MANDARIN AND ENGLISH USE DIFFERENT TEMPORAL MEANS TO MARK MAJOR PROSODIC BOUNDARIES

Chengxia Wang[1], Yi Xu[1], Jinsong Zhang[2]

[1]University College London, [2]Beijing Language and Culture University
chengxia.wang.15@ucl.ac.uk, yi.xu@ucl.ac.uk, jinsong.zhang@blcu.edu.cn

ABSTRACT

In this paper, we report findings of a major difference between Mandarin and English in terms of means of marking major prosodic boundaries. We performed detailed duration analysis on two large corpora, one in each language, using pre-labelled break indices as a reference indicator of boundary strength. Results showed that pre-boundary syllable duration stops increasing beyond break level 2 in Mandarin, but continues to increase in English. Meanwhile, the duration of silent pause significantly increases beyond break level 2 in Mandarin, as if to compensate for the lack of continuous syllable lengthening, while the increase in English is much less significant. Despite the robust difference, therefore, cross-boundary temporal distance, consisting of durations of both final syllable and silent pause, seems to be a common marker of boundary strength in both languages.

Keywords: Boundary, duration, final lengthening, silence, temporal distance.

1. INTRODUCTION

An important function of prosody is to provide cues for breaking up continuous speech into smaller chunks for ease of auditory comprehension. Of the variety of cues that have been reported, two are of particular importance, namely, pre-boundary lengthening and silent pause [4, 19]. Pre-boundary lengthening refers to the phenomenon that syllables and their component segments before a prosodic boundary are longer than they would be in other contexts [4, 9, 14, 20]. Also, the amount of pre-boundary lengthening is related to the strength of the boundary: the greater the strength, the longer the duration [9, 20]. Silent pause, the second important boundary cue, is often associated with a strong boundary [3, 10, 16]. It is not yet clear, however, how exactly these two kinds of cues are distributed across boundaries of different strengths.

Boundary strength is often represented by break index based on the ToBI annotation system [15]. In this system, break level 0 refers to syllable boundaries within word; level 1 refers to normal word boundary (or most phrase-medial word boundaries); break level 2 refers to a lower-level perceived grouping of words that does not have an intonational boundary marker; and break levels 3 and 4 are largely defined by intonational phrasing, referring to intermediate phrase and full intonation phrase, respectively [1].

For Mandarin, a slightly different break index system, namely, C-ToBI, is widely used [7]. In C-ToBI, break level 0 indicates the minimum break between syllables, usually within a prosodic word; level 1 indicates prosodic word boundary; level 2 refers to minor prosodic phrase boundary; level 3 refers to major prosodic phrase boundary; and level 4 refers to prosodic group boundary [7]. A major difference of C-ToBI from ToBI is therefore the lack of reference to pitch accents and boundary tones as determiners of break index levels.

There has been limited research on the relationship between boundary strength and the two kinds of temporal cues, especially for high-level boundaries. For English, [17] showed significantly different amounts of pre-boundary lengthening between all four levels of boundary strength: prosodic word, a group of words within a larger unit, intermediate phrase, and intonational phrase. But for Mandarin, [6, 21] showed no significant difference in pre-boundary lengthening between minor prosodic phrase and major prosodic phrase boundary. This raises the question of whether there is a difference between the two languages in terms of pre-boundary lengthening. There is even less research on the relationship between silent pause and boundary strength. For Mandarin, it is found that normally there is no silent pause after prosodic word, but silence duration increases with the break level beyond prosodic word [12, 18, 21]. This seems to suggest a trading relation between pre-boundary lengthening and silent pause for larger boundaries in Mandarin. No similar finding is made about English, although there have been suggestions that cues of lengthening and pausing may counterbalance each other [5, 13].

The present study is a preliminary corpus analysis aimed at examining the relationship between boundary strength and pre-boundary lengthening and silent pause in English and Mandarin. We are particularly interested in finding out if the distributions of the two types of cues across different boundary strengths are similar between the two languages.

2. METHODOLOGY

2.1. Corpus

2.1.1. English Corpus

For English, the Boston University Radio News Corpus [11] was used. It consists of news stories recorded by 3 female and 4 male FM radio news announcers. It had been annotated previously with orthographic transcriptions, phonetic alignments, part-of-speech tags and prosodic labels [11]. Only two levels of stress were distinguished: stressed and unstressed. The phonetic alignments were generated automatically [2]. Annotation for the news recorded in a lab was previously hand-corrected, while those recorded during actual broadcasts were not. The prosodic labels are previously marked by hand based on ToBI, and are available only for a subset of the corpus. 369 sentences which have both prosodic labels and syllable information were analysed in this study.

2.1.1. Mandarin Corpus

The Mandarin data were from Annotated Speech Corpus of Mandarin Discourse (ASCCD). There are 18 discourse structures, each containing 300-500 syllables and several paragraphs. 5 male and 5 female Beijing speakers who speak standard Mandarin read aloud the discourses naturally [8]. Four layers, including syllable tier, initial and final tier, break index tier and stress tier, were labelled previously [8]. Break indices were labeled previously based on C-ToBI[8]. Syllables with Neutral tone were excluded in the analysis.

2.2. Measurement

The measurements made were pre-boundary syllable duration, silence duration, and their sum as temporal distance.

3. RESULTS

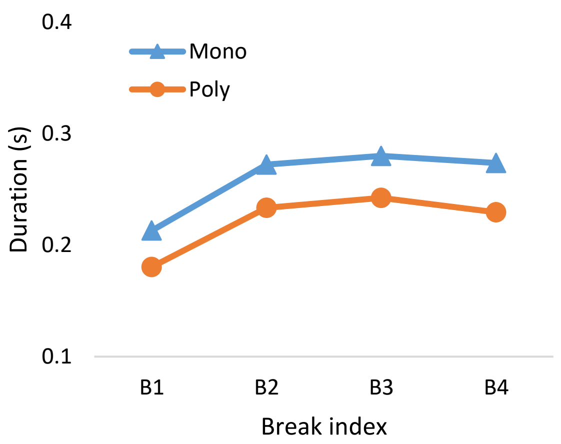
3.1. Mandarin results

The mean duration patterns are shown in Figs. 1-3. Fig. 1 shows that pre-boundary syllable duration ceases to lengthen beyond break level 2. In contrast, as shown in Fig. 2, temporal distance, which combines silent pause and pre-boundary duration, continues to increase beyond break level 2 for both monosyllabic and polysyllabic words.

A two-way repeated measures ANOVA was conducted, with the number of syllables (1 or more) in pre-boundary words and break index (1, 2, 3 and 4) as fixed factors, pre-boundary syllable duration as the dependent variable, and subjects as replication factor. The results showed a main effect of the number of syllables: $F(1, 9) = 70.700$, $p < 0.001$, partial $\eta 2 = .887$, and a main effect of break index, $F(3, 27) = 40.139$, $p < 0.001$, partial $\eta 2 = .817$. There was no interaction between the number of syllables and break index.

Bonferroni post-hoc analyses revealed that pre-boundary syllable before break 1 (M = 0.197, SD = 0.005) was significantly shorter than that before other breaks (break 2, M = 0.253, SD = 0.11, break 3, M = 0.261, SD = 0.008, break 4, M = 0.251, SD = 0.007). However, the other break levels do not differ from each other on pre-boundary syllable duration.

Figure 1: Pre-boundary syllable duration as a function of break index in Mandarin.



A two-way repeated measures ANOVA was conducted with the number of syllables in pre-boundary words (1 or more) and break index as fixed factors, temporal distance as the dependent variable, and subjects as replication factor. The results showed a main effect of the number of syllables, $F(1, 9) = 43.661$, $p < 0.001$, partial $\eta 2 = .829$, and a main effect of break index, $F(1.157, 10.409) = 86.737$, $p < 0.001$, partial $\eta 2 = .906$.

Bonferroni post-hoc analyses showed significant difference in each pairwise comparison between temporal distance at break 1 (M = 0.201, SD = 0.005), break 2 (M = 0.309, SD = 0.016), break 3 (M = 0.686, SD = 0.035) and break 4 (M = 0.912, SD = 0.070), $p < 0.01$.

There is an interaction between number of syllables and break index, $F(1.486, 13.370) = 10.393$, $p < 0.005$, partial $\eta 2 = .536$. A follow-up Paired-Samples t-Test showed that all paired samples are significantly different, $p < 0.05$. The effect of break index was more pronounced in syllables from monosyllabic words than polysyllabic words as break index increased.

Figure 2: Temporal distance as a function of break index in Mandarin.

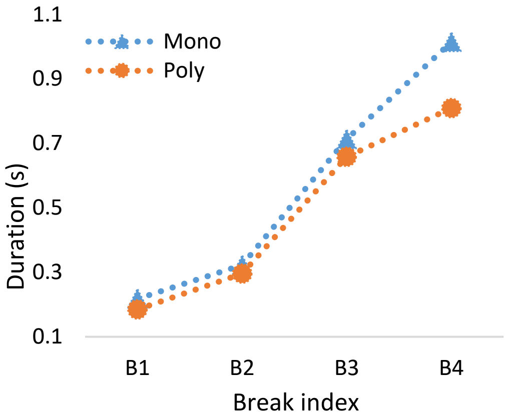
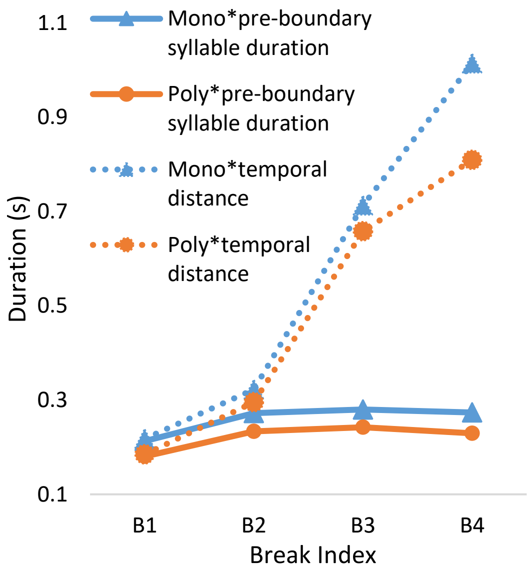


Figure 3: Pre-boundary syllable duration and temporal distance over break index in Mandarin.



3.2. English results

In English, stress is an important factor for syllable duration. Since polysyllabic words have stressed and unstressed syllables, we report results from monosyllabic words and polysyllabic words separately.

3.2.1. Monosyllabic Words

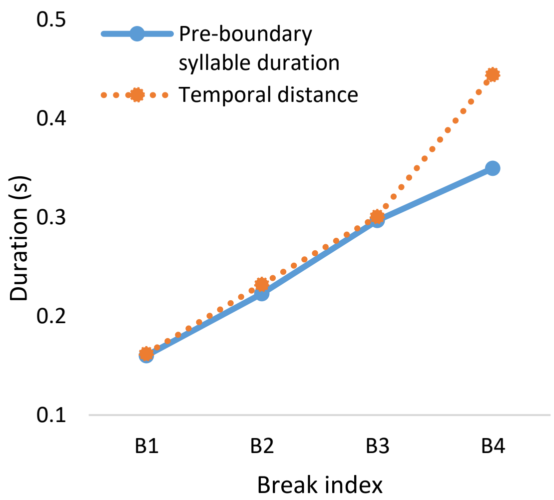
Fig. 4 shows that pre-boundary syllable duration increases gradually over break levels. It also shows that temporal distance has a similar trend and is largely overlapped with pre-boundary syllable duration except for break level 4.

Repeated-measures ANOVAs on pre-boundary syllable duration and temporal distance were conducted, with break level as a fixed factor and subjects as a replication factor. The results showed significant effects of break index on both pre-boundary syllable duration and temporal distance. Bonferroni post-hoc analyses revealed that each pairwise comparison was significant, $p < 0.05$.

Table 1: Results of repeated measures ANOVAs on the effect of break index on pre-boundary syllable duration and temporal distance.

Pre-boundary syllable duration	Temporal distance
F (3, 15) = 72.937, p < 0.001.B1 (0.160), B2 (0.223), B3 (0.297), B4 (0.350)	F (1.108, 5.540) = 38.903, p < 0.01. B1 (0.162), B2 (0.232), B3 (0.301) , B4 (0.444)

Figure 4: Pre-boundary syllable duration and temporal distance over break index after monosyllabic words in English.



3.2.2. Polysyllabic Words

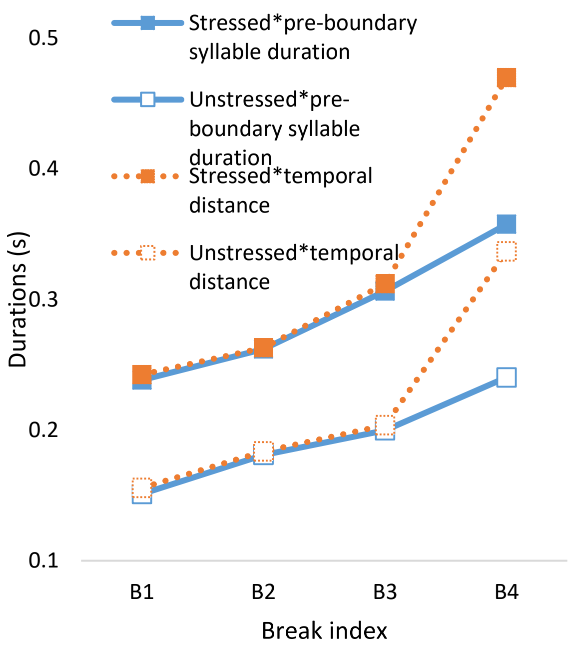
Fig. 5 shows that pre-boundary stressed and unstressed syllable duration increases gradually over break index. Also, temporal distance has a similar trend and is largely overlapped with pre-boundary syllable duration except for break level 4.

Repeated-measures ANOVAs on pre-boundary syllable duration and temporal distance were conducted with stress (stressed and unstressed) and break index as fixed factors and subjects as replication factor. The results indicated a main effect of stress and a main effect of break index on both pre-boundary syllable duration and temporal distance. There was no interaction between the two factors. Bonferroni post-hoc analyses showed that each pairwise difference was significant, $p < 0.05$.

Table 2: Results of repeated measures ANOVAs on the effect of break index and stress on pre-boundary syllable duration and temporal distance in English.

	Pre-boundary syllable duration	Temporal distance
Break index	F (3, 15) =90.651, p < 0.001. B1 (0.195), B2 (0.221), B3 (0.253), B4 (0.299)	F (1.117, 5.587) = 58.528, p < 0.001. B1(0.199), B2(0.223), B3(0.258), B4(0.403)
Stress	F (1, 5) = 303.664, p < 0.001	F(1, 5)=1309.778, p < 0.001

Figure 5: Pre-boundary syllable duration and temporal distance over break index after polysyllabic words in English.

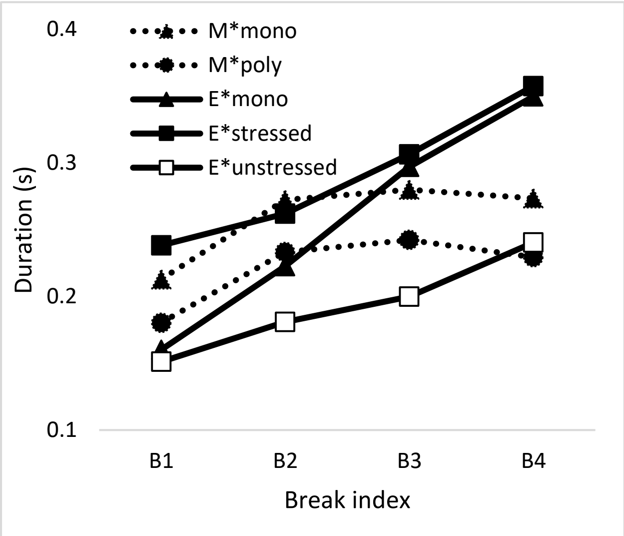


4. DISCUSSION

To highlight the main finding of the study, Figure 6 plots pre-boundary syllable duration in both Mandarin and English. As can be seen, in English pre-boundary syllable duration increases continuously with break index, whereas in Mandarin the duration increase stops beyond break 2. This is consistent with previous reports for Mandarin [6, 21] and English [17], respectively. But this is the first time that the difference between the two languages is clearly demonstrated. Also show for the first time is that duration of silent pause significantly increases beyond break level 2 in Mandarin, as if to compensate for the lack of continuous syllable lengthening, while the increase in English is less significant.

These results cannot be attributed to speaker differences between the two corpora. Previous research with professional radio broadcaster in Mandarin also showed no significant difference in pre-boundary lengthening between minor prosodic phrase boundaries and major prosodic phrase boundaries [6, 21].

Figure 6: Pre-boundary syllable duration in English and Mandarin as a function of break index.



A potential confound when comparing the two languages is the different criteria used in the labelling of the break indices between ToBI and C-ToBI. As mentioned in the introduction, the determination of break index in English depends heavily on intonation annotation [1]. Critically, break 3 is obligatory whenever a phrase accent is present, which by definition marks the end of an intermediate phrase even if there is no silent pause. The virtual overlap of temporal distance with break 3 in Fig. 5 shows that, indeed, little silence accompanied this break level. However, despite the lack of silence at break 3 in the English corpus, significant pre-boundary lengthening was found. This indicates that English syllables are much more flexible than Mandarin in terms of lengthening beyond break 2. On the other hand, despite the robust difference, cross-boundary temporal distance, consisting of durations of both pre-boundary syllable and silent pause, seems to be a common marker of boundary strength in both languages.

5. ACKNOWLEDGEMENT

This research was supported by China Scholarship Council to the first author. We thank Elizabeth Eden and Hao Liu for assistance with Python scripts, and Howon Kang for comments that greatly improved the manuscript.

6. REFERENCES

[1] Beckman, M. E., Ayers, G. 1997. Guidelines for ToBI labelling. The OSU Research Foundation, 3, 30.

[2] Kimball, O., Ostendorf, M., Bechwati, I. 1992. Context modeling with the stochastic segment model, IEEE Transactions on signal processing, 1584-1587.

[3] Lea, W. A. 1980. Trends in speech recognition. Prentice Hall PTR.

[4] Lehiste, I. 1972. The timing of utterances and linguistic boundaries. *J. Acoust. Soc. Am*. 516B, 2018-2024.

[5] Lehiste, I. 1979. Perception of sentence and paragraph boundaries. *Frontiers of speech communication research*, 191-201.

[6] Li, A. 1998. An analysis of Mandarin prosodic phrase duration. Institute of Linguistics, Chinese Academy of Social Sciences.

[7] Li, A. 2002. Chinese prosody and prosodic labeling of spontaneous speech. *Speech Prosody*.

[8] Li, A., Chen, X., Sun, G., Hua, W., Yin, Z., Zu, Y., Zheng, F., Song, Z., 2000. The phonetic labeling on read and spontaneous discourse corpora. *In Sixth International Conference on Spoken Language Processing.*

[9] Nakatani, L., O’Connor, K., Aston, C., 1981. Prosodic aspects of American English speech rhythm, Phonetica, vol. 38, no. 1-3, pp. 84-105.

[10] O'Malley, M., Kloker, D., Dara-Abrams, B. 1973. Recovering parentheses from spoken algebraic expressions. IEEE Transactions on audio and electroacoustics, 213, 217-220.

[11] Ostendorf, M. P. Price, J., Shattuck-Hufnagel, S. 1995. The Boston University radio news corpus, Linguistic Data Consortium, 1-19.

[12] Qian, Y.， Chu M., Pan, W. 2001. The acoustic analysis on prosodic boundaries of Chinese Mandarin. *The Proceeding of 5th National Conference on Modern Phonetics.*

[13] Scott, D. R. 1982. Duration as a cue to the perception of a phrase boundary. *J. Acoust. Soc. Am*. 714, 996-1007.

[14] Shattuck-Hufnagel, S., Turk, A. E. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of psycholinguistic research*, 252, 193-247.

[15] Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Hirschberg, J. 1992. ToBI: *A standard for labeling English prosody. In Second international conference on spoken language processing*.

[16] Swerts, M. 1997. Prosodic features at discourse boundaries of different strength. *J. Acoust. Soc. Am*. 1011, 514-521.

[17] Wightman, C., Shattuck - Hufnagel, S., Ostendorf, M., Price, P. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am*. 913, 1707-1717.

[18] Xiong, Z. 2003. An Acoustic Study of the Boundary Features of Prosodic Units. *Applied Linguistics*, 116-121.

[19] Xu, Y. 2009. Timing and coordination in tone and intonation—An articulatory-functional perspective. *Lingua*, 1196, 906-927.

[20] Xu, Y., Wang, M. 2009. Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin, *Journal of Phonetics*, vol. 37, 502-520, 2009.

[21] Yang, Y., Wang, B. 2002. Acoustic correlates of hierarchical prosodic boundary in Mandarin. Speech Prosody.