Temporal Cues of Discourse Boundaries in L1/L2 Mandarin Speech

Chunsheng Yang
The Ohio State University

This paper investigated the temporal cues of discourse boundaries in L1 and L2 Mandarin speech. Spontaneous speech in this study was elicited by asking subjects to retell a short non-verbal film, the *Pear Story* (Chafe 1980), after watching. The findings of the study are that 1) for L1 Mandarin speech, final syllable lengthening is a salient cue of prosodic boundaries and the pre-boundary syllable was lengthened the most; but the syllable lengthening does not reflect the discourse hierarchy. In addition, silent pause is another cue in L1 Mandarin speech and can reflect the discourse hierarchy; 2) for advanced American learners, final syllable lengthening is not a salient cue of prosodic boundaries, whereas silent pause duration is a salient temporal cue in L2 speech and can reflect the discourse hierarchy of L2 speech.

1. Introduction

Previous studies on L2 speech production and perception were mostly conducted on segmental levels, e.g. vowels and consonants. Recent years have witnessed the increase of studies on prosody, especially on the tonal phenomenon. Most prosodic studies on Mandarin Chinese were concentrated on tonal aspects, such as production and perception of Mandarin tones (Bent 2005), and perceptual training given to American listeners on Mandarin tones (Jongman and Wang et al., 2006; Sereno and Wang 2007, among others). By contrast, little study was conducted beyond the tonal prosody. However, as has been shown in some studies (Mareuil & Vieru-Dimulescu, 2005), prosody is more likely to lead to the perception of foreign accent than segmental deviations.

The present research investigates the temporal cues of discourse boundaries in L1 and L2 Mandarin Chinese speech. Specifically, the following questions are addressed:

(1) Are there temporal cues at various discourse boundaries in the speech of advanced American learners of Mandarin Chinese?

(2) If so, can the temporal cues reflect the discourse hierarchy?

(3) What is the difference in the above aspects between American learners and native Mandarin speakers?

2. Literature Review

Silent pause and pre-boundary lengthening are frequently suggested to be the most likely candidates for segmenting discourse (see Fon (2004) for a literature review on cues of prosodic boundaries in languages other than Mandarin Chinese).

2.1. Prosodic Boundary Cues in Mandarin

Shen (1992) used famous utterances of standing ambiguity in literary Beijing Mandarin to demonstrate that both silent pause and final syllable lengthening are robust cues in signaling sentential boundary locations. Tseng (2002) examined Taiwan Mandarin read speech and showed that pause is a major cue for prosodic structure in Mandarin spontaneous speech. Fon and Johnson (2004) examined the timing in Taiwan Mandarin spontaneous narratives and argued that the degree of syllable onset interval (SOI), the interval between the onset of one syllable and that of the next, which is essentially a combined measure of final lengthening and pause, is a good indicator for the structural organization of a discourse, namely, within a unit, the degree of lengthening reflects its positioning; the closer an SOI is to the end of a unit, the more likely it is to be lengthened. Across different units, the degree of lengthening can also reflect their hierarchical relationship, since final lengthening is proportional to the level of structural hierarchy. Liu and Li (2003) examined a corpus of Mandarin read speech and found that pause, pre-boundary syllable lengthening, F0 reset (F0 register), and F0 range are major cues of boundaries in Mandarin spontaneous speech. Yang and Wang (2002) examined the acoustic cues of hierarchical prosodic boundaries by investigating a large corpus of read Mandarin speech and found that pre-boundary lengthening is the acoustic correlate of weak boundary, pitch reset is that of medium boundary, and silence is that of strong boundary. In addition, the acoustic correlates of lower boundaries can also occur on larger boundaries, but the acoustic correlates of larger boundaries usually do not occur on lower boundaries.

Fon (2002) conducted a cross-linguistic study on syntactic and discourse boundaries cues in multi-lingual spontaneous speech (Taiwan Mandarin, Peking Mandarin, Japanese and English) and found that: in English, the structural boundaries are indicated by initial and final syllables and SOI lengthening. However, the two effects have different sources. Final lengthening is mainly due to the final lengthening effect while initial lengthening is incurred by pitch accents. The two lengthening effects have different scopes. The lengthening effect incurred by pitch accents is more focused on syllables than on SOI lengthening, while that incurred by boundary is more effective on SOI than on syllable lengthening. Discourse hierarch is indicated by degree of final SOI lengthening and initial syllable and SOI lengthening. But in Mandarin, the final lengthening effect is at work at P0[1] and P1 for syllables, but only at P0 for SOIs.

---

[1] P-1, P0 and P1 are syllables before or after the boundaries. P-1 is the second last syllable in a clause, P0 is the last syllable in the utterance whereas P1 is the first syllable after the boundary.

Discourse boundary strength is reflected only by SOIs. The following two tables illustrate the difference in cues of prosodic boundaries between Mandarin Chinese and English, adapted from Fon (2002, 129-132).

Table 1: Cues of Prosodic Boundaries in English and Mandarin Chinese

	English	Mandarin
Duration		
Syllable		
Final lengthening at P0	Yes	Yes
Final lengthening at P-1	No	Yes
Initial lengthening at P1	Yes	No
SOI (syllable onset interval)		
Final lengthening at P0	Yes	Yes
Initial lengthening at P1	Yes	No

Table 2: Prosodic Cues indexing Discourse Hierarchy in English and Mandarin Chinese

	English	Mandarin
Duration		
Syllable		
Final lengthening at P0	No	No.
Initial lengthening at P1	DSP 2 > DSP 1, 0	--
SOI (syllable onset interval)		
Final lengthening at P0	DSP2 > DSP1, 0	DSP2>DSP1> DSP0
Initial lengthening at P1	DSP2 > DSP1, 0	--

2.2. Problems in Previous Studies

Even though previous studies on Mandarin Chinese have found that final lengthening and silent pause are cues of discourse structure in Mandarin Chinese, the previous studies were conducted within different frameworks (Chinese ToBI, Mandarin ToBI, or mere impressionistic observation). Moreover, both read speech and spontaneous speech were used in previous studies. However, read speech is different from spontaneous speech in that the written script in read speech might eliminate the use of some cues in spontaneous speech, as has been argued in literature that spontaneous speech is a better source for researching discourse structure (Lee, 2005). Thus the present study only used spontaneous speech.

Meanwhile, as Fon (2002) has found, there are cross-linguistic differences in the prosodic cues of discourse structures. However, no previous studies have investigated the

effects of L2 prosodic cues in segmenting discourse and whether there are any differences between L1 and L2 cues and, if any, whether such differences in cues may contribute to the perception of foreign accent. The present study focuses on the first two questions. The third question will be left for future study.

3. Methodology
3.1. Subjects

Four male native Mandarin speakers and four male advanced American learners Mandarin Chinese participated in this study. The four advanced American learners are master or doctoral students at Chinese department. The four native Mandarin speakers are all from North China and all speak Standard Mandarin Chinese.

3.2. Materials

Spontaneous speech was elicited by playing a short film without spoken language, *The Pear Story* (Chafe, 1980), which can be downloaded for free. As mentioned in Fon (2002), to choose such as a soundless film has two advantages: 1) The film does not provide an explicit structure and viewers have to infer and reconstruct the structure of the film according to their understanding; 2) This movie was filmed to provide a relatively cultural-free story for discourse analysis studies, thus it is a good stimuli for this study as well.

3.3. Procedure

Subjects were given experiment instructions in English or Mandarin Chinese before watching the speechless film, *The Pear Story*. Subjects were told to retell the movie as if they were telling a friend who has never seen the movie. However, it would not matter even if they could not remember some details, for the purpose of eliciting their natural speech. All the recordings were done in the recording studio at the College of Humanities, the Ohio State University. The recordings were saved as wav. files, ready for data analysis in Praat.

3.4. Discourse Labeling

Drawing on Fon (2002), Fon and Johnson (2004) and Fon (2004), the present study labeled clauses as basic units. A clause is defined as having at least one main verb. All the utterances by the 8 subjects were segmented into clauses. The relationship between pairs of adjacent clauses was determined by using the guidelines outlined in Grosz and Sidner's (1986) computational psycholinguistic discourse framework. Discourse segment purpose (DSP) is defined as an intention that the speaker tries to convey to the hearer (Grosz and Sidner 1986). Three levels of discourse disjuncture were labeled, namely no discourse disjuncture (DSP0), low discourse disjuncture (DSP1), and high discourse disjuncture (DSP2).

Fon's original definition of DSP0, DSP1 and DSP2 merely considered the relationship between two adjacent clauses without putting them in the large context, e.g. the discourse level. The model proposed by Grosz and Sidner (1986) is a more comprehensive model, taking into account both the relationship between two adjacent clauses and the relationships of clauses with the whole discourse. In order to keep the same spirit of Grosz and Sidner and also make the discourse encoding simplistic, DSP0, DSP1 and DSP2 are defined as follows.

DSP2 refers to the relationship between two groups of clauses, each of which targets at a specific discourse purpose. Then DSP1 and DSP0 refer to the disjuncture between clauses within a DSP2: specifically DSP1 refers to the disjuncture between two related independent clauses, whereas DSP0 refers to the disjuncture between two minor clauses which are interrupted by only a clausal boundary. The relationships between DSP0, DSP1, DSP2 can be schematized as follows:

$[(\quad)_{DSP0} (\quad)_{DSP1} (\quad)_{DSP0}]_{DSP2} [(\quad)_{DSP0} (\quad)_{DSP1} (\quad)_{DSP1}(\quad)]$
One example of DSP encoding is shown as below:

他 把 帽子 拾 起来， 转回身 又 去 找 那个 少年，
{[Ta ba maozi shi qilai]$_{DSP1}$[zhuanhuishen you qu zhao neige shaonian]$_{DSP0}$
{[He picked up the hat]$_{DSP1}$[turned back and walked to the young man]$_{DSP0}$

把 帽子 还给 他。 那个 少年 很 感激， 送给 了
[Ba maozi huangei ta]}$_{DSP2}$ {[Neige shaonian hen ganji]$_{DSP1}$ [songgei le
[gave him the hat]}$_{DSP2}$ {[The young man was thankful]$_{DSP1}$ [gave

他 三个 梨…
ta sange li…]}
him three pears…]}

I believe that the encoding in this manner better reflects the arrangement of clauses within a discourse and more conforms to Grosz and Sidner's proposal (1986) than Fon (2002).

3.5. Measurements

Measurements were conducted on the two syllables before and after the boundaries. However, due to the nature of spontaneous speech, there are many fillers (e.g. *ranhou* "then"), hesitations and disfluency. If these syllables happen to be those to be measured, they were skipped and measurements were done on syllables after them (in most cases *ranhou* was used at the beginning of a clause).

Syllable duration, syllable onset interval (SOI) and silent pause duration were measured on the chosen positions across the boundaries. The following figure adapted from Fon (2004) illustrates the way how the measurement positions for syllable duration, SOI and silent pause duration were chosen.

Boundary

… σ   σ   σ   σ   σ]   [σ   σ   σ   σ   σ…

P$_{-2}$ P$_{-1}$ P$_1$ P$_2$
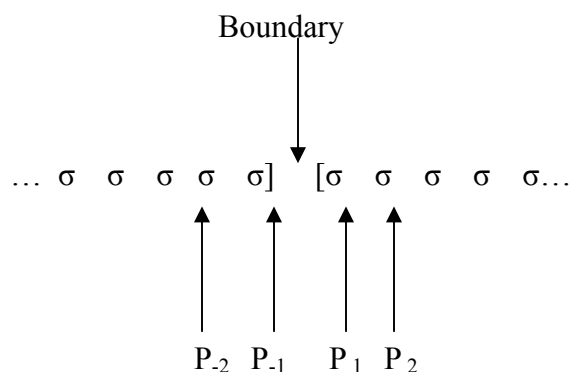
4. Data Analysis and Discussions

The purpose of this study is to examine whether there is any difference in temporal cues of prosodic boundaries between L1 and L2 Mandarin speech. Thus, statistical analysis is mainly conducted on each group respectively.

4.1. Syllable Duration

The following Table 3 shows the syllable duration for native Mandarin speakers and American learners.

Table 3: Syllable duration for Native Mandarin speakers and American learners (ms)

	Min	Median	Mean	Max
Native Chinese	0.04156	0.15837	0.16896	0.57516
Amercn learner	0.03919	0.17824	0.19436	0.77118

Before conducting ANOVA analysis on each group, an overall analysis of variance (ANOVA) was conducted on the syllable duration data. The ANOVA results show that: 1) there are main effects for native language (F=5.408, p-value=0.020), and position (F=3.349, p-value= 0.018); 2) no significant is found on disjuncture level; 3) there are interaction effects between native language and position (F=5.362, p-value=0.001). Post-hoc Bonferroni test was conducted on position and it showed that the differences between P-2 and P-1, between P-2 and P1, between P-2 and P2, between P-1 and P1 and P-1 and P2 are significant (p-values are 0.000,0.001,0.001, 0.000, 0.000).

Thus we can conclude that both syllable before the boundaries are lengthened and the last syllable is lengthened the most in both groups of speakers. However, syllable durations are not significantly different across disjuncture levels (DSP0, DSP1, and DSP2).

4.1.1. Native Mandarin Chinese Speakers

Analysis of Variance on syllable duration data of native Mandarin Chinese speakers shows that there is main effect of position at DSP1 and DSP2 (P-value = 0.00

234

and 0.00 respectively). The follow-up post-hoc test results show that :1) at DSP1, P-2 and P-1, P-2 and P1, P-2 and P2, P-1 and P1 and P-1 and P2 are significantly different in syllable duration (*p*-values are 0.000, 0.013, 0.047, 0.000 and 0.000 respectively); however, syllable duration for P1 and P2 are not significantly different; 2) at DSP2, P-2 and P1, P-2 and P2, P-1 and P1 and P-1 and P2 are significantly different in syllable duration (*p*-values are 0.000, 0.000, and 0.000 respectively); however the difference between P-2 and P-1 is only marginally significant (*p*-value=0.053); P1 and P2 are not significantly different.

Across syllable position, ANOVA shows that there is no significant difference for the syllable duration at the same position across disjuncture levels, which might suggest that even though there is final lengthening, the degree of lengthening does not index/reflect the discourse structure.

In summary, for the speech of native Mandarin Chinese speakers, we can see clearly the final lengthening of P-2 and P-1 and P-1 is lengthened the most. However, syllable duration across position does not seem to reflect the discourse disjuncture levels. The relationship between syllable position and disjuncture level can be seen clearly in Figure 1.
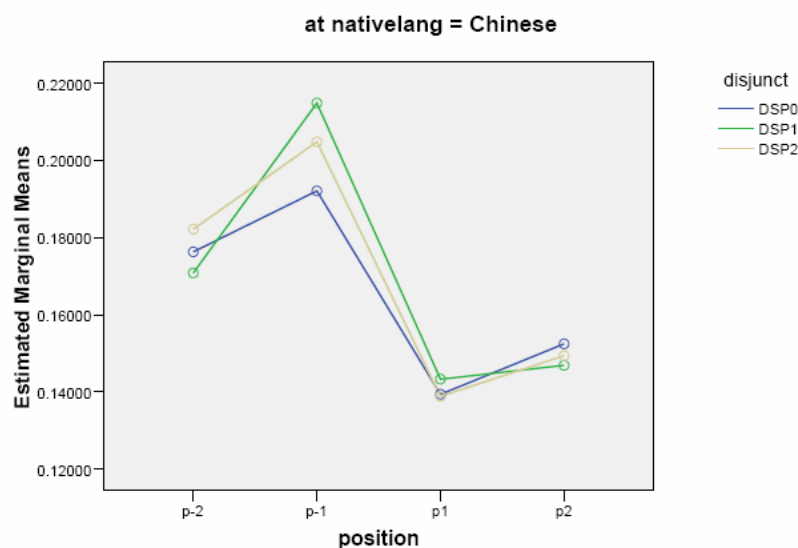

Figure 1: Profile plot of syllable duration of speech of native Mandarin speakers across position and disjuncture level

4. 1. 2. Advanced American Learners

For advanced American learners, ANOVA shows that at DSP0, DSP1 and DSP2, the syllable durations at P-2, P-1, P1 and P2 are not significantly different at all. And apparently the syllable durations across position does not register the discourse disjuncture

level. Figure 2 below shows syllable duration data of advanced American learners across syllable position and across discourse disjuncture level.

From Figure 2 it seems that at DSP0, syllable durations should be significantly different, at least between P-2 and P1. However, the statistical analysis shows that the seemingly greater different is not significant. The nature of the spontaneous speech might account for this difference, namely American learners tend to lengthen their syllables more while thinking. However, at both DSP1 and DSP2 in the above plot indicates that there is some final lengthening, even though not statistically significant.
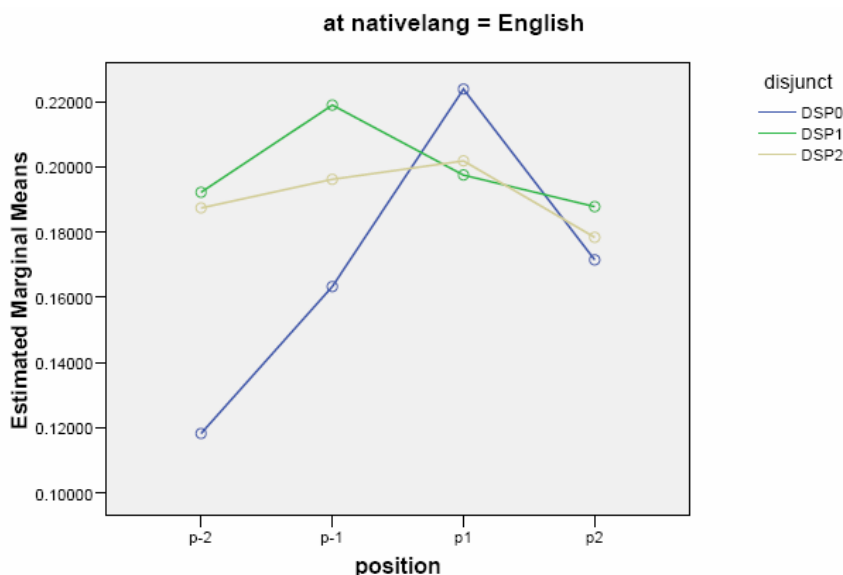

Figure 2: Profile Plot of syllable duration of speech of advanced American learners across syllable position and disjuncture level

4.1.3 Discussion

Incorporating the results in 4.1.1. and 4.1.2, we find that there is no syllable lengthening for the speech of American learners at all disjuncture levels, whereas for native Mandarin Chinese speakers, there is clearly final lengthening at P-2 and P-1 at DSP1 and DSP2 junctures. The findings of native Mandarin speakers conform to what Fon (2002) has found about Mandarin Chinese. However, as for the results from the advanced American learners, it is different from both the cues of English and the cues of Mandarin Chinese that Fon (2002) has found.

4.2. Syllable Onset Interval (SOI)

An overall analysis of variance was conducted on SOI data. The results show that: 1) there is main effect for disjuncture (F=33.673, p-value= 0.00) and position (F=92.210, p-value=0.00); 2) there is also interaction effect between disjuncture level and position

(F=33.413, *p*-value=0.00). Figures 3 and 4 below show the SOI of two groups of speakers.

However, one thing to note that the SOI data differ from syllable duration data primarily at P-1 position because there is nearly no pause in other positions (P-2, P1 and P2), except for after some syllables in the speech of native Mandarin speakers. Thus it is predicted that SOI at P-1 is the longest across different levels. Actually the effect of SOI is roughly the same as that of pause duration. Fon (2002) found that at P-1 in Mandarin Chinese, SOI can reflect the discourse disjuncture level. Similar findings were observed here. But I argue that this effect is more from pause duration than from SOI. Thus no further analysis was conducted on SOI.
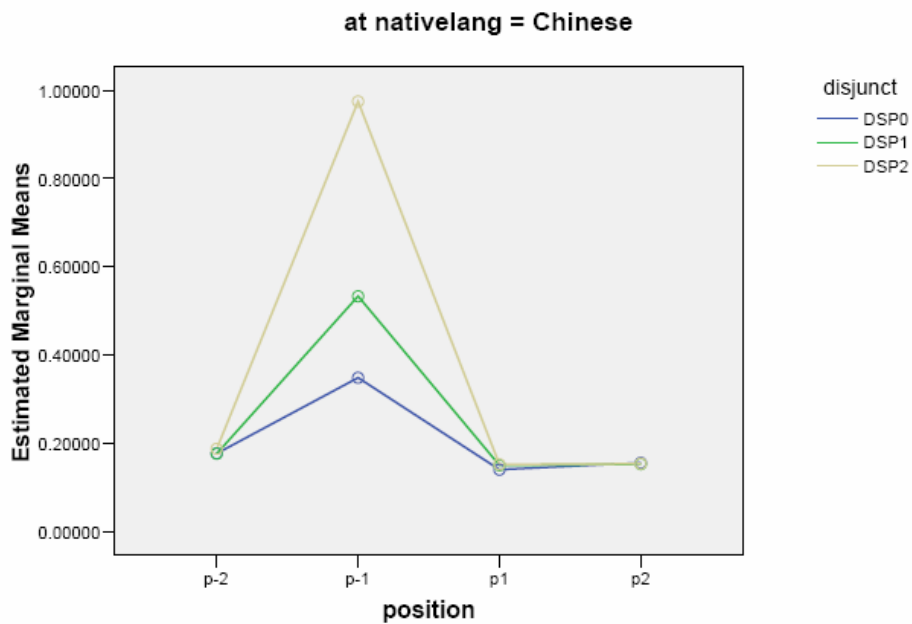

Figure 3: Profile Plot of SOI of speech of native Mandarin speakers across syllable position and disjuncture level
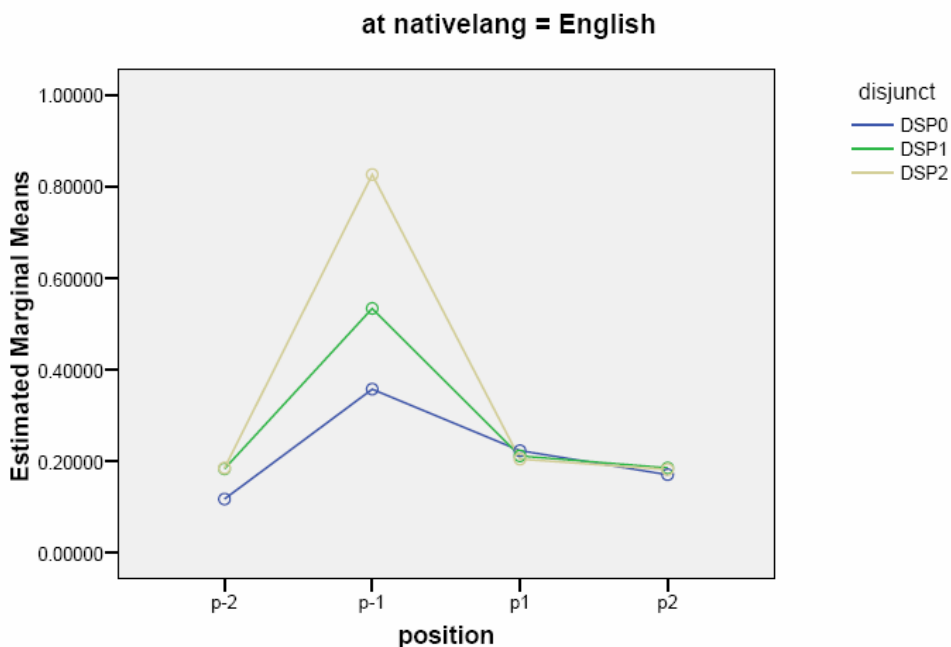
Figure 4: Profile Plot of SOI of speech of advanced American learners across syllable position and disjuncture level

4.3. Silent Pause

Table 4: Pause duration for native Mandarin speakers and American learners (ms)

	Min	Median	Mean	Max
Native Chinese	0.03932	0.54251	0.52238	2.25616
American learner	0.0374	0.5555	0.6227	2.4312

As mentioned above, pause predominantly occurred at prosodic boundaries. Thus the data for pause duration are only concerned with disjuncture level, not with syllable position. From Table 2 we can see that pause duration for American learners is generally longer than that for native Mandarin speakers.

An overall analysis of variance shows that there is significant difference in pause duration between DSP1 and DSP2 across groups (p-value=0.000) and that there is no significant difference between DSP0 and DSP1 in pause duration. The following profile plot illustrates the pause duration across disjuncture levels of two groups.
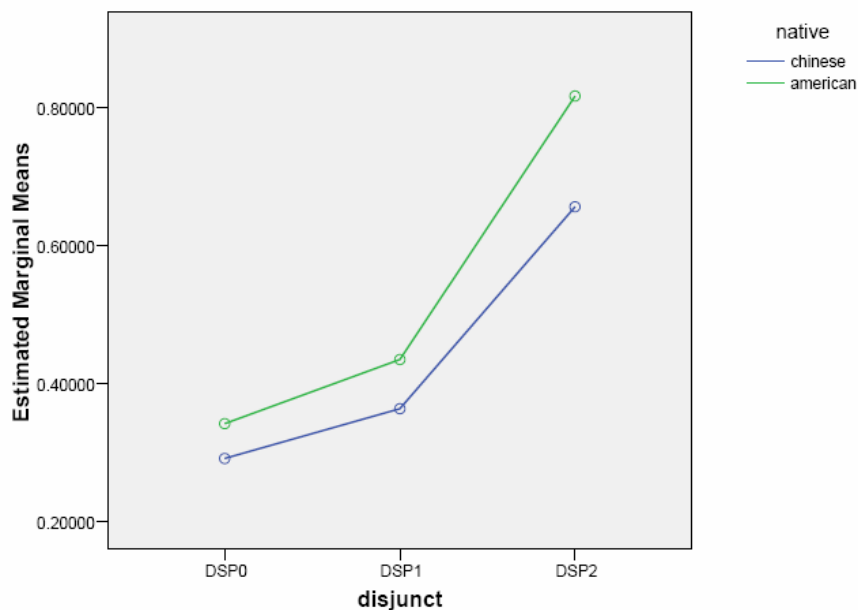
Figure 5: Profile plot of silent pause duration across language group and disjuncture level

Post-hoc tests indicate that for native Mandarin Chinese speakers, pause duration is significantly different between DSP1 and DSP2 (*p*-value=0.000), whereas for American learners, pause duration is significantly different between DSP0 and DSP1 (*p*-value=0.000) and DSP1 and DSP2 (*p*-value=0.000).

Thus the results here suggest that pause duration can reflect the discourse disjuncture levels of the speech by American learners at all three levels, whereas for native Mandarin speakers pause duration can only differentiate DSP1 and DSP2.

One thing to note that according to the criteria laid out previously in the paper, there is only a small number of DSP0, which might be problematic. In addition, as for advanced American learners, they tend to pause longer in their spontaneous speech. It is worth mentioning that Fon (2002) found that SOI is a better indicator of discourse disjuncture level. However, the present analysis shows that pause instead of SOI can reflect discourse disjuncture levels better.

4. 4. Summary

To make convenient comparison, the above findings are summarized in Table 5 and Table 6.

We can see there are many differences in the temporal cues of prosodic boundaries between L1 and L2 Mandarin speech. For native Mandarin speakers, final syllable lengthening is a salient cue of prosodic boundaries and the syllable at P-1 is lengthened the most. But the syllable lengthening does not reflect the discourse disjuncture levels. In addition, silent pause duration at the prosodic boundaries is another cue in L1 Mandarin

speech. Also the pause duration can reflect the discourse disjuncture level between DSP1 and DSP2. For advanced American learners, final syllable lengthening is not a salient cue of prosodic boundaries. However, we do see that there is final syllable lengthening at DSP1 and DSP2, though not statistically significant. And silent pause duration is a salient temporal cue in L2 Mandarin speech, which can reflect the discourse structure at all three levels.

Table 5: Significant pairs of syllable duration: across disjuncture level and native language

	Mandarin speakers	American learners
DSP0	No.	No.
DSP1	P-2 vs. P-1, P-2 vs. P1	No.
	P-2 vs. P2, P-1 vs.P1	
	P-1 vs. P2	
DSP2	P-2 vs. P1, P-2 vs. P2	No.
	P-1 vs. P1, P-1 vs. P2	

Table6: Significant pairs of pause duration: across language group at P-1

	Mandarin speakers	American learners
P-1	DSP1 vs. DSP2	DSP0 vs. DSP1
		DSP1 vs. DSP2

As mentioned above, the insignificance of final syllable lengthening in L2 Mandarin speech might be attributed to the fact that the mean syllable duration in L2 Mandarin speech is significantly longer than that in L1 speech. There are many factors which can account for this. The most likely one is that L2 speakers, even though they are advanced learners, still have some difficult in spontaneous speech, especially in the case of retelling a movie, when they need not only to organize the language, but also to organize the content.

As compared with what Fon (2002) has found about English, it seems that advanced American learners did not transfer the temporal cues in English to their Mandarin at all. However, I argue that, even though L2 learners did not transfer their L1 temporal cues in their Mandarin production, the difference in temporal cues of prosodic boundaries in the above aspects might contribute to the perception of their "foreign accent", because the difference in the above cues might render L2 Mandarin speech sounding rather different from L1 Mandarin speech, e.g. in terms of melody or temporal arrangement, etc.

5. Conclusion and Implications

The present research complements previous studies at tonal levels by providing a fuller picture of the prosody of Mandarin produced by L2 learners. The difference in temporal cues of prosodic boundaries found in this study suggests that prosodic studies in SLA, instead of focusing on tones and intonation only, should be broadened to incorporate temporal cues of prosodic boundaries so as to better understand the causes for foreign accent and come up with better ways to reduce foreign accent of learners. Also the present study suggests that the difference in temporal cues of prosodic boundaries between L1 and L2 Mandarin speech is likely to account for the foreign accent in L2 speech from a new perspective.

However, the present study has some limitations. One of the potential problems is the criteria of discourse encoding. Though the criteria used here drew on previous studies, some changes have been made and there might be some errors in implementing them, especially considering that only the researcher himself did the encoding due to the time limitation. Also potential problem may lie in the way fillers and disfluencies were treated in deciding on the syllables to be measured. Thus, a more elaborate set of criteria should be set up to encode discourse structure in future study.

REFERENCES

http://pearstories.org/docu/maps.htm
http://www.fon.hum.uva.nl/praat/

BENT, TESSA. 2005. *Perception and Production of Non-native Prosodic Categories*. Northwestern University dissertation.

CHAFE, WALLACE. L. 1980. The pear stories: Cognitive, cultural, and linguistic aspects of narrative production. *Advances in discourse process* 3. Norwood, NJ: Ablex.

FLEGE, JAMES E. 1987. The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15.47-65.

FLEGE, JAMES. E. 1995. Second-language speech learning: Theory, findings, and problems. *Speech Perception and Linguistic Experience,* ed. by W. Strange, 233-277. Timonium, MD: York Press.

FON, YEE-JEA JANICE. 2002. A cross-linguistic study on syntactic and discourse boundary cues in spontaneous speech. Columbus, OH: The Ohio State University dissertation.

FON, JANICE. 2004. A cross-linguistic study of discourse and syntactic boundary cues in spontaneous speech in the time domain. *Proceedings of the 1st session of the 10th International Symposium on Spontaneous Speech: Data and Analysis*, 1-27. The National Institute for Japanese Language, Tokyo, Japan.

FON, JANICE, and JOHNSON, KEITH. 2000. Speech timing patterning as an indicator of discourse and syntactic boundaries. *Proceedings of 6th International Conference on Spoken Language Processing* 2.555-558. Beijing, China.

FON, JANICE, and KEITH JOHNSON. 2004. Syllable onset intervals as indicator of discourse and syntactic boundaries in Taiwan Mandarin. *Language and Speech* 47.1.57-82.

JONGMAN, ALLARD; YUE WANG; C. MOORE; and J.A. SERENO. 2006. Perception and production of Mandarin Chinese tones. *The Handbook of East Asian Psycholinguistics*, by P. Li, E. Bates, L.H. Tan, and O. Tseng (eds.), 209-217. Cambridge University Press.

LEE, OK JOO. 2005. *The Prosody of Questions in Beijing Mandarin.* Columbus, OH: The Ohio State University dissertation.

LIU, YABIN, and AIJUN LI. 2004. Cues of prosodic boundaries in Chinese spontaneous speech. *ICPHS 2003*, (August 2003), Barcelona.

MAREUIL, P. B.; and B. VIERU-DIMULESCU. 2005. The contribution of prosody to the perception of foreign accent. *Phonetica 63*.247-267.

SERENO, J.A., and Y. WANG. 2007. Behavioral and cortical effects of learning a second language. *Language experience in second language speech learning, in honor of James E. Flege*, ed. by O. Bohn and M. Munro, 239-258. John Benjamins.

SHEN, XIAO NAN. 1992. A pilot study on the relation between the temporal and syntactic structures in Mandarin. *Journal of the International Phonetic Association* 22.1-2. 35-43.

TSENG CHIU-YU. 2002. The prosodic status of breaks in running speech：Examination and Evaluation. *Speech Prosody 2002,* Aix-en-Provence, France (April, 11-13, 2002, pp. 667-670.

WANG, YUE; A. JONGMAN; and J.A. SERENO. 2006. L2 acquisition and processing of Mandarin Chinese tones. *The Handbook of East Asian Psycholinguistics*, ed. by P. Li, E. Bates, L.H. Tan, and O. Tseng, 250-256. Cambridge University Press.

YANG YUFANG, and BEI WANG. 2002. "Acoustic correlates of hierarchical prosodic boundary in Mandarin". *Speech Prosody 2002.*