudioSegment.from_wav(fn)
            sound = s2 + shortsil + s3 + shortsil + s1 + longsil +
             s2 + shortsil + s3 + shortsil + s1 + longsil +
              s2 + shortsil + s3 + shortsil + s1
            sound.export("STIMULI-" + k + ".wav", format="wav")
        # Three syllable word. Presentation order: 312
        elif k == "caseybeer":
            for fn in filesdict.get(k):
                if "_s_1" in fn:
                    s1 = AudioSegment.from_wav(fn)
                elif "_s_2" in fn:
                    s2 = AudioSegment.from_wav(fn)
                elif "_s_3" in fn:
                    s3 = AudioSegment.from_wav(fn)
            sound = s3 + shortsil + s1 + shortsil + s2 + longsil +
             s3 + shortsil + s1 + shortsil + s2 + longsil +
              s3 + shortsil + s1 + shortsil + s2
            sound.export("STIMULI-" + k + ".wav", format="wav")
        # Three syllable word. Presentation order: 321
        elif k == "yokota":
            for fn in filesdict.get(k):
                if "_s_1" in fn:
                    s1 = AudioSegment.from_wav(fn)
                elif "_s_2" in fn:
                    s2 = AudioSegment.from_wav(fn)
                elif "_s_3" in fn:
                    s3 = AudioSegment.from_wav(fn)
            sound = s3 + shortsil + s2 + shortsil + s1 + longsil +
             s3 + shortsil + s2 + shortsil + s1 + longsil +
              s3 + shortsil + s2 + shortsil + s1
            sound.export("STIMULI-" + k + ".wav", format="wav")
        else:
            pass

if __name__ == "__main__":
    currentdir = os.getcwd()
    filesdict = create_files_dict(currentdir)
    createTestItems(filesdict)
    createStimuli(filesdict)
```

A.10 Praat script for acoustic analysis

```
# Script to query the duration and F0 values of stressed syllables.

writeInfoLine: "DURATION and F0 values of stress syllables"
```

```
4
5  selectObject: "Sound TwoSyllables", "TextGrid TwoSyllables"
6  # Query the stressed syllable in the word-initial position.
7  Extract intervals where: 2, "yes", "contains", "10"
8  sounds# = selected# ("Sound")
9  for i from 1 to size (sounds#)
10     selectObject: sounds# [i]
11     name$ = selected$ ("Sound")
12     To Pitch: 0.0, 130, 600
13     startTime = Get start time
14     endTime = Get end time
15     dur = endTime - startTime
16     startTimeBuffered = startTime + 0.025
17     endTimeBuffered = endTime - 0.025
18     stepDur = (endTimeBuffered - startTimeBuffered) / 2
19     aTimePoint = startTimeBuffered
20     bTimePoint = aTimePoint + stepDur
21     cTimePoint = bTimePoint + stepDur
22     aVal = Get value at time: aTimePoint, "Hertz", "Linear"
23     bVal = Get value at time: bTimePoint, "Hertz", "Linear"
24     cVal = Get value at time: cTimePoint, "Hertz", "Linear"
25     appendInfoLine: "_____"
26     appendInfoLine: "Stressed syl /", name$, "/", tab$, fixed$ (dur, 3), tab$, fixed$
27 endfor
28
29 appendInfoLine: "END OF REPORT"
```