Taiwan Journal of Linguistics
Vol. 21.1, 01-45, 2023
DOI: 10.6519/TJL.202301_21(1).0001

PERCEPTUAL COMPARISON OF WORD BOUNDARY SEGMENTAL CUES: ASPIRATION VS. GLOTTAL STOP*

Chiu-ching Tseng
Providence University

ABSTRACT

Previous studies on word-boundary perception in English have reported a preference for the use of the prevocalic glottal stop cue (e.g., 'seen [ʔ]ice' vs. 'see nice') over word-initial aspiration cue (e.g., 'keeps [t^h]alking' vs. 'keep s[t]alking') both by native speakers (Nakatani and Dukes 1977) and by L2 learners of various L1 backgrounds (Spanish: Altenberg 2005; Japanese: Ito and Strange 2009; French: Shoemaker 2014; Arabic: Alammar 2016). This study investigates how such phenomenon may apply in the case of Mandarin speakers, whose L1 uses stop aspiration, but not a glottal stop, contrastively. The question is whether their sensitivity to stop aspiration would help them use the cue in L2 word-boundary segmentation. The results showed that Mandarin speakers identified word boundaries more accurately when the stimuli had glottal stops than when they had aspiration stops. This outcome suggests that perceptual sensitivity to a particular acoustic cue in learners' L1 does not help them to use the cue readily in L2 perception. Both L1 and L2 groups performed significantly better with the glottal stop cue than with the aspiration cue, suggesting that the glottal stop may indeed be a universally unmarked acoustic cue for use in the task.

Keywords: speech perceptual cue, word boundary, aspiration, glottal stop, L2 word segmentation, Mandarin

* This work was supported by the Linguistics Program, English Department, George Mason University. I am extremely thankful to my advisors and all participants who kindly and willingly spared their valuable time, and to my wife, who has been my biggest fan since day one of my graduate study. All errors are my own.

1. INTRODUCTION

Speakers rely on multiple cues for word-boundary segmentation in their native languages. Cues may include lexical, semantic, syntactic, morphological, phonotactic, prosodic, rhythmic, and acoustic-phonetic information; e.g., word frequency, syllable structure, stress placement, intonation, phoneme duration, allophonic variations, and vowel harmony (Nakatani and Dukes 1977; Altenberg 2005; Ito and Strange 2009; Shoemaker 2014; Alammar 2016).

Is there a universally preferred cue for word-boundary segmentation, or do different languages prefer different cues? Altenberg (2005) reports that when listening to a continuous stream of English speech, L2 learners often found themselves unprepared to comprehend meaning, partly due to their inability to segment continuous streams into words. This viewpoint suggests that different languages rely on different cues for word-boundary segmentation, and that learners may rely on the cues preferred in their own L1. This proposition, however, could potentially interfere with boundary detection in an L2. This study explores the ability of Mandarin L2 learners of English to use the acoustic-phonetic cues for English word-boundary segmentation, specifically, those of the glottal stop and the voiceless stop aspiration.

The glottal stop is one of the most often used phonation types in world languages (Pennington 2005), but it may be absent or used differently in different languages. Garellek (2013) reports that only 47.9% of the world's languages use the glottal stop phonemically in their systems (c.f., Maddieson 1984). On the other hand, stop aspiration can be phonetic or phonemic across languages (Cho and Ladefoged 1999; Yavas 2011).

Studies have shown that glottal stops and stop aspiration both serve to mark word boundaries in English for native speakers (Nakatani and Dukes 1977; Bissiri et al. 2011). Nakatani and Dukes (1977) propose that a glottal stop at the onset of a word-initial vowel may be a cue for a boundary (see also Ito and Strange 2009). They also found that English speakers are sensitive to word-initial aspirated voiceless stops in segmenting potentially ambiguous phrases. This allophonic variation of voiceless stops plays an essential role in word-boundary segmentation, although it is not used phonemically in English. For instance, /kipstɔlkɪŋ/ (presented

in broad transcription) could be recognized as *'keeps talking'* or *'keep stalking'*[1]. Hence, the aspiration in [kips**tʰ**ɔlkɪŋ] vs. [kips**t**ɔlkɪŋ] becomes essential.

In regard to L2, studies have reported that L2 English learners rely more on prevocalic glottal stop cues to segment word boundaries than on word-initial aspiration cues (Altenberg 2005; Ito and Strange 2009; Shoemaker 2014; Alammar 2016).

Altenberg (2005) reports that when segmenting word boundaries of English speech, her Spanish participants performed, on average, at 76% accuracy, but they identified glottal stop cues much more precisely than aspiration ones (88.4% vs. 58.5%). She suggests that this discrepancy was due to L1 transfer because "… the glottal stop occurs in an emphatic speech in Spanish while aspiration does not …" (p.344).

Ito and Strange (2009) replicated Altenberg's findings with Japanese L2 learners of English. The control group (English native speakers) performed at the ceiling (96.8%), while the average accuracy of the Japanese L2 learners was 83.8%. Like the Spanish learners in Altenberg, the Japanese learners were more accurate in the case of the glottal stop cue (91.3%) than in that of the aspiration cue (73.1%). The authors suggest that L1 transfer may be the primary reason for their findings because the glottal stop can be inserted before a word-initial vowel and after a word-final vowel in an emphatic speech in Japanese (p.2350), and, stops in Japanese are generally weakly aspirated (see also Shimizu 2010).

Shoemaker (2014) reports that French does not systematically use either the glottal stop or aspiration for word-boundary segmentation, although both cues may be present in various phonological environments (p.714). For example, stop aspiration might occur in French in exaggerated exclamations, but it is not used in a systematic way (Shoemaker 2014, footnote 1). The glottal stop is also rarely exploited as a strategy for marking vocalic onset words. She states that strategies for detecting L2 word-boundary segmentation are language-specific, and that the use of these strategies is "… located in the listener, not in the signal"

[1] The realization between *'keeps talking'* and *'keep stalking'* might be dependent on other information and other factors than just stop aspiration; e.g., the maximum onset principle (Yavas 2011), lexical frequency (Shoemaker 2014), and/or vowel duration (Ito and Strange 2009), etc.

(p.711). This argument seems to suggest that what is featured in one's L1 system would be preferred for application in the task of L2 segmentation, i.e., L1 transfer (Major 2001). Shoemaker tested two groups of French English learners (differing by three years of English learning experience) and found that both groups did better with glottal stop cues than aspiration cues (84.8% vs. 59.3%). She thus proposed the universal saliency of the glottal stop for word-boundary segmentation as opposed to L1 transfer since "… French and Spanish employ glottal stop substantially less than Japanese… yet learners from both L1s are still significantly more sensitive to the presence of glottal stop than that of aspiration" (p.723).

All of the L2 learners in the previous studies predominantly preferred a glottal stop as the word-boundary cue over aspiration. The possible interpretations of these findings are:

a. Universal Markedness is in play (phonemically, the distribution of the glottal stop among the world's languages is 47.9%, and that of aspirated voiceless stop distribution is 28.7%; Maddieson 1984; Garellek 2013).
b. Full L1 transfer in L2 perception would assume that all languages studied previously prefer glottal stop as the word-boundary cue in their L1. If that is the case, then the comparison seems unfair because most of these languages do not feature aspiration; therefore, they will not transfer the feature that is not in their systems.
c. Methodological limitations: There can be a wide range of potential cues that signal the presence of a word-boundary to the listener. Therefore, control of independent variables such as lexical frequency and other potential word-boundary cues are essential (e.g., pitch-reset). Controlling for contextual information by using sliced words has been implemented in previous studies. The present study will also use sliced word stimuli and control for two other potential cues (lexical frequency and pitch contour).

The phenomenon of Universal Markedness leads us to predict that speakers of other languages would prefer the glottal stop cue in L2 English word-boundary segmentation. On the other hand, if the L2 learners of English, "… to some extent concerning perception, might be ... functionally monolingual..." (Altenberg 2005; c.f., Cutler et al. 1992),

then, they would, in that case, rely primarily on routines of in their L1 system in the L2. For example, according to PAM-L2 (i.e., Perceptual Assimilation Model for L2 Learners; Best and Tyler, 2007), Spanish L2 learners of English may assimilate the allophonic aspiration variants in English (i.e., [p] ~ [p^h]) into one category in their L1 (i.e., the Spanish short-leg /p/) resulting in functional difficulty in hearing the aspiration cue for word-boundary segmentation in the L2. On the other hand, if the speakers of a particular language functionally utilized aspiration in the L1, PAM-L2 would predict that they would not have much difficulty discriminating the English [p] from [p^h] because there would not be an assimilation of these two sounds into one category as in the case of Spanish.

As suggested by Shoemaker, to better understand the question of L1 transfer versus universality in differential sensitivity to word-boundary segmentation cues, we need to test English learners whose L1 uses aspiration in a contrastive manner. Mandarin seems to provide a good testing ground because Mandarin contrasts aspirated and unaspirated voiceless stops phonemically (Duanmu 2007; Maddieson 1984), e.g., /paŭ/ '*full'* and /p^haŭ/ '*to run*'; that is, Mandarin speakers distinguish /p/ and /p^h/ to render distinct meanings. In other words, Mandarin speakers are sensitive to the VOT difference in the positive long-lag region because it can change the meaning of the utterance. The remaining question is whether speakers of Mandarin would utilize this acoustic information for word-boundary segmentation in English.

Moreover, according to Duanmu (2007), a glottal stop is not typically inserted before a vowel in onset syllables in Mandarin. Wu (1992) also reports that in the consonantal prosthesis in vowel-initial syllables, the glottal stop is only used 0.4% of the time. If Mandarin speakers are somehow more sensitive to the glottal stop cue than to aspiration, we can further add evidence to the literature that the use of the glottal stop is indeed an unmarked cue; perhaps abided by Universal Markedness. However**,** if we were to suppose that speakers of Mandarin were more sensitive to the aspiration cue than the glottal stop cue, then in that case, we may propose that L1 transfer has a stronger influence than Universal Markedness in the perception of L2 word boundaries, and the glottal stop

is not a universally accepted preferred cue for word-boundary segmentation across the board.

2. THE TWO CUES IN ENGLISH AND MANDARIN

The two acoustic-phonetic cues, aspiration and the glottal stop, are utilized differently in English and Mandarin. First, both the glottal stop and aspiration cues are not phonemic in English, but aspiration is in Mandarin. As can be seen from Figure 1, the stop distribution is different in the two languages.

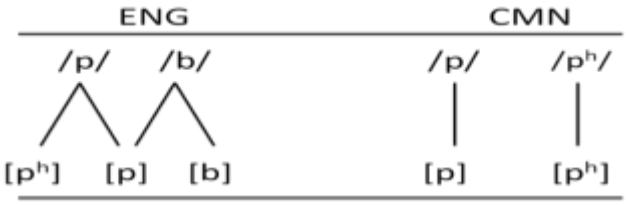

Figure 1. English stops vs. Mandarin Chinese stops

The VOT values of English voiceless stops are typically around 58 milliseconds (ms) to 80ms (Cho and Ladefoged 1999). Aspiration occurs in voiceless stops when they are at the beginning of a stressed syllable in English (Yavas 2011), e.g., [kʰipstʰɔlkɪŋ] *'keeps talking'* vs. [kʰipstɔlkɪŋ] *'keep stalking'*. Furthermore, glottal stops are commonly found word-initially when the words are without onset (Garellek et al. 2013), e.g., [sinʔaɪs] *'seen ice'* vs. [sinaɪs] *'see nice'*.

On the other hand, Mandarin contrasts aspirated and unaspirated voiceless stops phonemically. Therefore, it is reasonable to assume that Mandarin speakers are sensitive to the VOT difference. Although English does use aspiration phonetically, the VOTs of English aspirated stops are generally shorter than those of Mandarin aspirated stops (Liu et al. 2008) because Mandarin generally exhibits a long-lag VOT over 90ms, and stops would thus be categorized as "highly aspirated stops (Cho and Ladefoged 1999:223)". For instance, when the VOT value in '*pin'* [pʰɪn] is 75ms or 105ms, it will most likely not significantly affect the performance of

English native listeners. That is, the extra 30ms of a longer-lag VOT in [p^hɪn] does not alter the lexical representation of the word, and therefore it would be ignored.

The difference between a short-lag and a long-lag VOT is a meaningful contrast for English speakers (e.g., /p/ vs. /b/ in word onsets), and thus English speakers are presumably highly sensitive to the difference. However, the voiced stops VOT in English has never been found to be a cue for a word-boundary. Our focus is on a comparison of the glottal stop and aspiration cues for word-boundary segmentation. Therefore, the degree of sensitivity at the lower boundary is irrelevant to our study. Presumably, the long-lag VOT in English is more crucial to Mandarin speakers.

Thus, we may assume that Mandarin speakers are more attuned to aspiration than speakers of other languages that do not feature aspirated stops, such as Spanish, French, or Arabic. This leads us to predict that L2 learners of English whose native language is Mandarin may rely on aspiration when segmenting words more than Spanish, French, Japanese, and Arabic because of their greater attested sensitivity to this acoustic cue.

Testing Mandarin speakers can also allow for a constraint on the possibility of L1 transfer concerning the glottal stops because, unlike English, the glottal stop is not typically inserted before a vowel in an onset-less syllable in Mandarin. Duanmu (2007) proposes "…that the onset slot is optional [in Mandarin] for syllable structure (p.79)". Additionally, Wu (1992) reports that there might be an "optional" consonantal prosthesis in vowel-initial syllables, and the types of prosthesis might vary from environment to environment and from person to person. He reveals that the optional consonantal prosthesis before a vowel-initial syllable might be a velar approximant, a voiced velar fricative, a glottal stop, or zero onset (i.e., no insertion). He finds that the consonantal prosthesis rate is 18.6%, 19.1%, 0.4%, and 74.5%, respectively, suggesting that zero onsets occur predominantly in Mandarin. Even when the consonantal prosthesis occurs, the preferred prosthesis onset is not the glottal stop, as the rate of glottal stop insertion is only 0.4%. Although glottal stop insertion on onset-less syllables is optional and not predictable in English and Mandarin, it is much less frequent in Mandarin than in English, Spanish, French, Arabic, and Japanese. Therefore,

Mandarin speakers are presumably not accustomed to employing word-initial glottal stops and may not use the cue in the linguistic task of the detection of a word-boundary.

2.1 Word Boundary in Mandarin

What would then be the strategies for deciding word-boundary segmentation in Mandarin? Mandarin is a syllable-timed language (Mok 2009). In a Mandarin syllable, maximally four segments are allowed; i.e., any consonant (C) + non-consonantal, non-syllabic, high glide + vowel (V) + non-syllabic sonorant glide or nasal (Yin 1989; see examples in figure 2). Mandarin does not allow any consonant clusters, except for the consonant + glide combination in the onset position.

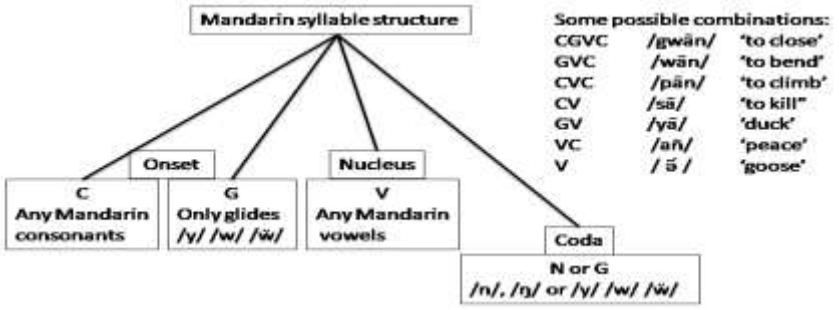

Figure 2. Example of Mandarin syllable structure (Adapted from Yip 1989)

The Mandarin coda is highly restricted to a singleton element with only either a nasal or glide. Therefore, the most common syllable structure in Mandarin is CV[2] (Duanmu 2009). Since each syllable corresponds to a Chinese character in Mandarin, a syllable boundary potentially provides the cut-off edge for word boundaries; i.e., any coda-less syllable, except for nasal-ending syllables, is a potential word boundary. Furthermore, Yang and Wang (2002) examined the acoustic correlates of the

[2] This phenomenon may present difficulties for Mandarin speakers in perceiving English words with onset clusters and codas; however, both instances can be equal in causing trouble because both are illegal in Mandarin (see section 4.2).

hierarchical prosodic boundaries[3] in Mandarin. They found that pitch-reset (i.e., the change of pitch between syllables), pre-boundary lengthening (i.e., the duration of the syllables preceding boundaries), and silence[4] are potential cues for prosodic and word-boundary segmentation (see Yang and Wang 2002 for detailed information). They claimed that "…the effect of pitch-reset and pre-boundary lengthening is more significant to [syllable and] prosodic word boundaries than intonational phrase boundaries… (p.3)", and the silence occurs at the larger phrase boundaries.

The concept of tonemes may provide insight into Mandarin word segmentation. According to Chen et al. (1997), tonemes, firstly mentioned by Pei (1966), treat lexical tones as phonemes consisting of specific pitch information in a tonal language. That is, with the same syllable structure, different pitch contours represent different tonemes. Each syllable bears certain tonal information for itself; that is, to look at the vowel /a/, for example, four tones can occur as four different /a/s. When moving from one syllable to the next one, such tonal information may need to be reset for the next syllable/word to bear the correct tone. Thus, the resetting of pitch may provide a cue for the boundary between syllables/words. Therefore, it is reasonable to predict that the pitch-reset between syllables is particularly important for word-boundary segmentation in the case of the speakers of a tone language, such as Mandarin. In two adjacent words in connected speech, if the pitch of the two syllables is slightly different, such difference can mark the word boundary for Mandarin listeners. The operation of pitch-reset might not be as apparent in between words of the same tones, for example, tone1 + tone1 (e.g., [kā:fē:] 'coffee' or [hwēfā]' to volatilize'); in that case, other acoustic information is needed. As pitch-reset is predicted to be an essential cue for Mandarin listeners, this study

[3] Typical prosodic boundaries are generally agreed to be applied at seven levels as: syllable, foot, prosodic word, prosodic phrase, intonational phrase, and sentence (Yang and Wang 2002).

[4] The duration of the silence in their study ranged from 273ms to 1151ms with a mean duration of 408ms, which is much longer than the typical duration of a glottal stop (55.10ms for in the present study). They found that there was no silence found in the case of prosodic word boundaries, with silences found for phrasal boundaries at M= 146ms, and silences found for intonational phrase boundaries at M = 408ms.

assumes that they may rely on this acoustic cue in English word segmentation; therefore, pitch contour will be controlled in this study.

Furthermore, Altenberg (2005) mentions that word intonation (or pitch) "can serve as [a] cue to syllable and word boundaries in a language" (p. 328; c.f., Ramana Rao and Srichand 1996). Shoemaker (2014) also notes that the "Speaker of English [may] use pitch movement as a segmentation cue in word-initial position …" (p.713). In other words, pitch/intonation information seems to be important in word-boundary perception. Also, Flege (1991, 1995) mentions that L2 learners need to rely on L1 phonotactics since the L2 phonotactics may not be available yet. The current study only focuses on two acoustic-phonetic properties (i.e., aspiration and the glottal stop). Therefore, it is crucial to control the potential exposure of the participants to additional cues and force them to rely only on the glottal stop or aspiration cues.

2.2 Outline of the Present Study

From the literature review, aspiration in stops is phonemic in Mandarin, and thus native speakers of Mandarin would be more sensitive to stop aspiration than, for example, Spanish, Japanese, French, or Arabic speakers. Mandarin speakers can perceive stop aspiration better than speakers of other languages who do not feature aspiration. The question is whether or not their sensitivity to this feature can help them use this aspect of L1 language production in L2 word-boundary segmentation. We predict that Mandarin speakers will perceive aspiration as a better cue than a glottal stop in word-boundary segmentation since they are more attuned to aspiration than the speakers referred to above.

The focus is to investigate which of the two acoustic cues is more preferred by the participants. To test the hypotheses, this study will control the potentially critical word-boundary cues, i.e., pitch-reset cues, by manipulating the pitch contour and lexical influence by using pseudo-words (see section 4.2). By doing so, the Mandarin speakers will be forced to use the two acoustic cues provided in the stimuli.

This study adopts the same stimulus conditions utilized in Altenberg (2005). Stimuli are placed within three major categories: aspiration, glottal stop, and double cue, in which both aspiration and the glottal stop are

present in two sliced phrase pairs (e.g., *'stroap inced'* vs. *'stro pinced'*, see section 4.2). It is predicted that the phrases of double cues will be easier than that of single cues, and native English speakers will perform better than L2 learners. Furthermore, research has shown that L2 proficiency generally increases as the length of residence increases (e.g., Guion et al. 2000). It is predicted that the longer the length of residence, the higher the level of the language skills that an L2 learner will have attained.

3. HYPOTHESES

The present study asks whether Mandarin L1-English L2 speakers can use aspiration as a better cue than the glottal stop in segmenting the word boundaries of English speech and investigates the chronology of the development of these learners' perception of the use of aspiration and the glottal stop within the use of their own inter-language.

It hypothesizes that:

1. Mandarin L1-English L2 speakers will prefer aspiration cues over glottal stop cues because they are accustomed to aspiration.
2. Double cue items will be easier to the participants than those with either aspiration or glottal stop cues since two cues should be better than one.
3. English native speakers will perform significantly better than L2 learners under the conditions of all three stimuli because the stimuli follow English phonetics and phonology.

4. METHODS

This study uses methods similar to Nakatani and Dukes (1977), Altenberg (2005), Ito and Strange (2009), and Shoemaker (2014)[5], but

[5] These four studies used similar stimuli originally suggested by Natatani and Dukes (1977). Altenberg (2005) created a different set of stimuli based on Natatani and Dukes's (1977) original conditions. Ito and Strange (2009) examined a new set of recordings using the same words as Altenberg. Shoemaker (2014) used the same recorded stimuli used in Ito and Strange (2009).

with a different set of stimuli. The stimuli employed in the present study are the same six conditions in three major categories as those in the previous studies, but are pseudo-words.

The three major groups are aspiration, glottal stop, and double cues, subdivided into six conditions:

1. aspiration group:
   a. VsC (e.g., *'loy spafes'* vs. *'loice pafes'*)
   b. CsC (e.g., *'keef stysk'* vs. *'keefs tysk'*)
   c. CsCC (e.g., *'twap skramth'* vs. *'twaps kramth'*)

2. glottal stop group:
   a. nasal (e.g., *'choln eeck'* vs. *'choll kneeck'*)
   b. obstruent (e.g., *'wrelf adged'* vs. *'wrell fadged'*)

3. double cue group (e.g., *'stroap inced'* vs. *'stro pinced'*).

All of the three voiceless stops, /p, t, k/, of English are included. Additionally, the stimuli are controlled for pitch level (see 4.4 for details). The procedure is also slightly modified from Altenberg's and Ito and Strange's in that this study uses a computer with a headset to present the stimuli to the individual participants. The participants are also tested one by one in a quiet room rather than in small groups.

4.1 Participants

Thirty-eight Mandarin L1-English L2 speakers (ME, the experimental group; mean age=24.99, SD=4.94) and twenty-eight native functionally monolingual English speakers, who only use English in daily life (NL, the control group; mean age=34.04, SD=14.17) were recruited. All of the MEs were born and had resided in China or Taiwan for at least 15 years[6] before coming to the US. Thirty of the MEs reported that they speak another Chinese dialect, whereas eight of them do not. All of the NLs self-reported as functionally monolingual speakers of American English; they have

[6] The criterion of 15 years is to ensure that the test subjects had fully acquired their native language.

learned a second language in school, but do not use it functionally in daily life. All of the participants orally and in writing reported no speech or hearing impairment before participation.

4.2 Stimuli Selection

The sliced two-word phrases are presented in Appendix C. The stimuli were created based on previously suggested conditions (Nakatani and Duke 1997 and Altenberg 2005; see Table 1 below). Previous studies used sliced words in order to control for contextual information. However, lexical information was still present. This is problematic because words differ in their lexical frequencies. For example, the two-word pair used in the previous studies, *'keeps talking'* vs. *'keep stalking'*, contains words that are not equal in lexical frequency. English L2 listeners are likely to select the former due to the frequency effect (Ellis 2002:151) since they might not have yet had the word *'stalking'* stored in their L2 lexicon because of its lower frequency in usage than the word *'talking'* in the language. Therefore, this study aims to further control for a frequency effect by using pseudo-word stimuli. By doing so, both of the two-word phrases in a pair will be equally unfamiliar to the participants' ears (for both native speakers and L2 learners).

One hundred fifty-two non-words were carefully selected from the 309,999 ARC non-words database based on the criteria of bigram frequency position-nonspecific (BFNC) and position-specific (BFSC) (Rastle, Harrington, and Coltheart 2002). For example, in the *'loice pafes'* two-word phrase, *'loice'* is the word1 and *'pafes'* is the word2, and for *'loice'* we find BFNC=516 and BFSC=101, whereas for *'pafes'* we find BFNC=796 and BFSC=248 (hereafter: *'loice pafes'* (516, 101:796, 248). The overall means for BFNC is 573 and for BFSC is 67 for word1, and 704 and 208 for word2, respectively. The selected words had bigram frequency values near the mean of the entire dataset; therefore, all of the words and paired phrases would have a similar bigram frequency according to the BFNC and the BFSC. Overall, the mean non-word frequency values for the entire dataset were 853.72 for the BFNC and 154.51 for the BFSC.

Table 1. Stimuli in two groups, three categories, six conditions

	Positive stimuli	Negative stimuli	conditions	categories
/pʰ /	*Loice pafes* /lɔɪs**pʰ**eɪfs/ (516,101:796,248)*	*loy spafes* /lɔɪs**p**eɪfs/ (256,13:1000,278)	VsC	Aspiration (*asp*)
	keefs tysk /kifs**tʰ**aɪsk/ (610,77:134,14)	*Keef stysk* /kifs**t**aɪsk/ (563,65:549,89)	CsC	*asp*
	twaps kramth /twæps**kʰ**ræmθ/ (458,131:780,103)	*twap skramth* /twæps**k**ræmθ/ (288,26:863,29)	CsCC	*asp*
/ʔ/	*choln eeck* /tʃɔln**ʔ**ik/ (680,99:527,54)	*choll kneeck* /tʃɔlnik/ (887,126:828,57)	nasal	glottal stop (*gl*)
	wrelf adged /wrɛlf**ʔ**ædʒd/ (602,87:1770,193)	*wrell fadged* /wrɛlfædʒd/ (588,26:1834,686)	obstruent	*gl*
	stroap inced /strop**ʔ**ɪnst/ (1173,146:1998,193)	*stro pinced* /strop**ʰ**ɪnst/ (932,15:2259,683)	double cues	double cue (*dc*)

* non-word frequencies are presented Elsewhere: [word1 $_{(BFNC,BFSC)}$: word2 $_{(BFNC,BFSC)}$].
* "positive stimuli" are instances with the experiment cues presented, and "negative stimuli" indicate the absence of the experiment cues in the sliced phrase.

The stimuli are VsC, CsC, and CsCC for the aspiration group, nasal and obstruent for the glottal stop group, and the double cue group.

One might notice in the VsC condition of the aspiration group that *'loice'* has a consonant coda structure, which is illicit in Mandarin phonotactics; thus, Mandarin listeners might be biased against the stimulus. However, in the same pair, the /s/ in *'spafes'* also creates a marked structure, CCV, for the Mandarin speakers. Although these two instances are phonotactically illegal in Mandarin for different reasons, it is possible that the participants would not be biased towards either stimulus based on the phonotactic familiarities.

To determine whether CVC (with the coda to be any obstruent except nasal) is a more marked structure than CCV (with a consonant onset cluster) or vice versa is not the aim of this study. Therefore, this study will assume that these structures are equally difficult for the Mandarin participants and focus on the word-boundary segmentation in question

here. Finally, 64 stimuli plus 12 fillers in each of the six conditions were selected and used; (conditions of: Vsc = 12, CsC = 12, CsCC = 12, nasal = 12, obstruent = 8, & *dc* = 8, also see Appendix C).

4.3 Stimulus Recording

One-hundred-fifty-two stimuli that constitute 76 pairs of potentially ambiguous phrases were recorded randomly by a phonetically trained female native speaker of American English in a sound-attenuated booth in an acoustics lab. Each of the 76 pairs was produced four times in the carrier phrase, "*I now say___again*", and recorded at 44,100Hz, 32-bit float, using Audacity (2.1.3) on a noise-free computer (Lenovo 110S) with an external microphone (MOVO MA200GY). The speaker read the entire list three times and then was instructed to read the phrases as naturally as possible at a normal rate of speech. The stimuli were randomly presented to the speaker using the PsychoPy program 1.8.3 (Peirce 2009). Four recordings were made on two separate days. We then selected two better-matched pairs from the four pair recordings based on a similar pitch level and duration, using Praat (Boersma and Weenink 2013).

4.4 Stimulus Manipulation

Two best pairs similar in pitch variation and syllable duration were sliced from the recorded carrier phrases and normalized for pitch level so that each of the two-word phrases had a relatively flat pitch contour. For example, the two instances of *'loice pafes'* had an average pitch of 179Hz for pair one and 173Hz for pair two and, similarly, 175Hz for the first *'loy spafes'* and 177Hz for the second *'loy spafes'*. Thus, these four related phrases were manipulated for a flat pitch at about 175Hz from the beginning to the end. Although not all the related pairs have the same flattened pitch of 175Hz in the case of the pairs of *'loice pafes'* - *'loy spafes'*, *'loice tafes'* - *'loy stafes'* and *'loice kafes'* - *'loy skafes'*, such manipulation might have altered the necessary F0 cue in the case of the initial stop consonants (Kim et al. 2002). In other words, there might be a concern about whether or not the participants can perceive the different stops accurately. We believe that the effect is minimal because the stimuli

are displayed to the participants on the computer screen; in other words, there is visual support, and the participants are told to focus on segmenting word-boundaries instead of discriminating stops. Figure 3 below shows the pitch contours before and after the manipulation.

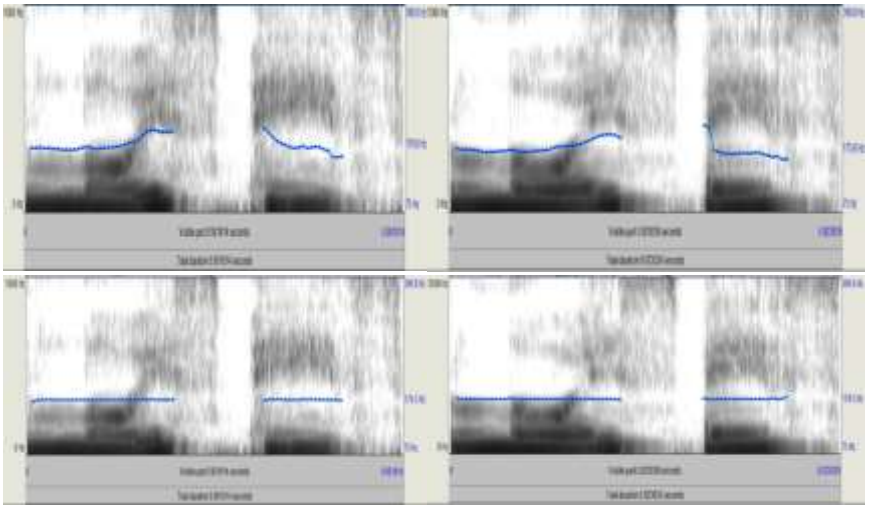

Figure 3. Pitch contours of '*loy spafes*' [lɔɪ speɪfs] (left) and '*loice pafes*' [lɔɪs pʰeɪfs] (right). The original pitch contour before the manipulation on the top panel and the pitch contour after the manipulations at the bottom.

As mentioned earlier, pitch bears tone and intonation information in Mandarin (Yang and Wang 2002). As the participants can perceive the change of the pitch between two words, as shown in Figure 3 before the manipulation, the pitch information was controlled. The mean duration of the sliced pitch was a little less than one second (M=885.72ms, SD=71.81).

One set of the carrier phrase tokens, '*I now say*' (988.03ms) and '*again*' (534.13ms), were selected from the recordings and used throughout the entire experiment. That is, the PsychoPy was programmed to repeat the same tokens for '*I now say*' and '*again*', but insert each sliced stimulus phrase for each trial. The average duration of the stimulus sentence was about 2.5 seconds (998.03ms for *'I now say'* + 885.72ms for the average length for 152 sliced-pairs + 534.13ms for *'again'*). The pitch of the

carrier phrase tokens was not manipulated so as to preserve some of the naturalness of the entire sentence.

We also analyzed the duration of the stop aspiration and the glottal stop in the 152 phrases. The mean VOT duration was 62.40ms (SD=19.63) for the aspirated stimuli and 15.24ms (SD=7.64) for the unaspirated stimuli. The difference between the positive and negative aspiration cues was significant at $p < 0.001$ (t (96) =16.378). The mean duration of the glottal stop was 55.01ms (SD=22.46). This was shorter than that of Altenberg's (77ms, SD=34.27, p.338), but was closer to what Garellek (2013) has suggested for the duration of a glottal stop (40 - 50ms, p.58).

4.5 Naturalness Rating of the Stimuli

The final 76 pairs were then rated for their naturalness by a different group of eight native English speakers. They were asked to rate each pair set for word1, word2, and the pair itself on a scale of 1-7, in which seven was the most natural-sounding to their ears. For example, for *'loy stafes'*, the mean rating was 4.875 for *'loy'*, 4.375 for *'stafes'*, and 3.75 for *'loy stafes'*; in other words, they rated *'loy'* and *'stafes'* as being somehow as sounding more natural than *'loy stafes'* together. The 76 pairs' overall rating showed a mean score of 3.81, which indicates that the pseudo-word stimuli were neither wholly natural nor too unnatural to native ears. This result coincided with the initial selection strategy based on the BFNC and the BFSC from the non-word database, in which the mid-point of naturalness was reached.

4.6 Task Procedures

The experiment was conducted at two different locations. All NL English participants (the control group) were tested at an acoustic lab. All ME participants (Mandarin L1-English L2 speakers, the experimental group) were tested in a quiet student consultation room at an ESL school. All of the participants were seated in a quiet room, wearing a headset (Mpow model#: BH059A), in front of a laptop (Macbook model: A1502) and a response box (Cedrus model#: RB-740). The same set of equipment was used regardless of the testing location. All of the participants were

given a copy of the written instructions and oral instructions in their native languages. The English instructions were: "You will hear two-word phrases and see words on the computer screen. Decide which phrase you have heard, and accordingly press the red or blue button on the response pad in front of you - red for the left phrase and blue for the right phrase". The same instructions were given to the ME group in Mandarin to avoid any possible misunderstanding of the task instructions. The interaction between the researcher and the participants was always in the native language of the participants throughout the entire test procedure.

There were eight practice trials for the participants to familiarize themselves with the task to adjust the volume of the auditory stimuli to their comfort. They were told to listen carefully to the stimuli, focus on the sounds (phonetic forms) rather than the meanings of the words, and not to worry about not understanding the unknown words. The two phrases in a stimulus pair were displayed on the screen for two seconds, and then the auditory sentence (e.g., '*I now say*' + *'loy stafes'* + '*again*', together for about 2.5 seconds) followed. As mentioned above, the stimuli were manipulated so that all were rendered with a flat pitch contour, while the pitch of the carrier phrase was not modified so as to preserve some naturalness of the sentence. In other words, the participants heard the normal pitch carrier phrase, '*I now say*', a pitch-controlled stimulus in the middle, and then the normal pitch carrier phrase of '*again*' at the end.

The participants were requested to choose one of the phrases by pressing the buttons on the response box. There was a 2-second interval between the visual stimulus and the auditory stimulus; that is, two phrasal tokens were first displayed for two seconds on the screen, and then the stimuli sentence was played through the headphone. There were four crucial time points during a trial - point 1: the onset of the visual stimuli, point 2: the onset of the auditory stimuli, point 3: the end of the auditory stimuli, and point 4: the end of the trial. The reaction time was recorded from the end of the auditory stimulus to the time point when a response was given. Therefore, the duration of *'again'* was not included in the reaction time.

The common practice is to allow the participant to respond to the stimuli from the onset rather than at the offset; however, the reason for the current design was to keep the task procedures and format as close to the

previous studies as possible (Altenberg 2005; Ito and Stranger 2009; Shoemaker 2014). We wanted to keep the modification of the task in the original design to the minimum amount possible, as Altenberg (2005) asked her test subjects to record their responses after they had heard the complete sentence. The current design also allowed us to avoid any mis-presses of the button before the participants had heard the auditory stimulus.

The average time for a single trial was 5.85 seconds, two seconds for the pause, 2.408 seconds for the stimulus playing time, and 1.4494 seconds for the mean reaction time. This timing was indeed a close inter-stimuli interval in Altenberg (2005), who suggested that "six seconds is a comfortable amount of time in which the participant can perceive and respond". On average, the experiment lasted about 14 minutes (M=14'14", SD=1'30"), excluding the self-paced break.

5. DATA ANALYSIS AND RESULTS

This section presents the results and data analyses of the experiment. The use of the implementation of the generalized linear mixed-effects models (*glmer*) for the binomial response accuracy data and the linear mixed-effects models (*lmer*) for the continuous reaction time data were done using the *lme4* (Bates, Maechier, Bolker, and Walker 2015) packages in R (R Core Team 2018). The pairwise comparisons were conducted using Tukey's HSD tests implemented in the *emmeans* (Lenth 2018) package. Three separate analyses were conducted and are presented in three sub-sections:

1. Mandarin group: focusing on accuracy (section 5.1.1) as a function of item groups, (i.e., aspiration (*asp*), glottal stop (*gl*), and double cue (*dc*)), the length of residence (LOR), the trial blocks, and the reaction time (section 5.1.2).
2. English group: the same modeling approaches as for the Mandarin group, but excluding LOR (section 5.2.1 and 5.2.2)
3. Between-group: focusing on the comparison of the effects of the native language.

The statistical modeling was based on Winter's (2013, 2015) suggestion using the likelihood ratio test to determine the effects of each fixed effect before implementing the linear mixed effect models. Detailed model settings are explained in each of the sub-sections.

5.1 Mandarin Group

5.1.1 Accuracy of response

There were 38 ME speakers in the experimental group. The age of onset of learning English was 11.35 (SD=3.69) years old, and the mean of the length of learning was 13.64 (SD=5.56) years. The average age of arrival was 22.78 (SD=2.58) years old. In other words, the participants have received English training from about 11 years old and came to the US around 23.

The LOR of this group ranged from five months to 252 months, with a mean of 30.9 months and a median of 12 months. We dummy-coded our ME participants into three groups of level of proficiency according to their LOR: 18 beginners (0-12 months), 10 intermediates (13-24 months), and 10 advanced learners (24+ months). The mean response accuracy for each group was 70.93%, 70.19%, and 75.60% for beginners, intermediate, and advanced, respectively. Regarding accuracy for item types, *asp* was 68.13%, *gl* was 77.27%, and *dc* was 94.08%; the overall accuracy was 72.75%.

A full mixed-effects model and two reduced mixed-effects models were designed to statistically test the significance for item groups, LOR conditions, and two trial blocks concerning accuracy in response for the ME group. In the full model, the accuracy in response served as a function of item groups * LOR * trial blocks. The random effects included random intercepts for by-participants and by-trial tokens. Adding random slopes resulted in non-convergence, and thus they were excluded from the full model. In the reduced models, one of each fixed effect was excluded. Holistically, the likelihood ratio test showed that the nature of the item group was a strong significant factor [χ^2 (12) =46.111, p<.0001] in perceiving word boundary cues, but, interestingly, LOR was not a significant factor in perceiving word-boundary cues [χ^2 (12) =12.21,

p=.3539]. The results for the trial block were also not significant [χ^2 (9) =5.2652, *p*=.8106].

As predicted, the post-hoc analysis showed that the ME group did significantly better with the *dc* items than with the *gl* and *asp* items. However, the Mandarin listeners performed significantly better in the case of stimuli with glottal stop cues than with aspiration cues, which was against our prediction. The difference was statistically significant (*p*=.0038); therefore, this study reveals a result with is consistent with those of previous studies in that, in perceiving word-boundary cues, glottal stop cues were used more accurately than aspiration cues when perceiving word-boundary cues by Mandarin L2 learners of English.

In regard to LOR, the prediction about level of proficiency was not borne out by the current outcomes, but advanced listeners were marginally more accurate than the beginners (*p* =.0749). This result presumably suggests that we did not have enough participants in each group. Had there been more participants in each group, or had we compared participants with even more (or less) English experience, the difference might have been significant. The current results also show no difference in the case of the block effect; that is, the ME groups did not learn to use these cues during the course of the experiment (although they showed a faster reaction time in the case of block B than in that of block A; see next section).

Table 2. Post-hoc results of accuracy in response for three fixed factors for the Mandarin group

Contrasts	Estimate	SE	z.ratio	p.value
Item group:				
asp – dc	-3.023024	0.5584672	-5.413	<0.0001***
asp – gl	-0.671306	0.2086300	-3.218	0.0038*
dc – gl	2.351718	0.5709432	4.119	0.0001***
Level of proficiency based on LOR:				
Advanced – Beginner	0.8162143	0.3746851	2.178	0.0749·
Advanced – Intermediate	0.6764640	0.3997750	1.692	0.2082
Beginner – Intermediate	-0.1397504	0.236884	-0.590	0.8255
Trial block:				
Block A – Block B	-0.0723598	0.186623	-0.388	0.6982
Post-hoc results of interaction across item group and LOR				
Within level of proficiency across item group:				

asp,Advanced - *dc*,Advanced	-4.16374430	1.0812510	-3.851	0.0038**
asp,Advanced - *gl*,Advanced	-0.50927721	0.2532437	-2.011	0.5355
dc,Advanced - *gl*,Advanced	3.65446709	1.0919707	3.347	0.0232*
asp,Intermediate - *dc*,Intermediate	-2.36470215	0.6440538	-3.672	0.0074**
asp,Intermediate - *gl*,Intermediate	-0.59647824	0.2510267	-2.376	0.2970
dc,Intermediate - *gl*,Intermediate	1.92699928	0.6475203	2.976	0.0719
asp,Beginner - *dc*,Beginner	-2.71991404	0.6379506	-4.264	0.0007***
asp,Beginner - *gl*,Beginner	-0.90816263	0.2264552	-4.010	0.0020**
dc,Beginner - *gl*,Beginner	1.47368891	0.5264195	2.799	0.1154
Within item group across level of proficiency:				
asp,Advanced - *asp*,Beginner	0.35521189	0.1605866	2.212	0.3977
asp,Advanced - *asp*,Intermediate	0.15877537	0.1824231	0.870	0.9944
asp,Beginner - *asp*,Intermediate	-0.19643652	0.1591763	-1.234	0.9491
gl,Advanced - *gl*,Beginner	-0.04367353	0.2187661	-0.200	1.0000
gl,Advanced - *gl*,Intermediate	0.07157434	0.2462523	0.291	1.0000
gl,Beginner - *gl*,Intermediate	0.11524787	0.2173554	0.530	0.9998
dc,Advanced - *dc*,Beginner	2.13710465	1.0492293	2.037	0.5173
dc,Advanced - *dc*,Intermediate	1.79904216	1.1091516	1.622	0.7930
dc,Beginner - *dc*,Intermediate	-0.33806250	0.5865619	-0.576	0.9997
Across item group across level of proficiency:				
asp,Advanced - *gl*,Beginner	-0.55295074	0.2640224	-2.094	0.4771
asp,Advanced - *dc*,Beginner	-2.02663965	0.5262510	-3.851	0.0038**
asp,Advanced - *dc*,Intermediate	-2.36470215	0.6440538	-3672	0.0074**
asp,Advanced - *gl*,Intermediate	-0.43770287	0.2869301	-1.525	0.8439
dc,Advanced - *asp*,Beginner	4.51895619	1.0863959	4.160	0.0011**
dc,Advanced - *gl*,Beginner	3.61079356	1.0947249	3.298	0.0271*
dc,Advanced - *asp*,Intermediate	4.32251968	1.0898213	3.966	0.0024**
dc,Advanced - *gl*,Intermediate	3.72604143	1.1006364	3.385	0.0204*
gl,Advanced - *asp*,Beginner	0.86448910	0.2740809	3.154	0.0427*
gl,Advanced - *dc*,Beginner	-1.51736244	0.5487868	-2.765	0.1259
gl,Advanced - *asp*,Intermediate	0.66805259	0.2874387	2.324	0.3273
gl,Advanced - *dc*,Intermediate	-1.85542493	0.6624682	-2.801	0.1150
asp,Beginner - *dc*,Intermediate	-2.71991404	0.6379506	-4.264	0.0007***
asp,Beginner - *gl*,Intermediate	-0.79291476	0.2727467	-2.907	0.0869
dc,Beginner - *asp*,Intermediate	2.18541502	0.5259009	4.156	0.0011**
dc,Beginner - *gl*,Intermediate	1.58893678	0.5482926	2.898	0.0891
gl,Beginner - *asp*,Intermediate	0.71172611	0.2632225	2.704	0.1462
gl,Beginner - *dc*,Intermediate	-1.81175140	0.652228	-2.778	0.1219

*significance codes: 0.05 / **significance codes: 0.01 / ***significance codes: 0.001

5.1.2 Reaction time of Mandarin-speakers

For reaction time, the average for the ME group was 1.6225 seconds, where:

- *asp* cue: M= 1.6414(SD=1.1098) seconds
- *gl* cue: M=1.6320 (SD=1.1290) seconds
- *dc* cue: M=1.4105 (SD= 0.9585) seconds

The average reaction time in the case of the trial blocks was 1.80187 (SD=1.217) seconds for block A and 1.4401 (SD=0.949) seconds for block B. They were about a little less than half a second faster in the case of block B (p<.0001). The mean reaction time by level of proficiency was 1.4501(SD=1.018), 1.6523 (SD=1.187), and 1.7013 (SD=1.099) seconds for advanced, intermediate, and beginners, respectively. In other words, the advanced learners were slightly faster than the beginners.

The likelihood ratio tests showed that the use of the trial blocks [χ^2 (9) =185.08, p<.0001] showed a significant effect, but LOR [χ^2 (12) =14.673, p=.2598] did not. This implies a difference in reaction time in different trial blocks, but no difference in level of proficiency. The post-hoc test results show that, within the item group, the reaction time for *dc* was significantly faster than those for *asp* (*SE*=.074, p=.0032) and *gl* (*SE*=.078, p=.0072), but, that there was no significant difference between the reaction times for the *asp* and *gl* items (*SE*=.041, p=0.9835). In other words, they responded the fastest in the case of *dc* items and there was no difference in performance in the cases of the *asp* and *gl* items. The experience of the ME participants with regard to English did not influence their reaction speed, but they did become faster in the course of the experiment as the reaction times in the case of block B were significantly faster than in the case of block A *(SE*=.0425, p<.0001).

Table 3. Reaction time for Mandarin group for trial blocks, item groups, and LOR

Contrasts	Estimate	SE	z.ratio	p.value
Item group:				
asp – dc	0.2440292	0.0748733	3.259	0.0032**
asp – gl	0.0071072	0.0409126	0.174	0.9835
dc – gl	-0.23692203	0.0785142	-3.018	0.0072**
Trial block:				
Block A – Block B	0.3395273	0.0425301	7.983	<.0001***
Proficiency level based on LOR:				
Advanced – Beginner	-0.2566722	0.234793	-1.094	0.5177
Advanced – Intermediate	-0.2138980	0.266025	-0.804	0.7005
Beginner – Intermediate	0.0427741	0.234613	0.182	0.9818

*significance codes: 0.05 / **significance codes: 0.01 / ***significance codes: 0.001

5.2 English Group

5.2.1 Accuracy in response

There were 28 native English speakers (NL) in the control group. This group performed at 75.46% overall accuracy (c.f., ME group: 72.75%). We first noticed that this number was very different from those in the results of previous studies (which were all near ceiling-performance above 95% for the native English speakers; see Altenberg (2005), Ito and Strange (2009), and Shoemaker (2014). The poor performance was not surprising because we additionally controlled for the pitch/intonation cue and the lexical information cue from the previous studies.

The mean response accuracy for the NL group was 73.16%, 76.49%, and 91.52% for the aspiration cue (*asp*), glottal stop cue (*gl*), and double cue (*dc*), respectively. Descriptively, the group performed the best with *dc* cues, with the *gl* stop cues the second, and with the *asp* cues the worst. The likelihood ratio tests showed that the item group was a significant effect [χ^2 (4) =10.322, p=.0353] in this group, and the trial block was not a significant effect [χ^2 (3) =2.0833, p=.5553] in regard to accuracy in response. This was similar to the results for the ME listeners. In terms of the three major categories, the results show that the performance in the case of *dc* was significantly better than those for *gl* and *asp*, but no significant difference was found in performance between *asp* and *gl* for

this group. For *dc* vs. *gl*, there was a marginal difference in performance. Table 4 provides the results.

Table 4. Tukey method for comparing a family of three estimates for the NL group

Contrasts	Estimate	SE	z.ratio	p.value
Item group:				
asp – dc	-1.5704778	0.5101474	-3.078	0.0059**
asp – gl	-0.3838651	0.2596084	-1.479	0.3012
dc – gl	1.1866127	0.5329534	2.226	0.0668·
Trial block:				
Block A – Block B	-0.1108962	0.1745789	-0.635	0.5253

*significance codes: 0.05 / **significance codes: 0.01 / ***significance codes: 0.001

5.2.2 Response timing of English group

The average reaction time for this group was 1.2115 seconds (compare to 1.6225 for the ME group), where:

- *asp* cue: M=1.2223(SD=0.740) seconds
- *gl* cue: M=1.2251 (SD=0.7149) seconds
- *dc* cue: M=1.0580 (SD= 0.563) seconds

The reaction times in regard to the trial blocks were 1.2829 (SD=0.7208) seconds for block A and 1.1397 (SD=0.7171) seconds for block B. Like the ME group, the NL group responded faster in block B than in block A, and the responses for the *dc* cues were the fastest among the three categories. The full model included reaction time as a function of item group * trial block and their interactions. The reduced models were built by removing each of the fixed effects from the full model.

The results for the trial block were a strong predictor for reaction time [χ^2 (1) =49. p<.0001], suggesting that the participants became faster. Also, the reactions to the *dc* items were significantly faster than those to the *asp* and *gl* items; that is, there was no significant difference between the reaction time to *asp* and *gl* shown in Table 5.

Table 5. Reaction time for NL group for trial blocks and item groups

Contrasts	Estimate	SE	z.ratio	p.value
Item group:				
asp – dc	0.17463915	0.06650206	2.626	0.0235*
asp – gl	-0.00562874	0.03657460	-0.154	0.9870
dc – gl	-0.01802679	0.06977768	-2.583	0.0265*
Trial block:				
Block A – Block B	0.1657823	0.03077178	5.387	<.0001***

*significance codes: 0.05 / **significance codes: 0.01 / ***significance codes: 0.001

5.3. Group Comparison

5.3.1. Accuracy in response

The results demonstrated that the ME group showed a similar performance to the NL group when perceiving word-boundary cues. We predicted that the NL group would outperform the ME group because the test tokens were English, although non-words, but the difference in performance was not significant.

Figure 4 below depicts a comparison of the average accuracy in response by language group and by cue types for both groups.

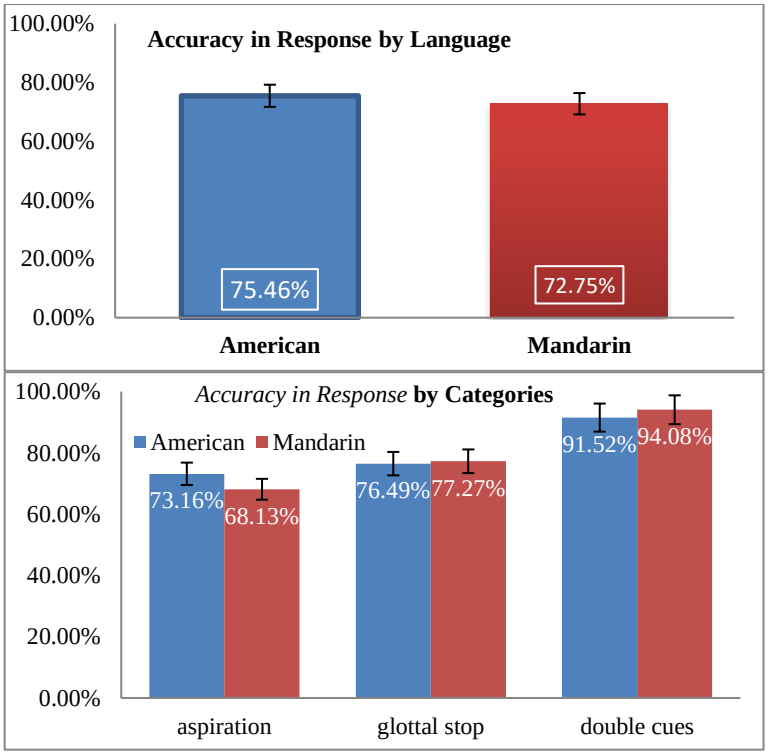

Figure 4. Group comparison of overall accuracy in response %

As shown, the two groups performed similarly, even though the NL group showed a slightly better performance in response to the aspiration cue than the ME group and the ME group performed better than the NL group in the case of the glottal stop cue and of the double cue categories.

The results of the likelihood ratio test were consistent with those of the single group analyses. The result of the item group was a significant factor in predicting the accuracy of response [χ^2 (8) =37.274, p<.0001], while the trial block [χ^2 (6) =3.5559, p=.7365] and language group [χ^2 (6) =8.822, p=.1838] were not significant predictors. In other words, there was no significant difference in performance between the NL and ME groups in terms of response accuracy [β=-0.07990455, SE=0.1479654, z= -0.54, p=.5892]; the trial blocks did not affect the response accuracy [β=-0.07178079, SE=.1258305, z= -0.57, p=.5684) either. The item group,

however, was shown to be a significant predictor (post-hoc results in Table 6). Overall, the result indicates no significant difference between the ME and NL groups in terms of accuracy in response.

Table 6. Tukey pairwise comparison across three item categories and two language groups

Contrasts	Estimate	St. Error	Z.ratio	p.value
Item group:				
asp – *dc*	-2.3640659	0.466123	-5.072	<.0.0001***
asp – *gl*	-0.6020921	0.2181879	-2.760	=0.0160*
dc – *gl*	1.7619738	0.4766402	3.697	=0.0006***
Language group:				
English – Mandarin	-0.01705293	0.2143732	-0.08	0.9366
Trial block:				
Block A –Block B	-0.08380838	0.1284567	-0.652	0.5141
Post-hoc results of interaction across item group and language group				
Within item group across language group:				
asp,English – *asp*,Mandarin	0.35945290	0.1621355	2.217	0.2298
gl,English – *gl*,Mandarin	0.07977071	0.2110378	0.378	0.9990
dc,English – *dc*,Mandarin	-0.49038242	0.5521378	-0.888	0.9495
Within language group across item group:				
asp,Mandarin – *dc*,Mandarin	-2.78898356	0.5108711	-5.459	<.0001***
asp,Mandarin – *gl*,Mandarin	-0.74193321	0.2237163	-3.316	0.0118*
dc,Mandarin – *gl*,Mandarin	2.04705035	0.5162491	3.965	0.0010***
asp,English – *dc*,English	-1.93914823	0.5800725	-3.343	.00107*
asp,English – *gl*,English	-0.46225102	0.2794895	-1.654	0.5625
dc,English – *gl*,English	1.47689721	0.5905049	2.501	0.1237

*significance codes: 0.05 / **significance codes: 0.01 / ***significance codes: 0.001

5.3.2. Reaction time between groups

Across the two listener groups, the mean reaction time was 1.4485 (SD=0.9844) seconds, and by blocks, was 1.5827 (SD=1.0680) seconds in block A and 1.3125 (SD=0.8711) seconds in block B. Figure 5 below shows the average reaction time for the two groups across the three categories of items.

As can be seen, the native English speakers (M=1.2115, SD=0.7224), responded faster than the L2 learners (M=1.6225, SD=1.107) in the case of all of the items. The likelihood ratio tests revealed that the language group was a significant factor [χ^2 (6) =39.517, p<.0001], suggesting that the NL group was significantly faster in performance than the ME group. The result of the trial blocks was also a significant factor [χ^2 (6) =235.26, p<.0001]. On the other hand, the item group was not a significant factor [χ^2 (8) =9.5072, p=.3013] in the case of reaction time.

Figure 5 provides a visual comparison of the reaction times between the groups.

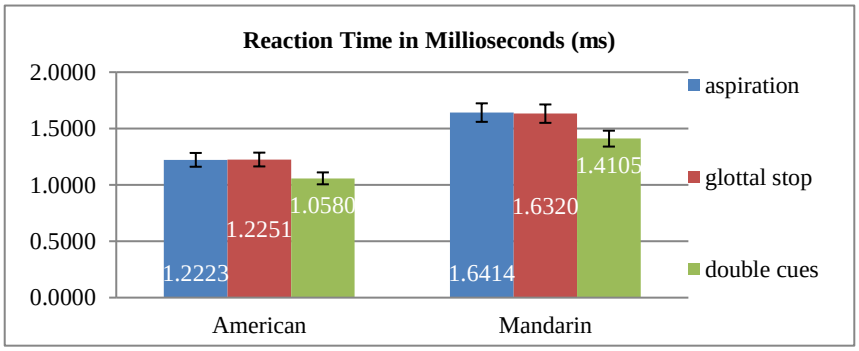

Figure 5. Group comparison of overall average reaction time across two groups and three categories

A mixed-effects model was built and analyzed using the *lmer* function in the lme4 package to determine the significance of the differences. The model included the reaction time as a function of item group * language group * trial block with random effects that included by-participants and by-trial random intercept and the by-participants random slopes for by item group and by-trial slopes for language groups.

The results provided in Table 7 indicate that the English speakers were significantly faster than Mandarin speakers (p=.0020) in responding to the stimuli. Both groups did better in block B than in block A. This might be because they had gotten used to the experiment task format in block A, and, therefore, performed better in block B. As found previously in the single group analyses, the responses to the *dc* items were significantly

faster than those to the *asp* and *gl* items, but no significant difference was found in the speed of the response between *asp* and *gl* items.

Table 7. The reaction time between language groups, trial blocks, and item groups

Contrasts	Estimate	SE	z.ratio	p.value
Language group:				
English – Mandarin	-0.4114373	0.1331092	-3.091	0.0020**
Trial block:				
Block A – Block B	0.2895731	0.03162431	9.157	<.0001***
Item group:				
asp – *dc*	0.24647974	0.08220462	2.998	0.0076**
asp – *gl*	0.01970952	0.04525639	0.436	0.9008
dc – *gl*	-0.22677022	0.08333510	-2.721	0.0179*
Within item group across language group:				
asp,English – *asp*,Mandarin	-0.41284518	0.14002738	-2.948	0.0376*
gl,English – *gl*,Mandarin	-0.41295101	0.14956224	-2.761	0.0639·
dc,English – *dc*,Mandarin	-0.40851568	0.14029657	-2.912	0.0419*
Within language group across item group:				
asp,English – *dc*,English	0.24431499	0.09765064	2.502	0.1234
asp,English – *gl*,English	0.01976244	0.05394679	0.366	0.9991
dc,English – *gl*,English	-0.22455255	0.09676739	-2.321	0.1857
asp,Mandarin – *dc*,Mandarin	0.24864449	0.09841392	2.527	0.1164
asp,Mandarin – *gl*,Mandarin	0.01965661	0.05421689	0.363	0.9992
dc,Mandarin – *gl*,Mandarin	-0.22898788	0.09911928	-2.310	0.1898

*significance codes: 0.05 / **significance codes: 0.01 / ***significance codes: 0.001

5.3.3. Trial conditions

Recall that we had six different conditions in three major categories: VsC, CsC, CsCC in the aspiration group (*asp*), nasal and obstruent in the glottal stop group (*gl*), and the double cue group (*dc*) (see Table 1). The overall accuracy percentages by condition were: 75.2% (VsC), 69.82% (CsC), 65% (CsCC), 85% (nasal), 66.86% (obstruent), and 92.99% (*dc*). Figure 6 shows the mean accuracy of response for the groups across six trial conditions. As depicted in the figure, the performance in the case of the *dc*-cue was the best, and then in the case of the nasal-cue results of which were consistent for both groups at 85%.

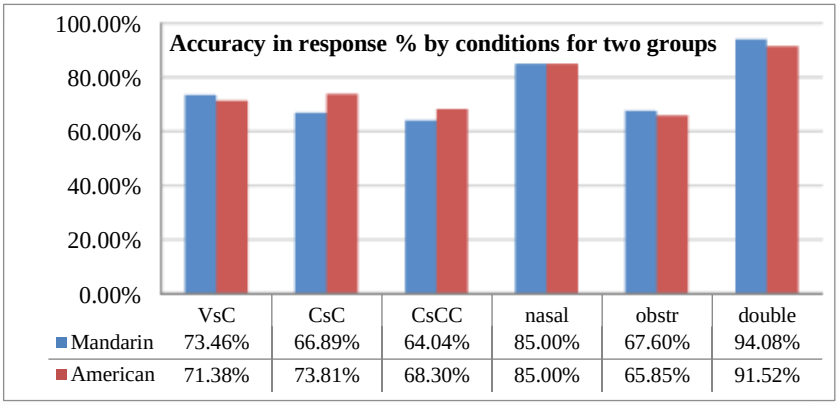

Figure 6. Group accuracy of response under six conditions

A mixed-effects model was built to investigate the performance of groups under each trial condition statistically. We modeled the accuracy of the response as a function of trial types * language group * trial blocks using the generalized linear mixed-effects model. The fixed effects included trial type (VsC, CsC, CsCC, nasal, obstruent, & *dc*), language group (ME & NL), trial blocks (block A & block B), and the interactions among them. Random intercepts were set for by-participants and by-trial tokens. Adding random slopes resulted in non-convergence and a "*singular fit"*.

Similar to the previous analyses in 5.3.1, the results of the performance of the two groups in both the trial block (p=0.4703) and language group (p=0.7264) were not significant. Concerning the trial types, the performances in the case of the *dc* and nasal-*gl* types showed that they functioned as significantly better cues than the others, but the results for the two types were not significantly different from each other (p=0.1079). Interestingly, we also found that both groups of listeners struggled with stimuli in the case of the obstruent-*gl* condition. Both groups performed significantly worse in the case of obstruent-*gl* conditions than in those of the *dc* and nasal-*gl* conditions, but their performance in the case of the obstruent-*gl* type was not significantly different from that of their performance in the cases of the three aspiration types, despite the fact that that the results show that the gl-cue was overall more salient with regard

to the accuracy in response than the asp-cue in both groups (p=.0160). It seems that the overall significance of the *gl* item was because the nasal-*gl* cue acted as a very strong, significant factor. Another interesting observation was that the NL group did better than the ME group in the case of only two conditions, (CsC, CsCC), while the ME group did better than the NL group in the case of VsC, obstruent-*gl*, and *dc*. The two conditions (CsC, CsCC) seem to be the most challenging for Mandarin listeners in regard to their phonotactics, as Mandarin allows neither consonant clusters nor consonant coda. Overall, no statistical significance was found for the aspiration groups for either of the two groups of listeners.

Table 8. Post-hoc comparison of six conditions across language groups

Contrasts	Estimate	SE	z.ratio	p.value
Language group:				
English – Mandarin	0.03867440	0.1105443	0.35	0.7264
Trial block:				
Block A – Block B	-0.05647813	0.0782332	-0.722	0.4703
Trial condition:				
CsC,*asp* – *dc*	-2.07967858	0.4066291	-5.114	<.0001***
CsC,*asp* – nasal,*gl*	-1.11291420	0.2768158	-4.120	0.0008***
CsC,*asp* – CsCC,*asp*	0.22893068	0.2548882	0.898	0.9471
CsC,*asp* – VsC,*asp*	-0.25780736	0.2563490	-1.006	0.9162
CsC,*asp* – obstruent,*gl*	-0.01743649	0.2911118	-0.060	1.000
CsCC,*asp* – *dc*	-2.30860927	0.4061960	-5.683	<.0001***
CsCC,*asp* – nasal,*gl*	-1.34184489	0.2761416	-4.859	<.0001***
CsCC,*asp* – obstruent,*gl*	-0.24636717	0.2904597	-0.848	0.9584
CsCC,*asp* – VsC,*asp*	-0.48673805	0.2555600	-1.912	0.3992
dc – nasal.*gl*	0.96676438	0.4190991	2.307	0.1912
dc – obstruent,*gl*	2.06224210	0.4288111	4.809	<.0001***
dc – VsC,*asp*	1.82187122	0.4069837	4.477	0.0001***
nasal,*gl* – obstruent,*gl*	1.09547772	0.3093073	3.542	0.0053**
nasal,*gl* – VsC,*asp*	0.85510684	0.2774005	3.084	0.0251*
obstruent,*gl* – VsC,*asp*	-0.24037088	0.2916837	-0.824	0.9632
CsC & CsCC in two groups:				
CsC,English – CsC,Mandarin	0.35596935	0.1403497	2.536	0.3171
CsC,English – CsCC,Mandarin	0.5077777	0.2791502	1.819	0.8079
CsCC,English – CsC,Mandarin	0.04991640	0.2779030	0.180	1.0000
CsCC,English – CsCC,Mandarin	0.20172482	0.1357000	1.487	0.9446
CsC,English – CsCC,English	0.30605295	0.2720119	1.125	0.9936

CsC,Mandarin – CsCC,Mandarin	0.15180842	0.2623723	0.579	1.000

*significance codes: 0.05 / **significance codes: 0.01 / ***significance codes: 0.001

6. GENERAL DISCUSSION

This study compared the performance of L1 and L2 English speakers in word-boundary perception, following an extensive body of works on the issue (e.g., Altenberg 2005; Ito and Strange 2009; Shoemaker 2014; Alammar 2016).

The present study examines the perceptually strong (or prominent) acoustic-phonetic cue for word-boundary segmentation for ME speakers. Secondly, it aims to investigate the connection between L1 phonemic knowledge and L2 word-boundary perception. It shows how ME speakers exploited two acoustic-phonetic cues perceptually to enable word-boundary segmentation in English pseudo-words. The main results of the previous and the present study are summarized in Table 9. As we can see, all of the L2 groups performed significantly better in the case of cues with glottal stops than in the case of those with aspirations.

The key finding of the current study provides additional evidence that the glottal stop may be "universal" as a cue for a boundary, as Altenberg (2005) and Shoemaker (2014) suggest. On the other hand, aspiration seems to be language-specific, and, in particular, specific to English. To the best of our knowledge, none of the languages examined in the literature of L2 word-boundary segmentation have this allophonic rule for the task of word-boundary segmentation.

Table 9. Results of previous studies and present study for accuracy in response

	L2 speakers	English	Stimuli	L2 language
	76.0%	97.0%	Real word	Spanish

Altenberg (2005)	***asp* < *gl*** ≈ *dc*	*asp* ≈ *gl* ≈ *dc*		
Ito and Strange (2009)	74.6% ***asp* < *gl*** ≈ dc	96.8% *asp* ≈ *gl* ≈ *dc*	Real word	Japanese
Shoemaker (2014)	74.6% ***asp* < *gl*** < *dc*	--- ---	Real word	French
Alammar (2016)	66.0% ***asp* < *gl***	80.0% *asp* < *gl*	Non-word	Arabic
Present study	72.75% ***asp* < *gl*** < *dc*	75.46% *asp* ≈ *gl* < *dc*	Non-word	Chinese Mandarin

---: not tested
<, =, ≈: "worse than", "equal to", and "approximately equal to"

This study also shows that second language learners do not necessarily perform in a manner that is less efficient than native speakers in segmenting speech when provided with an equal amount of acoustic-phonetic information (75.46% vs. 72 .75%; p=0.9366). Shoemaker (2014) mentions that L2 learners can use top-down information (e.g., lexical frequency, especially the advanced learners) to decode a linguistic message. Therefore, the previous studies controlled for a contextual cue using sliced phrases (e.g., *'keeps talking'* vs. *'keep stalking'*), but the lexical information was still preserved within these sliced phrases. We further controlled for two additional potential word-boundary cues - lexical frequency using English non-words and pitch cue by flattening the pitch contour.

A certain degree of caution should be considered when comparing data across studies; however, because this study controlled the lexical and pitch information, the level of the performance of our native English speakers was drastically lower when compared to that of the native speakers in previous studies, which did not control for lexical information and pitch contour. On the other hand, the accuracy of the response of the Mandarin speakers was comparable to that of the L2 speakers across all of the studies shown in Table 9 (around 75%, except for Alammar 2016). Alammar (2016) went a step further in controlling for lexical information, but her native English speakers were still able to perform (80%) significantly better than the L2 Arabic speakers (66%). It might be the case that the pitch contour was a significant word boundary cue for the native speakers in her study. When the pitch was controlled in this study, the native

English speakers did not do better than the L2 group. Our results here might suggest that the native speakers rely on lexical information and intonation/pitch in the normal course of things and were therefore influenced or biased by lexical and intonation information in the current design of the study. Further studies are needed to confirm the relative importance of lexical information, suprasegmental cues, and the two phonetic cues used in the current study in the word-boundary segmentation in the case of native speakers.

In terms of L1 transfer, holistically, our results do not support the hypothesis. Altenberg (2005) and Ito and Strange (2009) propose that L1 transfer might be the reason for why they found that the glottal stop cue was a better cue than an aspiration cue for Spanish and Japanese learners of English. Altenberg claims that L1 transfer might explain the low scores for her Spanish participants because the particular feature (i.e., aspiration) "does not occur in their L1 phonology (p.344)". However, Shoemaker (2014) noticed that Spanish, French, and Japanese speakers performed equally well in the case of the glottal stop, despite the different status of this feature in their languages. Therefore, L1 transfer seems to play a lesser role in L2 word-boundary segmentation. Our participants did not show L1 transfer, regardless of their English proficiency. Recall that we hypothesized that they would show an L1 transfer of the phonemic feature in question in the L2 perception task, but, instead, they did better with the glottal stop, which is not systematically used in Mandarin (see section 2.2). Perhaps having aspiration phonemically does not necessarily mean that it will be used as a word-boundary cue in an L2.

There might be a few explanations for the current results. First, Altenberg (2005) proposes that the glottal stop is "a universal phonetic default" (p.345) in that it can be inserted into an onset-less; and thus, it might often serve for use in word-boundary segmentation. Therefore, it may be the case that it is easier for an L2 learner to acquire a glottal stop than an aspiration for use in the segmentation of word boundaries, regardless of the learner's L1.

Secondly, the benefit of having the use of phonemic aspiration by our L2 group may be overshadowed by the phonotactic constraints of the syllable structures in Mandarin. Both complex onset clusters (except for consonant-glide) and non-nasal coda are not allowed in Mandarin.

Therefore, it is possible that when perceiving an illicit sequence of sounds, the system might be predominantly influenced by the phonotactic information, and acoustic details such as aspiration are ignored. If this is the case, once the L2 learners have fully acquired the complex L2 syllable structure, the sensitivity to aspiration may resurface. Suppose that the absence of the use of an aspiration cue in identifying word boundaries for the ME group was due to the phonotactics. In that case, it becomes challenging to argue either for or against whether Mandarin listeners did employ the aspiration cue under the conditions of the current design for eliciting information about the employment of aspiration cues. A study with further modification of the experiment and stimuli is needed.

Recall that we assumed that consonant coda and onset consonant clusters are equally marked for the Mandarin speakers. Our statistical models showed no significant difference in performance between the syllable structures of CsC and CsCC of the aspiration groups [β=0.15180842, *SE*=0.2623723, *z*= 0.579, *p*=1.0000]. These results call for a future study that examines languages that use aspiration phonemically, on a par to Mandarin, but that allow consonant coda and onset clusters, which differ from Mandarin.

Another possible explanation for why the Mandarin speakers in our study could not utilize aspiration cues better than the glottal stop in word segmentation might come from the duration of the VOT in the stimuli. The average VOT of the current stimuli was 62.40 ms. This might have been too short for the Mandarin participants to perceive the stimuli as aspirated. The Mandarin participants perceived aspiration for word-boundary segmentation for 68.13% of the time, but the results could have been different had we manipulated the VOT for even longer. This study was not designed to test how long the VOT should be for Mandarin listeners to perceive the items as aspirated reliably. In other words, with the current design, it was not possible to elicit reliable results for an aspiration cue for the Mandarin. Further study is needed to account for the results found in the current study.

There is a concern about the lack of a control for the level of intensity in the production of the stimuli since it can be essential in a stress-timed language. The level of intensity was analyzed for the 152 stimuli. The mean was 73.87 (SD= 2.38) dB, ranging from 68.67 to 78.84. The level of

intensity may be a potential cue for the use of word-boundary segmentation; however, it is unclear as to what level of intensity could be attributed to the task of word-boundary segmentation. The current study did not control for this factor. Future studies may compare any possible effect from the manipulation of the level of intensity for the same set of stimuli.

7. CONCLUSION

The present study investigated the use of two acoustic-phonetic cues, stop aspiration and the glottal stop, in tasks related to word-boundary segmentation. In conclusion, our findings indicate that the Mandarin group performed in a way similar to the NL group in terms of accuracy, but that the NL group performed better in regard to reaction time than the ME group. They both performed significantly the best in the case of the double cues and in the case of the glottal stop cues the second. The ME group performed significantly better in the case of the glottal stop cues than in that of the aspiration cues, whereas the NL group showed no significant difference in performance between the glottal stop cues and the aspiration cues.

While Mandarin speakers use aspiration contrastively in their native language, our results indicate that they rely on the glottal stop more than on aspiration in the L2 perceptual task. Thus, our findings add evidence to the claim that the glottal stop is a universally unmarked cue for the marking of word boundaries (Altenberg 2005; Shoemaker 2014).

Furthermore, the length of residence of the L2 learners in the US did not contribute to the accuracy in using the two phonetic cues, although advanced learners performed marginally better than the beginners and were slightly faster than the beginners in terms of reaction time. Our beginner participants had, on average, 2.57 months (ranging from zero to 12 months) of length of residence before the experiment. With only limited exposure to English, they were still able to perform at a somewhat high level of accuracy (70.60%). A future study might investigate learners with even less formal English experience; for example, learners who learn English as a Foreign Language in China or Taiwan without ample English

input. It is also essential to further our knowledge of Mandarin L1 word-boundary patterns in order to better understand the behavior of Mandarin learners in the L2 word segmentation task.

REFERENCES

Altenberg, E. P. 2005. The perception of word boundaries in a second language. *Second Language Research* 21(4):325-358.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1):1-48. (doi:10.18637/jss.v067.i01)

Bissiri, M. P., M. L. Garcia Lecumberri, C M.ooke, and J. Volín. 2011. The role of word-initial glottal stops in recognizing English words. Paper presented at the 12th Annual Conference of the International Speech Communication Association (INTERSPEECH 2011), August 27-31, 2011, Florence, Italy.

Boersma, P., and D. Weenink. 2013. *Praat: doing phonetics by computer [Computer program] (Version 6.0.21)*. 25 Sept. 2016. (http://www.praat.org/)

Broselow, E., S. I. Chen, and C. Wang. 1998. The emergence of the Unmarked in second language phonology. *Studies in second language acquisition* 20(2):261-280.

Chen, S. W. 2003. Acquisition of English onset clusters by Chinese learners in Taiwan. In *Postgraduate Conference*.

Cho, T., and P. Ladefoged. 1999. Variation and universals in VOT: evidence from 18 languages. *Journal of phonetics* 27(2):207-229.

Coblin, W. S. 2000. A brief history of Mandarin. *Journal of the American Oriental Society* 120(4):537-552.

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

Ellis, N. 2002. Frequency effects in language processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in Second Language Acquisition* 24(2):143-188. (doi:10.1017/S0272263102002024)

Garellek, M. 2013. *Production and perception of glottal stops*. Los Angeles: UCLA dissertation.

Gass, S. M. 2013. *Second Language Acquisition: An Introductory Course*. London: Routledge.

Ito, K., and W. Strange. 2009. Perception of allophonic cues to English word boundaries by Japanese second language learners of English. *The Journal of the Acoustical Society of America* 125(4):2348-2360.

Kim, M. R., P. S. Beddor, and J. Horrocks. 2002. The contribution of consonantal and Vocalic information to the perception of Korean initial stops. *Journal of Phonetics* 30(1):77-100.

Kuznetsova, A., P. Brockhoff, and R. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13):1-26. (doi:http://dx.doi.org/10.18637/jss.v082.i13)

Liu, H., M. L. Ng, M. Wan, S. Wang, and Y. Zhang. 2008. The effect of tonal changes on voice onset time in Mandarin esophageal speech. *Journal of Voice* 22(2):210-218.

Lisker, L., and A. S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical Measurements. *WORD* 20(3):384–422. (https://doi.org/10.1080/00437956.1964.11659830)
Maddieson, I. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.
Major, R. C. 2001. *Foreign accent: The ontogeny and phylogeny of second language phonology*. London: Routledge.
Nakatani, L. H., and K. D. Dukes. 1977. Locus of segmental cues for word juncture. *The Journal of the Acoustical Society of America* 62(3):714-719.
Peirce, J. W. 2009. Generating stimuli for neuroscience using PsychoPy. *Frontiers in neuroinformatics* 2(10):10.
Pennington, M. 2005. *The phonetics and phonology of glottal manner Features*. Bloomington: Indiana University Bloomington dissertation.
Rastle, K., J. Harrington, and M. Coltheart. 2002. 358,534 non-words: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology A* 55(4):1339-1362.
Shimizu, K. 2010. Acoustic analysis of English and Japanese stop voicing contrasts produced by Korean L2 learners. *名古屋学院大学論集 言語・文化篇* [*The Journal of Linguistics and Culture - Nagoya Gakuin Univeristy*] 22(1):1-10.
Shoemaker, E. 2014. The exploitation of subphonemic acoustic detail in L2 speech segmentation. *Studies in Second Language Acquisition* 36(4):709-731.
Winter, B. 2013. Linear models and linear mixed effects models in R with linguistic applications. *arXiv preprintarXiv:1308.5499*.
Winter, B. 2013. A very basic tutorial for performing linear mixed effects analyses. *arXiv preprintarXiv:1308.5499*.
Yang, Y., and B. Wang 2002. Acoustic correlates of hierarchical prosodic boundary in Mandarin. Paper presented at Speech Prosody 2002 (SP-2002), April 11-13, 2002, Aix-en-Provence, France.
Yavas, M. 2011. *Applied English Phonology*. Hoboken: John Wiley and Sons.
Yin, Y. M. 1989. *Phonological Aspects of Word Formation in Mandarin Chinese*. Austin: The University of Texas dissertation.
Yip, M. 2002. *Tone*. Cambridge: Cambridge University Press.

[Received 1 August 2020; revised 27 December 2020; accepted 5 June 2021]

Chiu-ching Tseng
Providence University
ctseng2@pu.edu.tw
ctseng2@gmu.edu

APPENDICES

A**:** Questionnaire for the Mandarin natives

Case#:_____

Date: _____

Gender: _____

1. 您的母语是哪种语言？ What is your native language? _____
2. 您有没有听力及说话能力上的问题？Do you have normal hearing and speech production? Y/N_____
3. 您的出身及成长地区？Where were you born and raised ?
4. 城市(City) _____省份(Province) _____;
5. 国家 (Country)_____
6. 除了普通话，您还会哪些方言？Other than Putonghua and your native tongue, what other Chinese languages do you know?_____
7. 您几岁来到美国的？At what age did you come to the U.S.?_____
8. 您来美国多久了？How long have you been in the U.S.?_____
9. 您学习英语多久了？How long have you studied English?_____
10. 你是怎么学习英语的 (上学校或自然学)？How did learn English? (academically or naturalistically) ?_____
11. 你学习英语的原因是什么？What are your reasons for learning English?

    _____
12. 你在课堂外使用英语吗？ Do you use English outside of the classroom? YES / NO
13. 除了英语您还会哪些外语? What other language(s) do you know?
14. _____

APPENDICES

B: Questionnaire for the English natives

Case#:_____

Date: _____

Gender: _____

1. Do you have normal hearing? YES/NO_____
2. What is your native language? _____
3. How old are you? _____
4. Where were you born?
5. City _____ State_____Country_____
6. What other language(s) besides English do you know?
   _____
7. Have you ever lived in a foreign country where the primary spoken language(s) is not English? YES/NO_____, If YES, where? _____
8. Are both your parents native speakers of English? YES/NO/BOTH NOT_____; if NO/BOTH NOT, what is/are his/her native language(s)?
   _____

APPENDICES

C: Stimuli

Practice items	spoot	/sput/	toab	/toʊb/			
	skook	/skuk/	kolb	/kɔlb/			
	rawps	/raʊps/	nage	/neɪdʒ/			
	rawp	/raʊp/	snage	/sneɪdʒ/			
	spoot	/sput/	toab	/toʊb/			
	skook	/skuk/	kolb	/kɔlb/			
	rawps	/raʊps/	nage	/neɪdʒ/			
	rawp	/raʊp/	snage	/sneɪdʒ/			
VsC group:				CsC group:			
loice	/lɔɪs/	pafes	/peɪfs/	keef	/kif/	spysk	/spaɪsk/
loice	/lɔɪs/	tafes	/teɪfs/	keef	/kif/	stysk	/staɪsk/
loice	/lɔɪs/	kafes	/keɪfs/	keef	/kif/	skysk	/skaɪsk/
loy	/lɔɪ/	spafes	/speɪfs/	keefs	/kifs/	pysk	/paɪsk/
loy	/lɔɪ/	stafes	/steɪfs/	keefs	/kifs/	tysk	/taɪsk/
loy	/lɔɪ/	skafes	/skeɪfs/	keefs	/kifs/	kysk	/kaɪsk/
theace	/ðis/	palt	/pælt/	chaic	/ʧeɪk/	speef	/spif/
theace	/ðis/	talt	/tælt/	chaic	/ʧeɪk/	steef	/stif/
theace	/ðis/	kalt	/kælt/	chaic	/ʧeɪk/	skeef	/skif/
thea	/ði/	spalt	/spælt/	chaicks	/ʧeɪks/	peef	/pif/
thea	/ði/	stalt	/stælt/	chaicks	/ʧeɪks/	teef	/tif/
thea	/ði/	skalt	/skælt/	chaicks	/ʧeɪks/	keef	/kif/
CsCC group:				nasal group:			
coophs	/kufs	prirp	/prɝp/	choln	/ʧɔln/	eeck	/ik/
coophs	/kufs	trirp	/trɝp/	choll	/ʧɔl/	kneeck	/nik/
coophs	/kufs	krirp	/krɝp/	claln	/klæln/	utched	/ʌtʃt/
cooph	/kuf	sprirp	/sprɝp/	clall	/klæl/	nutched	/nʌtʃt/
cooph	/kuf	strirp	/strɝp/	thaln	/ðæln/	eams	/imz/
cooph	/kuf	skrirp	/skrɝp/	thall	/ðæl/	neams	/nimz/
twaps	/twæps	pramth	/præmð/	rulm	/rʌlm/	arfes	/ɑɪfs/
twaps	/twæps	tramth	/træmð/	rull	/rʌl/	marfes	/mɑɪfs/
twaps	/twæps	kramth	/kræmð/	tewm	/tjum/	oltch	/ɔltʃ/

twap	/twæp	spramth	/spræmð/	tew	/tju/	moltch	/məltʃ/
twap	/twæp	stramth	/stræmð/	cew	/syu/	malk	/malk/
twap	/twæp	skramth	/skræmð/	cewm	/bdfh/	alk	/syu/
Obstruent group:				Double cues:			
wrelf	/wrɛlf/	adged	/ædʒd/	stroap	/strop/	inced	/aɪnst/
wrel	/wrɛl/	fadged	/fædʒd/	stro	/stro/	pinsed	/paɪnst/
bewsh	/bjuʃ/	aiche	/eɪʃ/	queap	/kwip/	abbed/	/æbd/
bew	/bju/	shaiche	/ʃeɪʃ/	quea	/kwi/	pabbed	/pæbd/
grauv	/grɔɪv/	aizzed	/eɪzd/	noit	/nɔɪt/	torched	/ɔɹtʃt/
graugh	/grɔɪ/	vaizzed	/veɪzd/	knoy	/nɔɪ/	orched	/tɔɹtʃt/
cheab	/tʃib/	indged	/aɪndʒd/	skoo	/sku/	corphs	/kɔɹfs/
chea	/tʃi/	bindged	/baɪndʒd/	skoock	/skuk/	orphs	/ɔɹfs/
Fillers:							
hoice	/hɔɪs/	kroice	/krɔɪs/	ces	/ses/	poock	/puk/
hoy	/hɔɪ/	skroice	/skrɔɪs/	cesp	/sesp/	oock	/uk/
highp	/haɪp/	spebbs	/spɛbz/				
highps	/haɪps/	pebbs	/pɛbz/				
corph	/kɔɹf/	strang	/stræŋ				
corphs	/kɔɹfs/	trang	/træŋ/				
smor	/smɔɹ/	naitched	/neɪtʃt/				
smorn	/smɔɹn/	aitched	/eɪtʃt/				
smoo	/smu/	dong	/doŋ/				

區分英語字句字段的聽力感知比較: 爆發停止音素對聲門停止音素

曾秋景
靜宜大學

關於英語單詞字段感知的研究報告表示，美語人士聽「聲門停止音」比「爆發停止音」更為精準（Nakatani and Dukes 1977）。從許多不同背景的第二語言學習者（西班牙語：Altenberg 2005；日語：Ito and Strange 2009；法語：Shoemaker 2014；阿拉伯語：Alammar 2016）等，也得到相同結論。本研究延續上述調查，報告對於中文人士對爆發停止音的敏感性是否會幫助他們在聽力感知上，用於英語字句分段時使用。結果顯示，當詞句中有聲門停止音時，他們區分字句的能力比那些有爆發停止音時能更準確。換言之，學習者母語中的特定音素的感知敏感性並不能幫助他們在第二語言聽力感知時輕鬆地使用該母語能力。這表明學習第二語言時「普遍文法」的影響。而使用聲門停止音確實可能是區分字段任務時最普遍及易用的音素手段。

關鍵字：聲門停止音、爆發停止音、字句分段、中文、英文、普遍文法