Perception of American English junctures by Chinese EFL learners

Qi Zhang[1,2] & Lei Wang[1,3]

Tongji University[1], Donghua University[2], Radboud University[3]
zhangqi@dhu.edu.cn, wl1410512@gmail.com

ABSTRACT

In this investigation, we report an identification experiment on the perception of American English junctures (i.e., word boundary) by advanced Chinese EFL learners. The stimuli contained thirteen phrase pairs that share the same segmental string, but differ in the word boundary location, e.g., *a nice man* vs *an ice man*, read by two native American speakers in a consistent stress pattern. Twenty-six participants listened to the randomly presented stimuli and performed a binary forced-choice task. The results showed that the Chinese listeners are able to utilize the acoustic information in correctly locating the word boundary. A talker effect was also reported that faster speech inhibits the correct segmentation. Finally, it was shown that some stimuli are easier to segment. This investigation may contribute to the learnability of L2 juncture cues and to issues in L2 speech perception in general.

Keywords: juncture cues, American English, word boundary, Chinese EFL learners.

1. INTRODUCTION

Juncture refers to 'any phonetic feature whose presence signals the existence of a grammatical boundary' [14, pp. 189]. More generally, it marks syllable boundaries and distinguishes word/phrase pairs that would otherwise be identical. For instance, the two phrases *keeps talking* and *keep stalking* in English share the same segmental composition /kipstɔkɪŋ/, but are pronounced differently in a number of (subtle) ways by native speakers. As summarized in Altenberg [1, pp. 331], among other differences, the /t/ in *keeps talking* shows a longer period of aspiration (hence a longer VOT) than that in *keep stalking*; the /s/ in *keep stalking* has a longer duration than in *keeps talking*; the /p/ in *keep stalking* has a longer closure duration than in *keeps talking*; the /s/ in *keep stalking* is greater in amplitude than in *keeps talking*.

Production and perception of word juncture by native speakers have been carried out in English [6,8,9], as well as in many other languages ([5] on Swedish; [11] 1980 on French; [10] on Dutch; [16] on Mandarin). Despite the fact that the syllable affiliation of speech sounds is controversial and is subject to a number of factors or principles, e.g., those summarized in Setter et al. [13, pp. 279], systematic juncture cues have been observed that distinguish otherwise ambiguous word/phrase pairs. In addition to the presence of aspiration of /p/ and the difference in (sub-)segmental duration and amplitude or intensity profile mentioned above, other cues include the variation in formant structure/transitions, fundamental frequency, and the presence of vowel laryngealization or a glottal stop before vowel-initiated syllables. There is little doubt that these junctures cues are reliably (though not equally) employed by native speakers in correctly locating the syllable boundaries (e.g., [6]) and some cues like glottal stop, laryngealization, aspiration have been shown to be stronger than other cues [3, 9]. However, little is known to what extent how non-native speakers rely on these juncture cues in speech segmentation.

We intend to provide a pilot study on the perception of American English juncture by advanced Chinese EFL learners, i.e., university students majoring in English in an attempt to open up avenues for future research. The major question we address is whether advanced Chinese EFL learners of English are able to segment phrase minimal pairs like *keeps talking* and *keep stalking*, based on the acoustic information in the speech signal. Though this experiment is not meant to test a certain hypothesis, we believe that the results may illuminate a number of theoretical issues that will be helpful for our future experiments.

2. METHOD

2.1. Preparation of stimuli

Thirteen phrase pairs were adapted from the previous studies [6,13], as shown in Table 1. Recordings of another word pair, *a name* vs *an aim*, were also collected and were used in the training. The word pairs were classified into three groups depending on whether the juncture consonant is an obstruent, a sonorant or a cluster. The word pairs 1–5 have a single obstruent consonant, the pairs 6–9 a sonorant consonant and the pairs 10–13 a consonant cluster. The members within each pair were also grouped depending on the number and the consonantal/vocalic status of the segment(s) right before and after the word boundary where juncture occurs, i.e., whether it

is V#C, C#V, C#C, C#CC, CC#C or V#CC (C and V stand for consonant and vowel, respectively).

Table 1: Thirteen phrase pairs.

	No.	V#C	C#V	C#C
single obstruent	1	why pink	wipe ink	
	2	my take	might ache	
	3	buy coil	bike oil	
	4	grey day	Grade A	
	5	why choose		white shoes
single sonorant	6		more ice	more rice
	7	see lying	seal eyeing	
	8	a nice box	an ice box	
	9	hoe maker	home acre	
		C#CC	CC#C	V#CC
cluster	10		might rain	my train
	11	keep sticking	keeps ticking	
	12	it sprays	it's praise	
	13	beer drips	beard rips	

Two native American speakers, one male (age: 54) and one female (age: 53), provided the recordings. The male speaker was born in Columbia and the female speaker in Albany Oregon. They spent most of their lives in America.

The digital recording was conducted in a quiet office/classroom, using a Shure SM10A head-mounted dynamic microphone connected to a laptop via a Shure X2u pre-amplifier. They were asked to pronounce the phrases both in isolation and in carrier sentences, using the same stress/pitch pattern that they feel comfortable with for each phrase pair. Only the isolated ones were used in the perception study. The speech sounds were digitalized at a sampling frequency of 12,000 Hz for the male speaker and 44,100 for the female speaker in Praat [2], the latter of which was later resampled at 12,000 Hz.

2.2. Subjects

Twenty-six university students, 21 females and 5 males (age, mean: 21.3, sd: 1.2, range: 19–23), from Donghua University (Shanghai) were recruited. Among them, 22 participants are majoring in English (sophomore 5, junior 2, senior 10, postgraduate 5) and the remainder four in engineering (postgraduate 4). They were paid a small fee for their participation. None of them had self-reported speaking or hearing problems. Sixteen subjects were born from the northern part of China and mastered a second Mandarin dialect besides Standard Mandarin. By contrast, the remainder subjects were from the southern part and spoke either a Wu dialect (9 speakers, e.g., Shanghai, Suzhou) or Min dialect (1 speaker) besides Standard Mandarin. The two native speakers who provided the recordings also paritipated in the experiment.

2.3. Procedure

The experiment was conducted in a studio for English listening and pronunciation practice centre at the School of foreign languages in Donghua University, where each participant was seated before a computer screen and equipped with a headset. Test materials were presented using ExperimentMFC implemented in Praat [2]. Before the experiment the participants were given a word list to familiarize themselves with the test phrases so as to minimize the word frequency effect. The stimulus was followed by a 600-ms 500-Hz beep and a 400-ms pause and was played once each time. There was a three-trial training session before the experiment. Upon hearing a sound, participants were instructed to perform a binary forced-choice identification task, where they have to press a button corresponding to one of the two phrases on the computer screen. The experiment was self-paced. After they have made a choice, they would click the 'next' button, situated in the lower right corner, to lead to the next stimulus. The 'previous' button was situated in the mirrored position in the lower left corner, the clicking of which would lead to the previous stimulus. There was also a 'replay' button, situated at the centre of the screen, which participants were allowed to click to play the current sound a second time. There are four repetitions for each phrase (208 stimuli: 13 phrase pairs * 2 members * 2 speakers * 4 repetitions), half of which were presented with the reverse order of the phrase buttons. The stimuli were randomly presented to the participants. The experiment lasted roughly half an hour.

3. RESULTS

3.1. Acoustic juncture cues

Visual inspection of the sound wave and the corresponding spectrogram revealed some widely attested acoustic juncture cues. In the word pairs 1–5, 10 and 13, the coda obstruents that resyllabify with the following segments show various degree of lenition and a shorter closure phase, compared with their word-initial counterparts (see Figure 1). In the word pairs 1–4 and 6–9, vowel-initiated words were often preceded by a short silent gap (see Figure 2) or laryngealization (see Figure 2); alternatively, a glottal stop was inserted before the vowel. In the word pairs 7 and 8, an earlier vowel formant transition into the consonant in *seal eyeing* and a longer portion of vowel nasalization in *an ice man* would appear to serve a salient acoustic cue (see Figure 2). In the word

pairs 11 and 12, aspiration would seem to play a primary role in segmentation.

Figure 1: *why pink* vs *wipe ink*

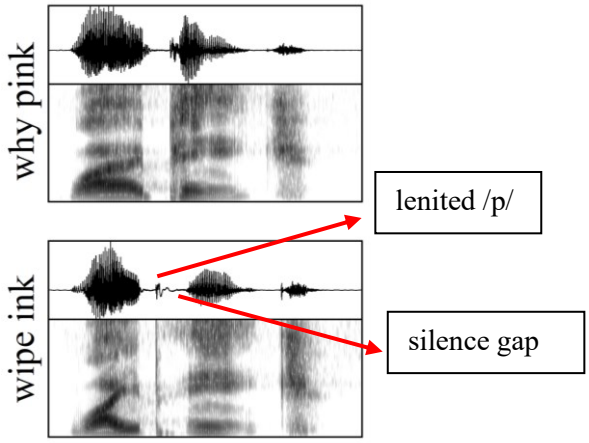
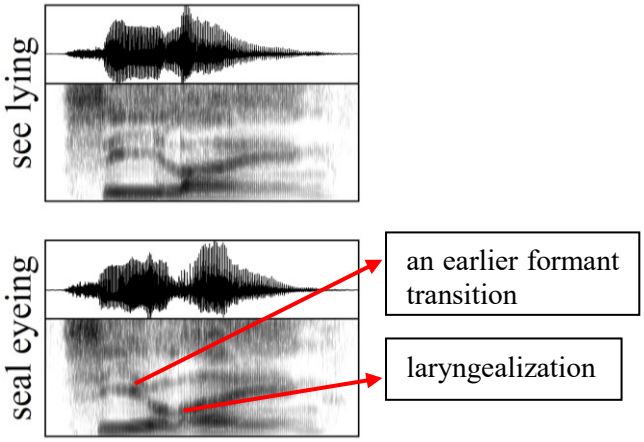


Figure 2: *see lying* vs *seal eyeing*

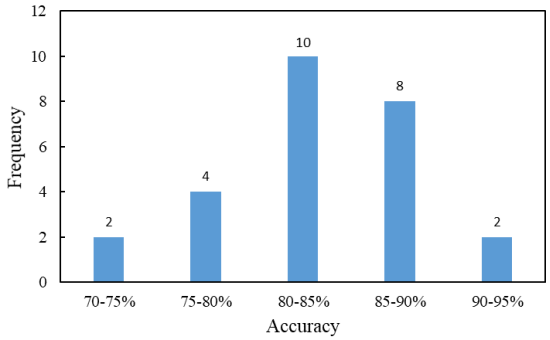


There was also a significant difference in the speech rate at which the two speakers pronounce the test material. The utterance duration was extracted for each phrase and for each speaker (the onset was located at the release burst if the utterance begins with a plosive). A paired-sample *t*-test yielded a significant effect of talker on the utterance duration with the significant level of 0.05, indicating that the female speech is faster than the male speech.

3.1. Overall accuracy

For the 26 EFL learners, the overall accuracy per subject over the entire dataset (208 stimuli) was calculated. The distribution was shown in Figure 3. With an average accuracy of 83.54% (standard deviation = 5.69%, range = 72.60–94.71%), their performance was comparable to the two native speakers (85.10% and 83.17% for the male and the female speaker, respectively).

Figure 3: Overall accuracy (% correct) for the 26 EFL learners



3.2. Speaker- and listener-dependent patterns

To assess the effect of *talker* (i.e., fast talker vs slow talker) on the overall accuracy across the phrases, a paired-sample *t*-test was conducted, with *talker* as a within-subjects factor. The results showed that the effect of voice is significant ($t(1,25) = -13.812$, $p < .001$), indicating that listeners performed better when exposed to the slow male speech (mean accuracy = 90.61%) than the fast female speech (mean accuracy = 76.48%).

3.3. Stimulus-dependent patterns

Table 2: Percentage correct for each phrase

	No.	C	Chinese listener (mean % correct)	English listener
single obstruent	1	p	94.47%	78.13%
	2	t	92.07%	100%
	3	k	79.81%	87.5%
	4	d	82.93%	87.5%
	5	tʃ	75%	78.13%
single sonorant	6	r	92.79%	100%
	7	l	91.35%	78.13%
	8	n	87.5%	84.38%
	9	m	69.95%	56.25%
cluster	10	tr	97.60%	96.88%
	11	st	79.09%	81.25%
	12	spr	70.67%	87.5%
	13	dr	72.84%	78.12%

The mean percentage correct of both Chinese and English listeners for each phrase pair is listed in Table 2 and some pairs would appear to be easier to segment than others. Within the 'single obstruent' pairs, juncture consonants /k, d, tʃ/ are more difficult to segment than /p/ and /t/. When the juncture consonant is a single sonorant, nasals /m/ and /n/ are more difficult to segment than /r/ and /l/.

4. DISCUSSION

The overall accuracy of the Chinese EFL learners is 83.54%, far above chance level (50% correct),

comparable to that of the two native speakers, who partially listened to their own speech and were thus presumably expected to achieve their optimal performance. Though the number of native American listeners is not large enough to allow a direct statistical comparison between the native and non-native listener groups, we would like to conclude that by and large Chinese EFL learners are able to use acoustic juncture cues in word segmentation of American English, at least for the young advanced learners of Chinese that were tested.

One logical reason to their attainment of the L2 juncture cues lies in the fact that some cues in Mandarin Chinese are readily and positively transferable to L2 speech segmentation. For instance, aspiration distinguishes voiceless aspirated plosives (e.g., /p/ vs /p^h/) and affricates (e.g., /ts/ vs /ts^h/) in Mandarin. Onsetless non-high vowels are often preceded by a glottal stop [ʔ] or some laryngealiztion and more rarely a velar fricative [ɣ] or a velar nasal [ŋ] [4]. Sensitivity to such cues in their first language could aid their L2 segmentation, especially when they serve strong cues for English juncture [9].

There is also a talker effect that stimuli produced by a naturally slow talker induced a greater percentage of correct identification than by a naturally fast talker. Similar results were obtained by Schwab et al. [12]. In our experiment, for one thing, it is evident that the allophonic realizations, i.e., the position- or context-dependent variation of segments, at the word boundary are more or less hyperarticulated and hence more canonically realized in the speech of the slow talker than in that of the fast talker. The mean increase in the percentage correct is 14.1% (sd = 5.22%, range = 4.81–25.96%) from the fast speech to the slow speech. For the two native speakers in the current experiment, the female talker showed a weak increase (4.81%), while the male speaker a more sizable one (24.04%). This indicates that there is a huge inter-speaker variation in the performance for both the L2 and the native listeners.

Some juncture patterns would appear to be easier to segment than others. Interestingly, the disparity in accuracy within the 'single obstruent' type follows from the markedness scale of these consonants, if the assumption is that less marked segments are easier to acquire, as suggested by e.g., [7]. Specifically, the affricate /tʃ/ is more marked than the singleton plosives; voiced /d/ is more marked than the voiceless /t/; the velar voiceless plosive /k/ is more marked than the voiceless labial and dental plosives /p, t/. Alternatively, the low accuracy in the pair *why choose* and *white shoes* may also be attributed to the fact that neither *choose* nor *shoes* begins with a vowel, so that additional salient cue like the insertion of [ʔ] or laryngealization that appear in the pairs 1–4 are not applicable. The high accuracy in /l/ and /r/ juncture consonants suggest that their allophonic variations are strong juncture cues compared with nasal juncture consonants [1,9, pp. 719].

The vocoid nature of Mandarin Chinese nasal coda has been shown in [15,16], which blocks its resyllabification with the following vowel across the word boundary. This lack of resyllabification has also been shown to be negatively transferred to the L2 speech of English such that the first /n/ in *an ice man* is often pronounced without a complete oral closure by Chinese EFL learners [15]. However, our investigation suggests that the /n/ juncture consonant was readily recognizable (with an accuracy of 87.5%). This may reflect a misconnection between speech production and perception. The successful perception is probably because similar cues in Chinese were also employed, though not explicitly in distinguishing this juncture type. Xu [16] showed that Chinese words /fa#nan/, /fan#an/ and /fan#nan/ have distinct acoustic patterns. The distinction between /fa#nan/ and /fan#nan/ is acoustically identical to that in *a nice box* and *an ice box*.

5. CONCLUSION

The recognizability of the acoustic juncture cues in American English by twenty-six advanced Chinese EFL learners was evaluated using a binary forced-choice identification test. Listeners heard recordings of thirteen minimal phrase pairs produced by one naturally fast male speaker and one slow female speaker, both of which are native Americans. The results showed that the Chinese listeners recognized the juncture pairs at above-chance levels, suggesting that L2 juncture cues are attainable. The slow speech was more easily identified than the fast speech. It was also shown that some phrase pairs were easier to segment than others, suggesting that some cross-linguistic and language-specific factors like markedness and transfer may play a role in L2 speech segmentation. An improved understanding of this matter can benefit from a larger collection of data, especially from native English speakers, as well as from future production data of such sentences from these Chinese speakers.

6. ACKNOWLEDAGEMENTS

Constructive comments by three anonymous reviewers on an earlier version are gratefully acknowledged (Some of them are not fully addressed due to space limitation). This research is supported by Fundamental Scientific Research Funds for Central Colleges and Universities awarded to the first author (18D111404).

7. REFERENCES

[1] Altenberg, E. 2005. The perception of word boundaries in a second language. *Second Language Research* 21(4), 325–358.

[2] Boersma, P. & Weenink, D. 1992–2015. Praat: doing phonetics by computer [Computer program]. Version 5.4.08, retrieved from http://www.praat.org/

[3] Christie, W. 1974. Some cues for syllable juncture perception in English. *The Journal of the Acoustical Society of America* 55(4), 819–821.

[4] Duanmu, S. *The phonology of Standard Chinese*, 2nd edn. Oxford: Oxford University Press.

[5] Gårding, E. 1967. *Internal juncture in Swedish* (Travaux de l'Institut de phonétique de Lund 6). Lund, Sweden: CWK Gleerups.

[6] Lehiste, I. 1960. An acoustic-phonetic study of internal open juncture. *Phonetica* 5(Suppl. 1), 1–54

[7] Major, R. & Faudree, M. 1996. Markedness universals and the acquisition of voicing contrasts by Korean speakers of English. *Studies in Second Language Acquisition* 18, 69–90.

[8] Mattys, S. & James, M. 2007. Sentential, lexical, and acoustic effects on the perception of word boundaries. *The Journal of the Acoustical Society of America* 122(1), 554–567.

[9] Nakatani, L. & Kathleen, D. 1977. Locus of segmental cues for word juncture. *The Journal of the Acoustical Society of America* 62(3), 714–719.

[10] Quené, H. 1993. Segment durations and accent as cues to word segmentation in Dutch. *The Journal of the Acoustical Society of America* 94(4), 2027–2035.

[11] Rietveld, A.C.M. 1980. Word boundaries in the French language. *Language and Speech* 23(3), 289–296.

[12] Schwab, S., Miller, J., Grosjean, F. & Mondini, M. 2008. Effect of speaking rate on the identification of word boundaries. Phonetica 65, 173–186.

[13] Setter, J., Mok, P., Low, E., Zuo, D. & Ao, R. 2014. Word juncture characteristics in world Englishes: A research report. *World Englishes* 33(2), 278–291.

[14] Trask, R. 1996. *A dictionary of phonetics and phonology*. London: Routledge.

[15] Wang, Z. 1997. 英汉音节鼻韵尾的不同性质 [Different natures of nasal codas in English and Chinese]. *现代外语*[Modern Foreign Languages] 4, 17–29.

[16] Xu, Y. 1986. 普通话音联的声学语音学特征 [Acoustic-phonetic characteristics of junctures in Mandarin]. *中国语文* [Chinese Linguistics] 5, 353–360.