

# Durational Patterning at Syntactic and Discourse Boundaries in Mandarin Spontaneous Speech

Language and Speech

54(1) 5–32

© The Author(s) 2010

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0023830910372492

las.sagepub.com



**Janice Fon and Keith Johnson**

National Taiwan University, Taiwan

**Sally Chen**

University of California at Berkeley, USA

## Abstract

This study focused on durational cues (i.e., syllable duration, pause duration, and syllable onset intervals (SOIs)) at discourse boundaries in two dialects of Mandarin, Taiwan and Mainland varieties. Speech was elicited by having 18 participants describe events in *The Pear Story* film. Recorded data were transcribed, labeled, and segmented into clauses. Discourse boundary indices were used to label discourse disjuncture levels. Results showed that the scope of lengthening included both the final and the penultimate syllables. Pause was a robust but optional indicator for discourse boundaries, and Taiwan Mandarin preferred unfilled over filled ones, while the Mainland variety did not show such strong preferences. Discourse hierarchy corresponded consistently with occurrences of pauses and duration of SOIs, within which pause was the main contributor. Higher discourse levels were more likely to be accompanied by pauses, and were indicated by longer pauses and SOIs. Syllable duration only played a secondary role in indicating discourse disjuncture size when pause was absent. When there was an accompanying pause, Mainland Mandarin relied solely on it to indicate discourse hierarchy, while Taiwan Mandarin used both syllable and pause duration, a dialectal difference that seemed to be rhythm-related. By shortening the degree of lengthening in the boundary syllable and lengthening the following pause at the same time, Taiwan Mandarin increased the absolute and relative duration of pause and maximized its role in indicating discourse hierarchy. The results of this study implied that the use of pause in discourse disjuncture demarcation was more of a language-specific choice while its role in discourse hierarchy encoding is more language-universal.

---

## Corresponding author:

Janice Fon, Graduate Institute of Linguistics, National Taiwan University, 1 Roosevelt Rd. Sec. 4, 317 Lexue Bldg., Taipei, Taiwan, ROC 10617

Email: jfon@ntu.edu.tw

## Keywords

discourse structure, final lengthening, hierarchical encoding, pause duration, spontaneous speech

## Introduction

Unlike written English and many other written languages using phonetic alphabets of some kind, where words are conveniently separated by spaces, sentences by periods, and paragraphs by paragraph breaks, spoken languages seem to be more like Chinese writings, where no apparent word boundary cues are provided, and sometimes even more like ancient Chinese writings, where both word and sentential boundaries are left unmarked. Therefore, the fact that listeners can segment incoming stretches of speech with graceful ease is phenomenal. It leads many researchers to suspect that some spoken equivalents of punctuation marks might actually exist in speech, but are nevertheless obscured by other production phenomena such as repetitions, self-corrections, slips-of-the-tongue, etc.

One type of cue that has been proposed to be robust in indicating sentential and discourse boundaries is in the dimension of time, whose main phonetic realization is pause and final lengthening. Goldman-Eisler (1972) and Shen (1992) showed that there was a correlation between the existence of a silent pause and clausal boundaries in read and prepared speech. Grosjean (1980) further found that breathing pauses are more likely to occur at major syntactic boundaries (i.e., sentential and clausal) while non-breathing ones are more likely to occur at minor syntactic boundaries (i.e., phrasal and word level). Breathing in pauses can be annihilated by faster speech rate only at minor but not major boundaries, implying that it is at least in part required in speech planning. The existence and the duration of silent pause are also indicative of structural hierarchy. Pauses are more likely to occur at sentential boundaries than clausal ones (Stenström, 1986). In addition, longer pauses are more likely to follow phrases of higher attachments (O'Malley, Kloker, & Dara-Abrams, 1973).

Silent pause is also found to be a robust cue for discourse unit delimitation. In spontaneous speech, discourse-initial phrases are preceded by longer pauses and followed by shorter ones as compared to other utterance-initial phrases, while discourse-final phrases have shorter preceding pauses and longer subsequent ones compared to other utterance-final phrases (Grosz & Hirschberg, 1992; Hirschberg & Nakatani, 1996; Swerts & Geluykens, 1993; Swerts & Ostendorf, 1997). Similar results have also been found in dialogs (Traum & Heeman, 1997). Longer pauses are more likely to follow utterances related to the immediate preceding utterance by another speaker than ones unrelated. In addition to segmenting discourse boundaries, pause is also used for topic demarcation. In Dutch, topic-introducing phrases or clauses are shown to be flanked by pauses, with the preceding pause being longer than the following one (Swerts & Geluykens, 1993, 1994).

The study of final lengthening also has a long tradition. As early as the 1970s, researchers have shown that lengthening in English tends to fall on sentence-final units, including the rhyme (Campbell & Isard, 1991; Oller, 1973) and the foot (Lehiste, 1975; Scott, 1982). As for Mandarin, a language that is prosodically different from English, Shen (1992) found the syllable to be the scope. Final syllables were lengthened by about 47% compared to their non-final counterparts.

In order to account for the often-concurrent and interdependent nature of pause and lengthening, some researchers have devised indices to accommodate both. Ladd (1988)

used “boundary duration”, defined as the interval between the onset of the last stressed syllable preceding a clausal boundary and the onset of phonation of the first syllable after the boundary. In other words, this would include the duration of the final stressed syllable, the following unstressed syllables if there are any, and the following pause if there is any. Ladd found that boundary duration was a good indicator of hierarchical organization of clauses in English. Clausal boundaries of higher disjuncture had longer boundary duration than those of lower disjuncture.

In order to study Taiwan Mandarin, a language that is more syllable-than stress-timed as compared to English (Duanmu, 2002), Fon and Johnson (2004) used a slightly modified measure of syllable onset interval (SOI), defined as the interval between two contiguous syllable onsets, which would include the duration of the syllable itself and any potential following pause. Similar to Ladd (1988), they also found the measure to be highly reflective of the hierarchical organization of discourse and syntax. Discourse-final SOIs are longer than clause-final ones, which are in turn longer than phrase-final ones. In addition, there is also a mild lengthening effect in non-final SOIs. Unit-internal ones are slightly longer than unit-initial ones at all three levels. Based on the findings, Fon and Johnson concluded that the effect of hierarchical encoding is fairly robust, despite active performance factors in spontaneous speech.

The high correlation between timing and structure is to a large extent mediated by prosody. Syntactic and discourse boundaries are likely to be accompanied by prosodic boundaries, which are often signaled by special pitch contours, final lengthening, and pause, among others (Butcher, 1980; Grosz & Hirschberg, 1992; Hirschberg & Nakatani, 1996; Keller, Zellner, Werner, & Blanchoud, 1993; Ladd & Campbell, 1991; Stenström, 1986; Zellner, 1994). However, as prosodic phrasings are also influenced by other non-syntactic factors such as individual speaking styles, speech rate, the interval between one prosodic boundary and the next, etc., not all prosodic boundaries coincide with syntactic and discourse boundaries (Frazier, Carlson, & Clifton, 2006; Ladd, 1996). If prosodic boundaries that coincide with major syntactic and discourse boundaries are ones that would definitely occur, then one could view a speech flow as being segmented into several obligatory prosodic phrasings which may potentially be interspersed with optional prosodic boundaries. In this study, we are mainly interested in the obligatory prosodic boundaries that accompany major syntactic and discourse boundaries and how they reflect these boundary strengths through phonetic realizations. In particular, we will mainly deal with acoustic cues from the time domain, even though pitch movements are also important cues for boundary demarcation, since pitch contours are largely dependent on accent types and lexical tonal categories (as in the case of Mandarin), and are thus highly variable. This is probably also why pause and duration are shown to be the two most prevalent cues in machine-segmenting spontaneous speech, which naturally displays a wide spectrum of accent types, while pitch features are only reliable for news speech, a genre that includes only a small inventory of pitch accent types (Shriberg, Stolcke, Hakkani-Tur, & Tur, 2000).

Based on the above studies, it is clear that pause and final lengthening are robust indicators for clausal and sentential boundaries. In terms of discourse boundaries, however, pause seems to be a stronger cue than lengthening.

## Specific aims

There are three specific aims in this study. The first is to investigate whether pause and syllable duration are good discourse boundary indicators in Mandarin spontaneous speech. Previous studies

showed that both cues are important in standing syntactic ambiguities (Shen, 1992),<sup>1</sup> and a combination of the two (i.e., SOI) is a robust indicator of hierarchical organization in speech among phrasal, syntactic, and discourse levels (Fon & Johnson, 2004). However, it is unclear whether discourse boundaries are encoded equally by both temporal measures in spontaneous speech, especially when standing syntactic ambiguity is not an issue, and if so, whether it is encoded in the same manner. If Mandarin has a similar temporal encoding pattern of discourse boundaries as Indo-European languages, then one would expect to find pause to be a more indispensable cue than lengthening, and most of the discourse boundaries should be accompanied by pauses. On the other hand, if the obligatory occurrence of pause is due to the need to disambiguate different ways of parsing, then lengthening might be enough for encoding discourse boundaries in regular non-ambiguous speech, leaving pause an optional mechanism.

The second aim of the study one would like to look into is whether temporal cues can be indicative of discourse hierarchy. Previous studies looking at hierarchical encoding mainly focused on syntactic organizations (Grosjean, 1980; O'Malley et al., 1973), and for those that looked into discourse hierarchy, they mainly compared syntactic units that are at discourse boundaries with those that are not, and did not further look into potentially different layers of discourse hierarchical levels (Fon & Johnson, 2004; Grosz & Hirschberg, 1992; Hirschberg & Nakatani, 1996; Ladd, 1988; Traum & Heeman, 1997). However, discourse hierarchy is not a discrete, all-or-none variable, but should instead be more like syntax, where boundaries of various sizes exist (Grosz & Sidner, 1986). Therefore, it is quite possible that temporal encoding of speech would also reflect the hierarchy and incorporate such gradation into the system, as is suggested by Swerts and Geluykens (1994). If such a relationship exists, then one would expect discourse boundary strength to be mirrored by acoustic strength, be it final lengthening, pause, or SOI, the realization of which being a (usually positive) correlation between the two at boundary positions.

The third specific aim one would like to study is the dialect effect. Mandarin is the official language of both Taiwan and Mainland China. However, due to 60 years of political separation between the two places, the two Mandarins have developed independently so that dialectal variations are obvious to speakers of either variety (Cheng, 1985). Lexicon aside, the most prominent difference probably lies in rhythm and prosody. Mainland Mandarin more often distinguishes between stressed and unstressed syllables (Lin, 1985), while Taiwan Mandarin has more equally stressed syllables except for perhaps particles and function words (Duanmu, 2002; Hsu, 2006; Kubler, 1985). For example, *dong<sup>1</sup>-xi<sup>1</sup>* 'east and west' and *dong<sup>1</sup>-xi<sup>0</sup>* 'thing' are distinguished by stress in the second syllable in Mainland Mandarin, but are both indistinguishably pronounced as *dong<sup>1</sup>-xi<sup>1</sup>* in the Taiwan variety.<sup>2</sup> In other words, on the rhythmic spectrum, Mainland Mandarin is more stress-timed while Taiwan Mandarin is more syllable-timed, resulting in a somewhat faster speech tempo in the former dialect.<sup>3</sup> If prosodic and rhythmic differences can influence encoding of discourse organization, then one would expect to find boundary signals to pattern differently in the two dialects.

## Methods

### Subjects

Eighteen native speakers of Mandarin were recruited in this study. Half of them were from the Taipei metropolitan area in Taiwan (age:  $M = 28.33$ ,  $SD = 1.5$ ), and half of them were

from the Beijing metropolitan area in China (age:  $M = 29$ ,  $SD = 3.74$ ). There were 5 female and 4 male speakers in each group. Contact with other Chinese languages was essentially unavoidable, as all of the Taiwan subjects have at least passive knowledge of Min, a powerful substrate language on the island.<sup>4</sup> One Taiwan subject speaks also Shanghaiahua, and one Mainland subject speaks also Anhuihua and Pekingese. In addition, all of the subjects are proficient in English, as it is a mandatory subject in high school education. However, regardless of subjects' language backgrounds, Mandarin was the dominant language for all of them. On a scale of 1 to 7 (1 = very influent; 7 = very fluent), they had an average self-rated score of 6.89 ( $SD = 0.32$ ).

## Materials

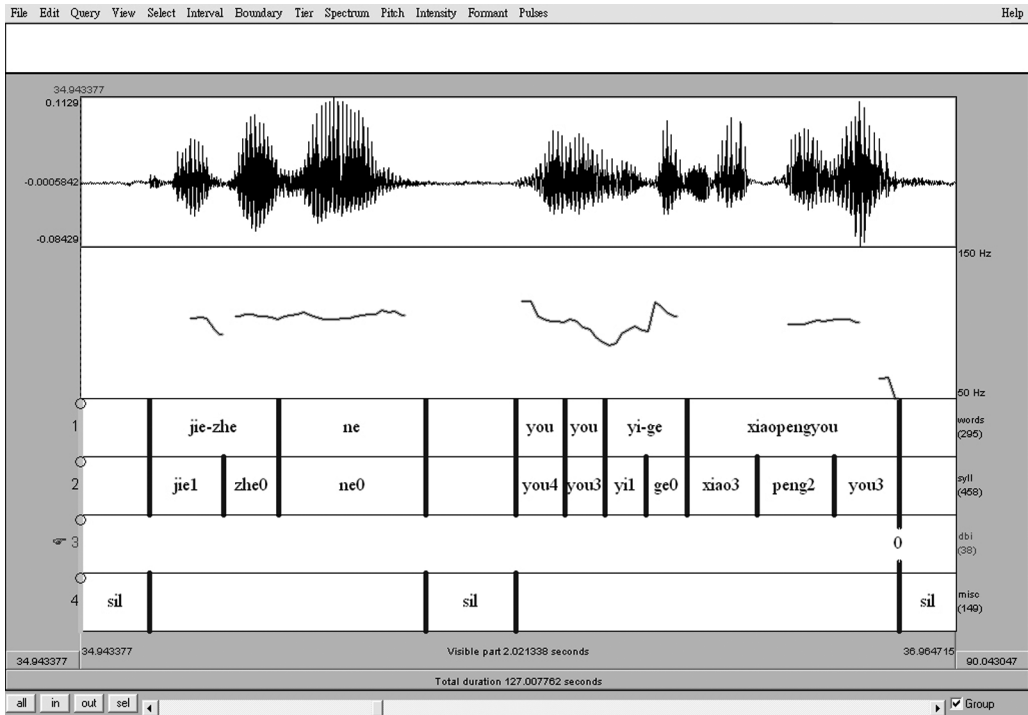
In order to facilitate spontaneous speech elicitation from speakers of the two dialects, *The Pear Story* film (Chafe, 1980) was used. A short movie without any lines, the story is proven to be relatively culture-free and is successful in eliciting speech in various languages of various cultural backgrounds (e.g., English: Kärkkäinen, 1996; Mandarin: Erbaugh, 1990; Japanese: Clancy, 1980; Polish: Pakosz & Flaschner, 1988). This is important in this study considering the possible cultural differences between Taiwan and China.

## Equipment

Recordings were done with a SHURE SM10A head-mounted microphone connected to a SONY DAT DTC-790 recorder through a Symetrix SX202 Dual Mic Preamp preamplifier using Maxell R-64DA DAT tapes. The recording sampling rate was set at 44,100 Hz. A D-to-D transfer from the DAT recorder to the computer was performed using CoolEdit, and the speech signal was later downsampled to 22,050 Hz for further analyses.

## Procedure

Subjects were tested individually in a quiet room. To avoid any accommodation effect that subjects might have, they were tested by a (near-)native speaker of the target dialect. The first author, a native speaker of Taipei Mandarin, did the recordings for the Taiwan speakers, while a fluent near-native speaker of Beijing Mandarin recorded the Mainland subjects. Subjects were first shown the short film and were then asked to describe it afterwards as if talking to a friend. Most of the subjects had no trouble with the instruction. Some subjects told the stories more than once due to various reasons. Subject WCF requested to do a practice story-telling first before the actual recording. Therefore, her second try was used for later analyses. Subject HPH recorded three times because he did not like the first try due to its informal register. The second try was self-interrupted and the third try was rather formal. The first recording was used for further analyses. Subject FBS also recorded three times. The first time was interrupted by the experimenter due to its very formal and literary register. He was told to restart. The second try was somewhat better, but still formal. The subject was more at ease in the third try, which was used for further analyses. The experiment took about 20–30 min, and monetary rewards were given to each subject. Appendix A shows the duration of the recording of each subject.



**Figure 1.** An illustration of labeling using Praat. Please see the text for description

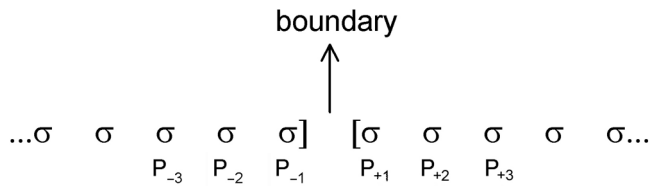
## Transcription and labeling

Recordings were orthographically transcribed in Pinyin romanization by the first author and hand-labeled using Praat (Boersma & Weenink, 2007). A labeling example is shown in Figure 1. There were two orthographic transcription layers, the words tier and the syll tier, which provided the word-level and the syllable-level transcription, respectively. Due to the syllable-per-character writing system in Mandarin, the definition of a word is a matter of debate. This study adopted the definition provided by Li and Thompson (1981) as a convenient solution.

## Measurements

After labeling, durational measurements of intervals corresponding to syllables and pauses were extracted using customized Praat scripts. Mandarin syllables are fairly easy to identify phonologically and the most complicated syllable structure is CGVC, where the coda consonant is either a nasal or a glide (Lin, 1989).

Silent pause duration was measured whenever there was a perception of such by the transcriber, instead of setting an arbitrary cutoff point. Since speech rate fluctuates to a great extent in



**Figure 2.** A schematic illustration of the relevant syllables chosen around the boundary region

spontaneous speech, it is more reasonable to judge silent pause in context and use trained ears as a way of normalization to compensate for what a fixed cutoff point could not have achieved. Silent pause was labeled as *sil* in the misc (= miscellaneous) tier, as shown in Figure 1. Both syllable and pause were measured in terms of milliseconds. As a result, three duration measures were taken, the syllable duration, the pause duration, and the SOI (Fon & Johnson, 2004), which includes essentially the duration of a syllable and whatever pause that follows it.

Since what is of interest in this study is discourse boundary cues, syllables and SOIs at the boundary region, three before and three after, were chosen as the target for analyses. The position immediately before the boundary was designated  $P_{-1}$  (P for position), and the two positions before that were then conveniently referred to as  $P_{-3}$  and  $P_{-2}$ . Similarly, the position immediately after the boundary was designated as  $P_{+1}$ , and the two positions after that were referred to as  $P_{+2}$  and  $P_{+3}$ . The reason for including six syllables in the boundary region for analyses was twofold. First of all, although previous studies showed that final lengthening is the most effective on the last rhythmic unit, be it the rhyme (Campbell & Isard, 1991; Oller, 1973), the syllable (Shen, 1992), or the foot (Lehiste, 1975; Scott, 1982), it is unclear whether its scope would also be limited to such units only, and whether it would be the same for the two rhythmically distinct dialects. Secondly, syllables that are not under the effect of final lengthening, such as those at  $P_{+1}$ ,  $P_{+2}$ , and  $P_{+3}$ , could serve as reference measures for evaluating the magnitude of the lengthening effect.<sup>5</sup> A schematic illustration of the six positions chosen for further analyses is shown in Figure 2.

### Discourse labeling

Since sentence is difficult to define in spontaneous speech, clause was used as the basic unit in discourse labeling. A clause was usually defined by having at least one main verb. However, since Mandarin has no morphological markings differentiating serial verbs from sequences of main verbs, it is sometimes difficult to segment utterances into clauses. In this study, the guidelines laid out by Li and Thompson (1981) were adopted. All utterances in the database were segmented into clauses by the first and the third authors independently, both of whom are native speakers of Mandarin. The resulting Cohen's (1960)  $\kappa$  of 0.94 indicated high inter-rater agreement,  $p < .0001$ . As a consequence, segmentation by the first author was adopted for further analyses.

After clausal segmentation, discourse relationship between every pair of adjacent clauses was determined based on Grosz and Sidner's (1986) computational discourse framework. There are



three separate but interrelated components in their discourse model—the linguistic structure, the intentional structure, and the attentional state. What is of concern in this study is the intentional structure, which indicates the goals that the talker intends to achieve when conducting a particular discourse. For example, narrating a movie may achieve the goal of describing the story. This goal is fairly broad and has to be broken down into several subgoals, and maybe even sub-subgoals (e.g., describing the cast, the plot, evaluation of the story, etc.). For each (sub-)subgoal, there will be a discourse segment fulfilling the intention, called the discourse segment purpose.

Since this study used *The Pear Story* movie for narrative elicitation, and movies generally involve many scenes, each scene was conveniently considered a subgoal of the intentional structure of discourse. In addition, narration devices related to the structuring of the story, such as the introduction and the ending, were also deemed as subgoals. In total, there were nine subgoals identified in the corpus (Table 1). Within each subgoal, there were usually more than one sub-subgoals. For example, the pear-picking scene would usually elicit not only narrations about the farmer picking pears, but also descriptions of the farmer and even those of the pears. Evaluations of the scene were also fairly common, as well as projected thoughts to the characters.

Based on how adjacent clauses were related to each other in terms of their discourse segment purposes, three levels of discourse boundary indices (DBIs) were devised.  $DBI_0$  refers to the lowest possible degree of disjuncture between two adjacent clauses in a discourse. Namely, no discourse level disjuncture exists. The only boundary is the syntactically defined clausal boundary. In other words, the two flanking clauses belong to the same subgoal and the same sub-subgoal. An example is shown in (1). The two clauses share the same subgoal of the pear-picking scene, and the same sub-subgoal of describing the farmer picking pears. Therefore, the clausal boundary was designated  $DBI_0$ . In Mandarin,  $DBI_0$  is often signaled by zero anaphora. Although  $DBI_0$  does not entail a discourse boundary, it is still relevant because it represents a potential discourse disjuncture and could serve as a baseline for higher levels of discourse hierarchy.

(1) The pear-picking scene: the farmer picking pears

[ <i>ta</i>	<i>jiu</i>	<i>pa</i>	<i>dao</i>	<i>shu-shang</i>	<i>a</i> ] <sub>DSP<sub>0</sub></sub>	[ <i>qu</i>	<i>cai</i>	<i>xiyangli</i>	<i>a</i> ]
[3 <sup>rd</sup> sg.	then	climb	to	tree-above	PART:RF <sup>6</sup> ] <sub>DBI<sub>0</sub></sub>	[go	pick	pear	PART:RF]
[he then climbed up the tree] <sub>DBI<sub>0</sub></sub> [went to pick pears]									

$DBI_1$  refers to the next higher level of disjuncture. It is designated for a boundary break between two adjacent clauses that is interrupted by not only a clausal boundary, but also a small discourse boundary. The two clauses communicate two different sub-subgoals but share the same subgoal of the intentional structure. An example is shown in (2). The two clauses here do not share exactly the same discourse purpose. The first describes the stealing action of the main character, while the second conveys the talker's own evaluation of action. However, the two are related in purpose because the second clause, although not a description of the stealing event itself, is a comment on the event and also contributes to the pear-stealing scene.

(2) The pear-stealing scene: pear-stealing + comment

[ <i>jiu</i>	<i>ba</i>	<i>ta</i>	<i>zai-zou-le</i> ] <sub>DBI<sub>1</sub></sub>	[ <i>zhe-ge</i>	<i>jiao</i>	<i>touqie</i>	<i>ba</i> ]
[then	ACC <sup>7</sup>	it	carry-walk-PFV <sup>8</sup> ] <sub>DBI<sub>1</sub></sub>	[this-CL <sup>9</sup>	call	stealing	PART:SA <sup>10</sup> ]
[(he) then carried it away] <sub>DBI<sub>1</sub></sub> [this is called stealing, no?]							



**Table 1.** The general intentional structure of *The Pear Story* narratives

Scene/subgoal	Content/sub-subgoal
1. Introduction	(1) The talker declaring that he is going to tell a story (2) The talker's overall comment about the story
2. Background	(1) The setting of the story (2) Other background information
3. Pear-picking	(3) Other related comments/evaluations (1) The farmer picking pears (2) Description of the farmer (3) Description of the pears (4) Other related comments/evaluations
4. Goat	(1) The goat herder and the goat passing by (2) The farmer's reaction (3) Description of the goat (4) Other related comments/evaluations
5. Pear-stealing	(1) The boy stealing the pears (2) Description of the boy (3) Description of the bicycle (4) Other related comments/evaluations
6. Fall	(1) The boy running away (2) The boy meeting the girl (3) The boy losing the hat (4) The boy falling (5) Other related comments/evaluations
7. Help	(1) The boy's reaction to the fall (2) The three children helping the boy (3) The boy giving the three children pears (4) Other related comments/evaluations
8. Suspicion	(1) The three boys walking by the pear tree (2) The farmer's reaction (3) Other related comments/evaluations
9. Ending	(1) The talker declaring the ending of the story (2) Other related comments/evaluations

DBI<sub>2</sub> refers to a high level of disjuncture. Similar to DBI<sub>1</sub>, DBI<sub>2</sub> is designated for a boundary break between two adjacent clauses that is interrupted by both a clausal and a discourse boundary. However, the two clauses are devoted to different sub-subgoals and subgoals. Example (3) is an illustration of DBI<sub>2</sub>. The first clause describes the pear-stealing scene while the second clause describes the falling scene. In Mandarin, DBI<sub>2</sub> is often indicated by a full anaphora and a different subject.

(3) The pear-stealing scene + the fall scene

[zhunbei	zai-hui	jia	qu] <sub>DBI<sub>2</sub></sub>	[jieguo	lu-shang	you	yi-ge	ren]
[prepare	carry-	home	go] <sub>DBI<sub>2</sub></sub>	[as	a road-on	exist	one-CL	person]
	return			result				

[(he) prepared to carry (it) home]<sub>DBI<sub>2</sub></sub> [then on the road there was a person (i.e. the girl)]

Since DBIs were determined by comparing adjacent clauses, narration-final clauses were excluded from such assignment, as no following clauses existed. As shown in Figure 1, DBIs were labeled numerically after each clause in the dbi tier. Using criteria described above, the first author labeled all the data. The third author also served as an independent labeler, and labeled 50% of the data. To avoid judgments based on prosodic cues, labeling for both authors was text-based. Tinsley and Weiss' (1975) T index showed a high inter-labeler agreement of 0.99,  $p < .0001$ . As a result, the labeling of the first author was adopted for further analyses.

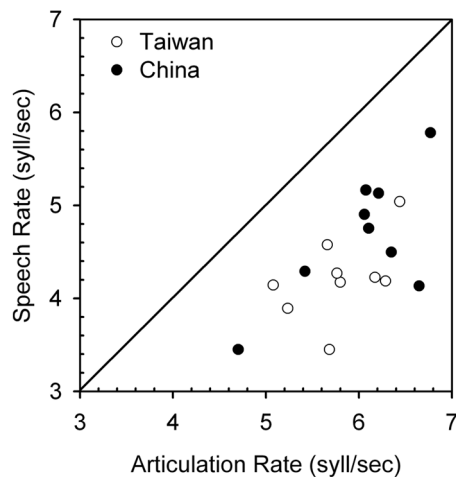
## Results

### Summary statistics

On average, Taiwan speakers uttered 862.67 syllables,  $SD = 576.97$ , and Mainland speakers uttered 759.78 syllables,  $SD = 340.36$ . Independent  $t$  test showed that there was no significant difference between the two groups,  $t(16) = .46$ , ns. For the number of syllables uttered by each speaker, please see Appendix A.

Figure 3 is a scatterplot of the speech rate and the articulation rate of the two groups of speakers. Speech rate was calculated by dividing the total number of syllables by the total amount of time spent inclusive of silent pauses, while articulation rate was defined by dividing the total number of syllables by the total amount of articulation time, excluding silent pauses. There was a medium correlation between the two measures,  $r(18) = .66$ ,  $p < .01$ . The speech rate of Taiwan Mandarin fluctuated between 3.45 syll/sec and 5.04 syll/sec, with an overall mean of 4.22 syll/sec, while the articulation rate fluctuated between 5.08 syll/sec and 6.44 syll/sec, with an overall mean of 5.79 syll/sec. On the other hand, the speech rate of Mainland Mandarin fluctuated between 3.45 syll/sec and 5.78 syll/sec, with an overall mean of 4.68 syll/sec, while the articulation rate fluctuated between 4.70 syll/sec and 6.77 syll/sec, with an overall mean of 6.03 syll/sec. The pattern was fairly comparable to previous studies (Fon & Johnson, 2004). For both measures, Taiwan speakers tended to be somewhat slower than those from China.

Table 2 shows the distribution of different levels of DBIs in the two dialects with regards to pausing patterns at clausal boundaries. Only instances that were free of hesitation pauses in the boundary region were included. This was conveniently achieved by excluding instances that contain pauses at locations other than after the boundary syllable. As can be seen from the table, there were more instances of  $DBI_0$  than any other categories. For the number of cases each subject contributed, please see Appendix B. In addition, most of the clausal boundaries were either without any pause ( $\sigma_{-1}$ ), or with an unfilled pause ( $\sigma_{-1}+UP$ ). With higher levels of DBIs, clausal boundaries were more likely to be accompanied by silent pauses (Taiwan:  $\chi^2(2) = 84.11$ ,  $p < .0001$ ; China:  $\chi^2(2) = 63.63$ ,  $p < .0001$ ).<sup>11</sup> Other combinations of filled and unfilled pauses did exist, but did not constitute significant proportions. Interestingly, pausing patterns were more varied for Mainland speakers than for Taiwan speakers. Patterns not found in Taiwan Mandarin, such as  $\sigma_{-1}+FP$  and  $\sigma_{-1}+FP+UP$ , occurred sporadically in Mainland Mandarin. In order to facilitate cross-dialectal comparisons, only  $\sigma_{-1}$  and  $\sigma_{-1}+UP$  cases were included for all subsequent analyses. Syllable, pause, and SOI duration were examined for the  $\sigma_{-1}+UP$  cases, while only syllable (= SOI) duration was examined for the  $\sigma_{-1}$  cases.



**Figure 3.** A scatterplot of speech rate and articulation rate of two groups of speakers

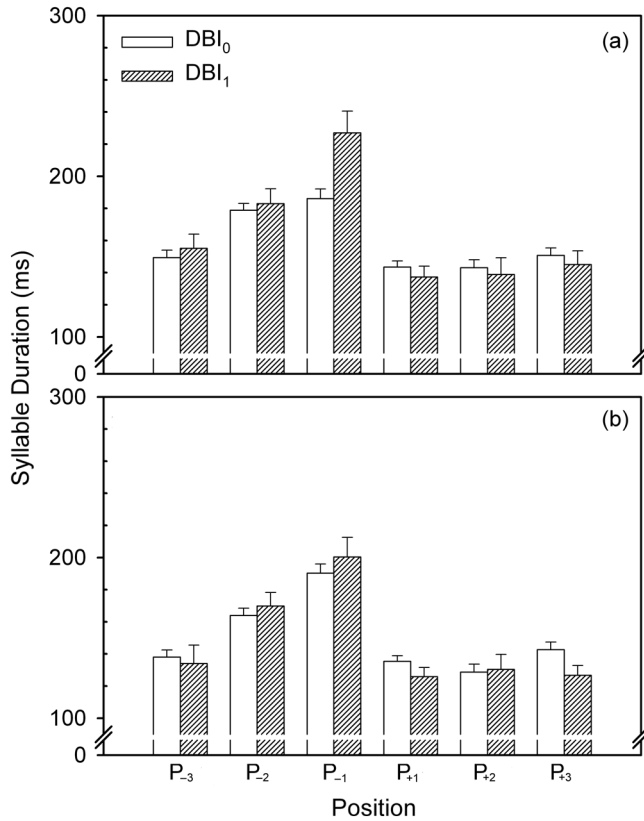
**Table 2.** Number of DBIs in the two dialects

Boundary syllable	Taiwan			China			Total
	DBI <sub>0</sub>	DBI <sub>1</sub>	DBI <sub>2</sub>	DBI <sub>0</sub>	DBI <sub>1</sub>	DBI <sub>2</sub>	
$\sigma_{-1}$	185	37	3	176	47	4	452
$\sigma_{-1}$ +UP	118	125	34	111	120	29	537
$\sigma_{-1}$ +UP+FP	3	2	0	13	13	6	37
$\sigma_{-1}$ +UP+FP+UP	1	1	1	3	9	7	22
Others	0	0	0	4	1	1	6
Total	307	165	38	307	190	47	1054

Note.  $\sigma_{-1}$ : the syllable at  $P_{-1}$ , right before the boundary; UP: unfilled pause; FP: filled pause.

### *Syllable duration at boundary region with $\sigma_{-1}$*

Figure 4 shows the duration patterning of  $\sigma_{-1}$  at the boundary region for both Mandarin dialects. DBI<sub>2</sub> was not included because there were too few cases and the results would not be reliable. The effect of syllable final lengthening was fairly consistent across the two discourse levels. Its scope included both  $P_{-1}$  and  $P_{-2}$ . In addition, the degree of lengthening showed a positive correspondence with discourse disjuncture at  $P_{-1}$  and  $P_{-2}$  for both varieties of Mandarin. Syllables became longer with higher discourse levels.



**Figure 4.** Syllable duration of (a) Taiwan and (b) Mainland Mandarin at the boundary region of  $\sigma_{-1}$ . The error bars represent standard error

In order to confirm the above observations, an overall POSITION (6)  $\times$  DBI (2)  $\times$  DIALECT (2) three-way mixed design ANOVA was performed on syllable duration, with the within-subject factor being POSITION. The main effects of POSITION and DIALECT were significant (POSITION:  $F(4.63, 2043.00) = 54.28, p < .0001, \eta^2 = .11$ ; DIALECT:  $F(1, 441) = 10.47, p < .01, \eta^2 = .02$ ).<sup>12</sup> The two-way interaction regarding POSITION and DBI was also significant,  $F(4.63, 2074.26) = 3.15, p < .01, \eta^2 = .002$ . Other two-way interactions and the three-way interaction were not significant.

Post hoc analyses horizontally across positions regarding the interaction effect showed that syllables at P<sub>-1</sub> were the longest at both DBI<sub>0</sub> and DBI<sub>1</sub>, and those at P<sub>-2</sub> were the second longest,  $p < .05$ .<sup>13</sup> Syllable duration at other positions was not significantly different from one another. Hierarchically, syllables at P<sub>-1</sub> and P<sub>+3</sub> showed significant effects, but in opposite directions. At P<sub>-1</sub>, syllables at DBI<sub>1</sub> were longer than those at DBI<sub>0</sub>,  $p < .05$ , while for P<sub>+3</sub>, syllables at DBI<sub>1</sub> were shorter than those at DBI<sub>0</sub>,  $p < .05$ . No other effects were significant.

**Table 3.** Distribution of boundary syllable types in the two dialects in  $\sigma_{-1}$  situation

Boundary syllable	Taiwan		China		Total
	DBI <sub>0</sub>	DBI <sub>1</sub>	DBI <sub>0</sub>	DBI <sub>1</sub>	
V/CV/VC	87	25	85	22	219
CGV/CVC/CVR	81	12	55	13	161
CVCR/CGVC/ CGVCR	17	0	36	12	65
Total	185	37	176	47	445

Note. C: consonant; G: glide; V: vowel; R: er-ization (i.e., roticization) of the vowel.

With regards to the DIALECT effect, the post hoc test showed that syllable duration was significantly longer in Taiwan Mandarin,  $p < .01$ . Mean difference was about 13 ms.

Since syllable duration is also influenced by segmental duration, and more complex syllable structures tend to be intrinsically longer (Feng, 1985; Shih & Ao, 1997), one would like to make sure that the duration difference between DBI<sub>0</sub> and DBI<sub>1</sub> was not due to these factors. Table 3 shows the distribution of phonotactic structures of the boundary syllables. Chi-square tests showed that the distribution for the Taiwan variety was significant, while that for the Mainland variety was not (Taiwan:  $\chi^2(2) = 6.93$ ,  $p < .05$ ; China:  $\chi^2(2) = .62$ , ns). For the Taiwan dialect, there were fewer V/CV/VC syllables, and more complex ones than expected at DBI<sub>0</sub>, while there were more V/CV/VC syllables, and fewer complex ones than expected at DBI<sub>1</sub>. If differential lengthening at boundary syllables of DBI<sub>0</sub> and DBI<sub>1</sub> were due to an imbalanced distribution of complex syllables, one should have found the opposite trend instead. Therefore, the extra lengthening of boundary syllables at DBI<sub>1</sub> could not have been due to the differential distribution of boundary syllable types.

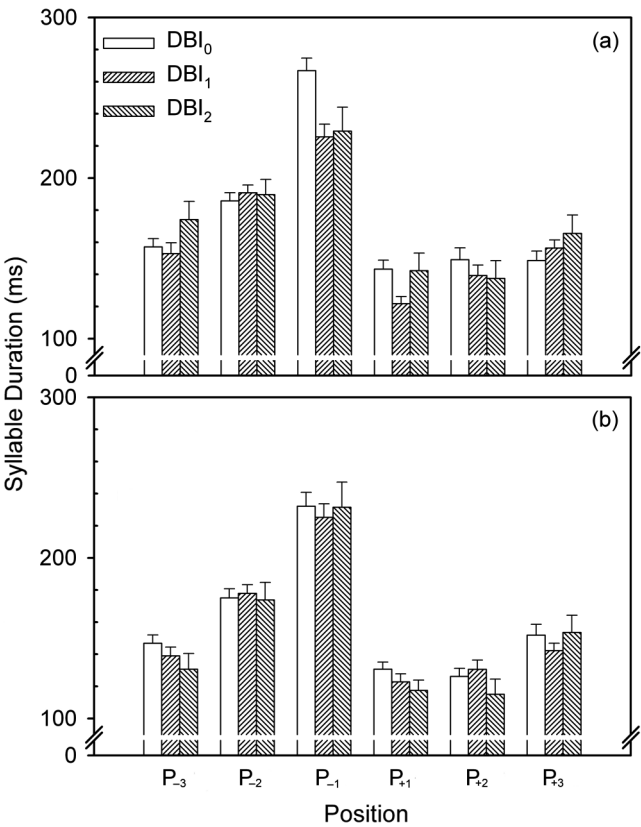
In addition, tone and word class are also likely to influence syllable duration. At utterance-final positions, Tone 2 is the longest, followed by Tone 1, Tone 4, and Tone 3 in Taiwan Mandarin (Fon, 1997), while Tone 3 is the longest, followed by Tone 2, Tone 1, and Tone 4 in Mainland Mandarin (Feng, 1985). Neutral tones are short when they are not sentence particles, but become long when they are (Shih & Ao, 1997). Table 4 shows the distribution of boundary syllables with regards to their tonal categories and word classes. Chi-square tests showed that the distribution was not significant for either dialect (Taiwan:  $\chi^2(5) = 7.21$ , ns; China:  $\chi^2(5) = 5.17$ , ns). Therefore, longer boundary syllables at DBI<sub>1</sub> could not have been due to differential distributions of tonal categories and word classes.

### *Syllable duration at boundary region with $\sigma_{-1}$ +UP*

Figure 5 shows the duration patterning of  $\sigma_{-1}$ +UP at the boundary region for both varieties of Mandarin. Like in the case of  $\sigma_{-1}$ , the effect of lengthening was fairly consistent across all three levels. The scope of syllable final lengthening also extended from P<sub>-1</sub> to P<sub>-2</sub>. However, unlike in syllables of  $\sigma_{-1}$ , the degree of lengthening only seemed to correspond to discourse boundary strength in the Taiwan variety. Syllable lengthening at P<sub>-1</sub> tended to decline as discourse disjuncture became larger, especially between DBI<sub>0</sub> and other higher discourse juncture levels. No such correspondence was found in the Mainland variety.

**Table 4.** Distribution of boundary syllables with regards to their tonal categories and word types in  $\sigma_{-1}$  situation

Boundary syllable	Taiwan		China		Total
	DBI <sub>0</sub>	DBI <sub>1</sub>	DBI <sub>0</sub>	DBI <sub>1</sub>	
Tone 1	30	5	16	1	52
Tone 2	35	6	51	13	105
Tone 3	16	1	9	2	28
Tone 4	52	7	44	15	118
Neutral tone particle	25	10	32	6	73
Neutral tone non-particle	27	8	24	10	69
Total	185	37	176	47	445



**Figure 5.** Syllable duration of (a) Taiwan and (b) Mainland Mandarin at the boundary region of  $\sigma_{-1}$ +UP. The error bars indicate standard error

In order to confirm the above observations, an overall POSITION (6)  $\times$  DBI (3)  $\times$  DIALECT (2) three-way mixed design ANOVA was performed on syllable duration, with the within-subject factor being POSITION. Results showed that all of the main effects were significant (POSITION:  $F(4.36, 2315.70) = 148.56, p < .0001, \eta^2 = .22$ ; DBI:  $F(2, 531) = 3.12, p < .05, \eta^2 = .01$ ; DIALECT:  $F(1, 531) = 17.36, p < .0001, \eta^2 = .03$ ). In addition, the two-way interaction between DBI and POSITION was significant,  $F(8.72, 2315.70) = 1.96, p < .05, \eta^2 = .01$ . Finally, the three-way interaction was also near significant,  $F(8.72, 2315.70) = 1.75, p = .07, \eta^2 = .01$ .

Regarding the three-way interaction, two post-hoc POSITION (6)  $\times$  DBI (3) two-way mixed design ANOVAs were performed separately on the two varieties of Mandarin. For Taiwan Mandarin, results showed that the main effect of POSITION was significant,  $F(4.43, 1214.77) = 75.06, p < .0001, \eta^2 = .22$ , and the main effect of DBI was near significant,  $F(2, 274) = 2.89, p = .06, \eta^2 = .02$ . The interaction effect was also significant,  $F(8.87, 1214.77) = 3.26, p < .001, \eta^2 = .02$ .

Post hoc analyses horizontally across different positions showed that syllable duration was the longest at  $P_{-1}$  regardless of disjuncture levels,  $p < .01$ . For  $DBI_0$  and  $DBI_1$ , syllables at  $P_{-2}$  were also the second longest,  $p < .001$ . No other significance was found. Hierarchically, syllables at  $P_{-1}$  reflected boundary strength. Those at  $DBI_0$  were longer than those at  $DBI_1$  and  $DBI_2, p < .001$  for  $DBI_1$  and  $p = .08$  for  $DBI_2$ . In addition, syllables at  $P_{+1}$  were the shortest at  $DBI_1, p < .01$ . No other significant difference was found.

For Mainland Mandarin, results showed that only the main effect of POSITION was significant,  $F(4.15, 1069.70) = 74.72, p < .0001, \eta^2 = .22$ . No other effects were significant. Post hoc analyses showed that syllables at  $P_{-1}$  were the longest,  $p < .0001$ , and those at  $P_{-2}$  were the second longest, regardless of discourse disjuncture levels,  $p < .001$ .

With regards to the DIALECT effect, the post hoc test showed that syllable duration was significantly longer in Taiwan Mandarin,  $p < .0001$ . Mean difference was about 14 ms.

As with the boundary syllables in the  $\sigma_{-1}$  situation, one would like to make sure that the difference found among  $DBI_0, DBI_1$ , and  $DBI_2$  was not due to other factors such as syllable structure, word type, and tone (Feng, 1985; Fon, 1997; Shih & Ao, 1997). Table 5 shows the distribution of boundary syllable types in the two dialects. Chi-square tests showed that the distributions were not significant in either dialect (Taiwan:  $\chi^2(4) = 7.88, ns$ ; China:  $\chi^2(4) = 9.50, ns$ ), indicating that syllable structure could not have been the main cause for the differential boundary lengths at various DBI levels.

Table 6 shows the distribution of boundary syllables with regards to their tonal categories and word types. One instance of  $DBI_0$  from the Taiwan dialect was excluded because the speaker code-switched to English for the boundary syllable and its tonal category could not be identified. Regardless of dialects, no significance was found using chi-square tests (Taiwan:  $\chi^2(10) = 16.63, ns$ ; China:  $\chi^2(10) = 6.03, ns$ ). In other words, tonal categories and word classes could not have been the main causes for durational differences of the boundary syllables.

### *Syllable duration at $P_{-1}$ : $\sigma_{-1}$ vs. $\sigma_{-1}+UP$*

One would also like to see if compensatory effects exist between boundary syllable and the following pause. If there indeed exists a compensatory relationship, then one would expect boundary



**Table 5.** Distribution of boundary syllable types in the two dialects in  $\sigma_{-1}$ +UP situation

Boundary syllable	Taiwan			China			Total
	DBI <sub>0</sub>	DBI <sub>1</sub>	DBI <sub>2</sub>	DBI <sub>0</sub>	DBI <sub>1</sub>	DBI <sub>2</sub>	
V/CV/VC	56	68	25	58	68	18	293
CGV/CVC/CVR	53	46	8	39	28	10	184
CVCR/CGVC/CGVCR	9	11	1	14	24	1	60
Total	118	125	34	111	120	29	537

Note. C: consonant; G: glide; V: vowel; R: er-ization (i.e., rhoticization) of the vowel.

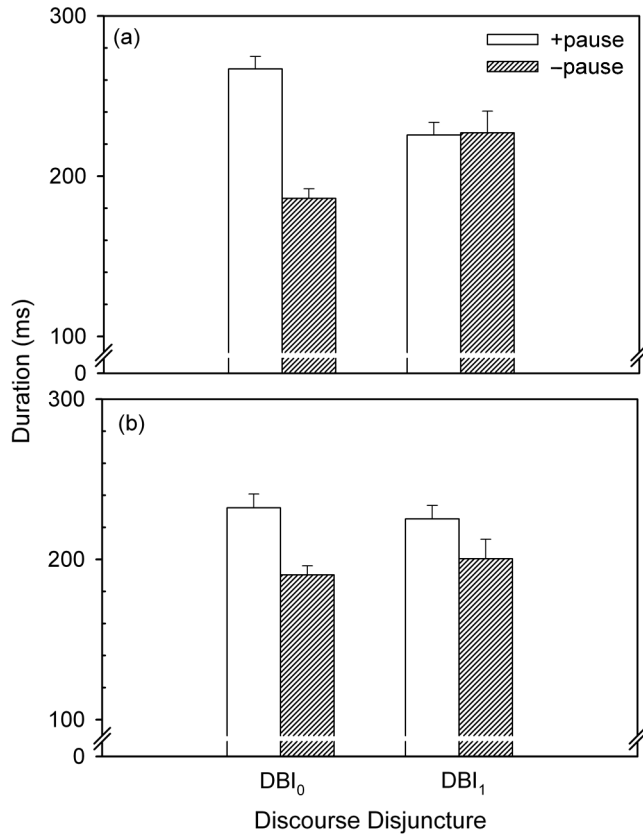
**Table 6.** Distribution of boundary syllables with regards to their tonal categories and word types in  $\sigma_{-1}$ +UP situation

Boundary syllable	Taiwan			China			Total
	DBI <sub>0</sub>	DBI <sub>1</sub>	DBI <sub>2</sub>	DBI <sub>0</sub>	DBI <sub>1</sub>	DBI <sub>2</sub>	
Tone 1	22	15	3	11	17	6	74
Tone 2	19	17	0	27	29	4	96
Tone 3	10	15	4	12	8	2	51
Tone 4	34	44	13	28	27	9	155
Neutral tone particle	13	19	10	18	20	5	85
Neutral tone non-particle	19	15	4	15	19	3	75
Total	117	125	34	111	120	29	536

syllables of  $\sigma_{-1}$ +UP to be shorter than those of  $\sigma_{-1}$ . As shown in Figure 6, there seemed to be an exactly opposite trend. For both Mandarin varieties, boundary syllables of  $\sigma_{-1}$ +UP tended to be longer than those of  $\sigma_{-1}$ .

In order to confirm the above observations, an overall PAUSE (2)  $\times$  DBI (2)  $\times$  DIALECT (2) three-way independent design ANOVA was performed on boundary syllable duration. Only DBI<sub>0</sub> and DBI<sub>1</sub> were included in the analysis since there were not enough cases for DBI<sub>2</sub> of the  $\sigma_{-1}$  condition. Effects involving PAUSE included a main effect,  $F(1, 911) = 32.17, p < .0001, \eta^2 = .03$ , a two-way interaction effect (PAUSE  $\times$  DIALECT:  $F(1, 911) = 14.85, p < .001, \eta^2 = .02$ ), and the three-way interaction effect (PAUSE  $\times$  DBI  $\times$  DIALECT:  $F(1, 911) = 6.38, p < .05, \eta^2 = .01$ ).

Regarding the three-way interaction, two post-hoc PAUSE (2)  $\times$  DBI (2) two-way independent design ANOVAs were performed separately on the two varieties of Mandarin. For Taiwan Mandarin, results showed that both effects involving PAUSE were significant (PAUSE:  $F(1, 461) = 18.24, p < .0001, \eta^2 = .04$ ; PAUSE  $\times$  DBI:  $F(1, 461) = 19.53, p < .0001$ ). Post hoc independent  $t$  tests showed that boundary syllables of  $\sigma_{-1}$ +UP were significantly longer than those of  $\sigma_{-1}$  at DBI<sub>0</sub>,  $t(301) = 8.32, p < .0001$ , but no such difference was found at DBI<sub>1</sub>. For Mainland Mandarin, only the main effect of PAUSE was significant,  $F(1, 450) = 14.01, p < .001, \eta^2 = .03$ . No higher-level

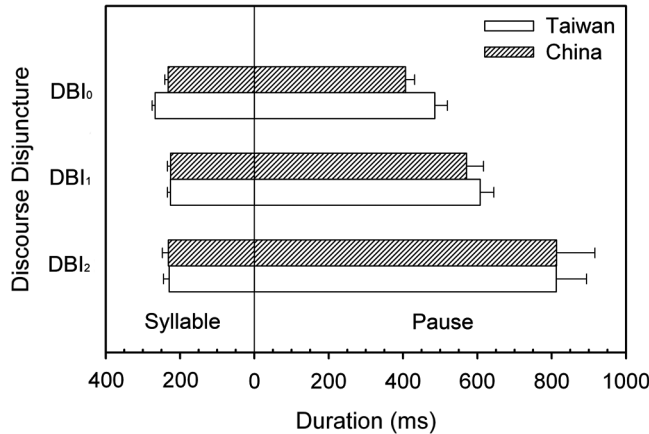


**Figure 6.** Boundary syllable duration of  $\sigma_{-1}$  and  $\sigma_{-1}+UP$  in (a) Taiwan and (b) Mainland Mandarin. The error bars indicate standard error

interaction was found. The post hoc independent *t* test showed that boundary syllables of  $\sigma_{-1}+UP$  were always longer than those of  $\sigma_{-1}$ , regardless of discourse juncture levels.

#### *Pause and SOI duration at $P_{-1}$ for $\sigma_{-1}+UP$*

Figure 7 shows the syllable and pause duration at the boundary location of  $\sigma_{-1}+UP$  for the two dialects of Mandarin. The left side of the graph indicates syllable duration while the right side of the graph indicates pause duration. Therefore, the total length of the bars indicates SOI. Compared to syllable duration, pause and SOI seemed to be much stronger indicators of discourse hierarchy in both dialects. The two measures increased as discourse disjuncture levels became higher. The degree of lengthening was also much stronger than that of syllable.



**Figure 7.** Syllable and pause duration at the boundary location of  $\sigma_{-1}$ +UP with regards to discourse disjuncture levels. The left half of the graph indicates syllable duration while the right half of the graph indicates pause duration. The error bars indicate standard error

Two overall DBI (3)  $\times$  DIALECT (2) two-way ANOVAs were performed to confirm the above observations, one on pause and the other on SOI. With regards to pause, only the main effect of DBI was significant,  $F(2, 531) = 21.27, p < .0001, \eta^2 = .07$ . Post hoc comparisons showed that pauses at DBI<sub>2</sub> were the longest while those at DBI<sub>0</sub> were the shortest,  $p < .001$ . As for SOI, the pattern was fairly similar. Only the main effect of DBI was significant,  $F(2, 531) = 17.62, p < .0001, \eta^2 = .06$ . Post hoc comparisons showed that SOIs were the longest at DBI<sub>2</sub> and shortest at DBI<sub>0</sub>,  $p < .01$ . No dialectal differences were found.

## Discussion

Although both pause and final lengthening are shown to be reliable boundary cues in Mandarin read speech (Shen, 1992), they do not seem to have equal weights in spontaneous narratives. Unlike read speech, in which all boundaries are reported as accompanied by pauses, about 43% of the boundaries were not so in this study (Table 2), indicating that pause is probably not a necessary component at Mandarin syntactic and discourse boundaries, and lengthening alone could potentially be a sufficient cue for boundary signaling, even for higher-level discourse disjunctures. The prevalence of pause in Shen's study might thus be due to the special genre used, obliging the speakers to employ cues that are otherwise optional in regular everyday speech to help listeners disambiguate standing ambiguities in the sentences.

The realization scope of final lengthening is also interesting. Unlike previous studies, in which the scope of final lengthening is fairly localized on the boundary unit at  $P_{-1}$  (Campbell & Isard, 1991; Goldman-Eisler, 1972; Lehiste, 1975; Oller, 1973; Scott, 1982; Shen, 1992), this study found the scope of the effect in spontaneous Mandarin to be extending from  $P_{-1}$  to  $P_{-2}$ . Although the

absolute lengthening duration was not extraordinary, about 67 ms (Taiwan) and 74 ms (Mainland) at  $P_{-1}$ ,  $SE = .005$ , and 31 ms (Taiwan) and 32 ms (Mainland) at  $P_{-2}$ ,  $SE = .004$  (Figure 4 & Figure 5),<sup>14</sup> it was already beyond the just noticeable difference level for speech (ca. 10 to 40 ms, Klatt & Cooper, 1975; Lehiste, 1970). The lengthening ratios were even more phenomenal. There were a 70% (Taiwan) and an 83% (Mainland) increase at  $P_{-1}$  ( $SE = .05$ ), and a 41% (Taiwan) and a 48% (Mainland) increase at  $P_{-2}$  ( $SE = .04$ ),<sup>15</sup> the former of which were significantly larger than Shen's 47% increase of final syllables (Taiwan:  $t(501) = 4.74$ ,  $p < .0001$ ; China:  $t(486) = 7.00$ ,  $p < .0001$ ). Besides genre differences, the much larger lengthening ratios might be one way of compensating for possible processing difficulties due to faster rates and higher unpredictability in spontaneous speech.<sup>16</sup>

The heavy reliance on final lengthening in Mandarin boundary demarcation is somewhat unique, since pause seems to be more indispensable than final lengthening in many Indo-European languages (Goldman-Eisler, 1972; Grosjean, 1980; Grosz & Hirschberg, 1992; Hirschberg & Nakatani, 1996; O'Malley et al., 1973; Scott, 1982; Stenström, 1986; Swerts & Geluykens, 1993, 1994; Swerts & Ostendorf, 1997; Traum & Heeman, 1997). One possibility for this discrepancy might lie in the differences in speech rhythm. Compared with Mandarin, Indo-European languages are more stress-timed (Mok & Dellwo, 2008), and have a richer inventory of phonotactic combinations and vowels that contrast in length. As a consequence, syllable duration becomes more variable (Dauer, 1983), possibly precluding listeners from making reliable boundary judgments based on the cue alone. On the other hand, simple phonotactic structures and lack of vowel length contrasts in Mandarin not only create a more typical syllable-timed rhythm, but might also likely facilitate comparisons among syllables of different lengths without much complex calculation necessary for mental normalization, making final lengthening a potentially dependable cue.

In addition, lack of frequent stress contrasts might further reinforce this reliance on duration.<sup>17</sup> Unlike languages that utilize prosodic prominence in boundary demarcation (e.g., English: Beckman, Hirschberg, & Shattuck-Hufnagel, 2005; Dutch: Gussenhoven, 2005), Mandarin does not have a stress system that could be readily used as a convenient indication for upcoming boundaries (Peng et al., 2005). The prevalence of lexical tones also prevents employment of boundary tones that are common in a number of languages (e.g., Japanese: Venditti, 2005). Therefore, turning to durational cues would be a natural choice for speakers of the language. One suspects that widening the scope of final lengthening might also stem from this reliance tendency. By starting the *ritenuto* process somewhat earlier, listeners might thus receive more facilitative cues for boundary detection, compensating for the restrictive usage of the pitch aspect of the signal. Since Shen (1992) did not include measurements of penultimate syllables in her study, it is unclear whether the scope of final lengthening is genre-dependent. More production studies on Mandarin and other relatively syllable-timed languages, and perceptual studies on differential scopes of final lengthening would be needed in order to better understand the possible underlying mechanism for scope widening.

For boundaries that were accompanied by pauses, various combinations of filled and unfilled pauses were observed. Boundary syllables followed by a silent pause were the most common type in both Mandarin varieties, more so in Taiwan Mandarin (97%) than the Mainland variety (82%) (Table 2). Boundaries with filled pauses were more varied in pattern. In Taiwan Mandarin, filled pauses seemed to only occur after silent pauses, while in Mainland Mandarin, both orders could be found, although unfilled pauses followed by filled ones were still more common. In general, the

usage of filled pauses at discourse boundaries seemed to be more common and thus more acceptable in the Mainland variety, as all except for one speaker showed frequent usages. On the other hand, Taiwan speakers preferred to use silent pauses instead, as boundary filled pauses were mainly contributed by two speakers only.

The distribution of boundary syllable patterns also corresponded to discourse hierarchy (Table 2). For both dialects, boundary syllables became less likely to be without pauses as discourse disjuncture increased (DBI<sub>0</sub>: 60% (Taiwan) and 57% (China); DBI<sub>1</sub>: 22% (Taiwan) and 25% (China); DBI<sub>2</sub>: 8% (Taiwan) and 9% (China)). At higher discourse levels, boundary syllables tended to be followed by silent pauses. The trend was especially prominent in the Taiwan variety (DBI<sub>0</sub>: 38% (Taiwan) and 36% (China); DBI<sub>1</sub>: 76% (Taiwan) and 63% (China); DBI<sub>2</sub>: 89% (Taiwan) and 62% (China)). There was a similar trend for filled pauses, but it was much more obvious in Mainland Mandarin (DBI<sub>0</sub>: 1% (Taiwan) and 7% (China); DBI<sub>1</sub>: 2% (Taiwan) and 12% (China); DBI<sub>2</sub>: 3% (Taiwan) and 30% (China)). In other words, pauses were used profusely in Mandarin to indicate higher-level discourse disjunctures, as are the cases in English and Dutch (Grosz & Hirschberg, 1992; Hirschberg & Nakatani, 1996; Swerts, 1998; Swerts & Geluykens, 1993; Swerts & Ostendorf, 1997). However, there were dialectal differences regarding pause types. Mainland Mandarin employed both filled and unfilled pauses at discourse disjunctures, much like Dutch (Swerts, 1998; Swerts & Geluykens, 1993; Swerts & Ostendorf, 1997), while Taiwan Mandarin preferred unfilled pauses to filled ones, perhaps more like English (Grosz & Hirschberg, 1992; Hirschberg & Nakatani, 1996).

Although a performance limit on the elasticity of the syllable might restrict its degree of lengthening required for higher-level boundary indication, and thus summoning the help of pauses to fulfill the duration requirement is an unsurprising natural solution for language, the differential preference between the two dialects of Mandarin with regards to pause types at higher-level discourse disjunctures is still intriguing. This could not have been due to differences in speech fluency between the two groups of speakers, since the occurrences of filled pauses at non-boundary syllables were about the same, 0.95% (Taiwan) and 0.96% (China).<sup>18</sup> Instead, one suspects that it might be related to dialectal preferences regarding speech rhythm (Scollon, 1985; Tannen, 1981). As Taiwan speakers tend to have slower speech rates than those from China (Figure 3), they might also prefer and tolerate a longer void of speech (Figure 7), and would therefore feel more comfortable with unfilled pauses. On the other hand, Mainland speakers might feel compelled by their faster rhythm to intersperse unfilled pauses with filled ones in order to break the long signal void, even though it is used for boundary indication.

For boundaries without pauses, syllable duration at P<sub>-1</sub> correlated positively with discourse hierarchy in both dialects (Figure 4). Bigger boundaries were accompanied by longer syllable duration, but only for the utterance-final syllable, not the penultimate. The increase from DBI<sub>0</sub> to DBI<sub>1</sub> was longer in Taiwan Mandarin (41 ms) than in the Mainland variety (10 ms), indicating a robust dialect effect. For boundaries with silent pauses, there was a bigger dialectal difference (Figure 5). Taiwan Mandarin still showed a correlation at P<sub>-1</sub>, but in a negative fashion. Syllable duration was longer at DBI<sub>0</sub> than DBI<sub>1</sub> and DBI<sub>2</sub> by around 40 ms. However, no such correlation was found in the Mainland variety. Syllable duration was not a reliable indicator of discourse hierarchy when the boundary was followed by a silent pause in this dialect.

On the other hand, silent pauses at  $P_{-1}$  showed a consistent effect of discourse hierarchy across the two dialects (Figure 7). In both varieties, disjunctures of higher levels were accompanied by longer pauses. The increase between adjacent discourse levels was phenomenal, and the expansion from  $DBI_1$  to  $DBI_2$  was 55% larger than that from  $DBI_0$  to  $DBI_1$  ( $DBI_0 \rightarrow DBI_1$ : 143 ms,  $DBI_1 \rightarrow DBI_2$ : 222 ms).

Judging from the seemingly contradictory patterns between boundary syllable duration and discourse disjunctures of  $\sigma_{-1}$  and  $\sigma_{-1}+UP$ , one suspects what was important in durational patterning of Mandarin was perhaps not the rigid positive or negative correlation between discourse hierarchy and boundary syllable duration. Instead, what speakers of Mandarin were trying to achieve was a positive correlation between discourse hierarchy and boundary SOIs, regardless of boundary types. Therefore, when boundaries were without pauses, positive correlation was found between discourse hierarchy and boundary syllable duration since boundary syllables and SOIs were coterminous in this case. However, when boundaries were followed by pauses, positive correlation was found in pause and SOI duration only. The positive correlation between boundary duration and hierarchical encoding corresponds well with previous studies on Mandarin and English (Fon & Johnson, 2004; Ladd, 1988), implying that it is probably a fairly universal pattern. What would be different among individual languages might be how one defines the domain of boundary duration. For stress-timed languages like English, the domain would be closely tied to stress assignment, while for syllable-timed languages like Mandarin, the domain would be more closely attached to the syllable itself. However, in either case, pause plays a more significant role than syllable, especially for bigger boundaries. This is similar to the trend found in Indo-European languages (Grosz & Hirschberg, 1992; Hirschberg & Nakatani, 1996; Swerts & Geluykens, 1993, 1994; Swerts & Ostendorf, 1997).

The importance of pause in hierarchical encoding might also explain why there existed a curious negative correlation between boundary syllables at  $P_{-1}$  and discourse boundary size in Taiwan Mandarin, contrary to our hypothesis. By reducing the degree of final lengthening at larger boundaries, one is essentially increasing both the relative and absolute duration of pause within a given SOI, thereby creating the percept of a long pause and a large disjuncture. We suspect that the reason why such a pattern was not found in the Mainland variety might also be related to differences in speech rhythm. As mentioned above, Mainland speakers prefer a faster speech rate and thus a lower tolerance level for long silent pauses (Scollon, 1985; Tannen, 1981). This implies that their sensitivity for increments of pause duration might be higher than their Taiwan counterparts', making enlarging relative pause proportions an unnecessary implement. In addition, the smaller degree of boundary lengthening might also imply a more restricted degree of syllable elasticity (Figure 7), rendering differential lengthening across various discourse levels rather difficult. On the other hand, Taiwan speakers' high tolerance for long silent pauses might grant themselves with a higher threshold for detecting pause increments. Greater syllable elasticity allowed in this dialect also facilitates differential lengthening across discourse levels. Therefore, extra mechanisms were possible and perhaps also preferred for clearer boundary hierarchy demarcation. In other words, due to differential preferences for speech rhythm and disparate degrees of syllable elasticity, syllable duration seemed to be indicative of only boundary location, but not boundary size in the Mainland variety, while it seemed to be reflective of both in Taiwan Mandarin.

If what is important for hierarchical encoding is the duration of SOI and perhaps also pause, but not so much the syllable, then one would expect to find some kind of compensatory relationship existing between syllable and pause duration. In other words, boundary syllables of  $\sigma_{-1}$  should be longer than those of  $\sigma_{-1}+UP$ , since the former rely fully on syllable duration to reach the required SOI. However, such intuition was not confirmed (Figure 6). For Mainland Mandarin, boundary syllables of  $\sigma_{-1}+UP$  were always longer instead, regardless of discourse hierarchy. For Taiwan Mandarin, differences were only found for  $DBI_0$ , but not for  $DBI_1$ , again with syllables of  $\sigma_{-1}+UP$  being longer. In other words, both syllable and SOI duration tended to be counter-intuitively longer in  $\sigma_{-1}+UP$ . It is unclear why different amounts of durational cues were used for the same level of discourse disjuncture. One suspects that this might have something to do with perceptual organization of “subjective” rhythm (Bolton, 1894; Woodrow, 1909), by which lengthening of a syllable might create a percept of a following pause that is physically nonexistent (Woodrow, 1951). Therefore, although the two boundary types did not have the same length of SOI, they might have created a subjective illusion of such, and thus were judged as appropriate to encode the same level of discourse disjuncture. More studies will be needed in order to see whether perceptual weighting of syllable and pause would change for the different compositions of final SOIs.

## Conclusion

Although prosodic boundaries do not necessarily coincide with syntactic and discourse boundaries, for those that do, the manner in which prosodic organization is realized is to a large extent reflective of syntactic/discourse hierarchy. This study showed that the hierarchical encoding of durational patterning not only existed in spontaneous speech, but was sensitive enough to signal discourse structures above the clausal level. In both Mandarin varieties, the location of syntactic/discourse boundaries was indicated by a *ritenuto* tempo starting from the penultimate syllable, while the size of the discourse disjuncture was indicated by the duration of boundary SOIs. Pauses were preferred but not required for higher hierarchical levels, with Taiwan Mandarin having a strong preference for unfilled pauses, and the Mainland variety a liking for both filled and unfilled ones. Within a given SOI, pause played a more important role than syllable duration in hierarchy indication, with the latter being significant only when pause was nonexistent. The interaction between syllable and pause duration at the boundary position was dialect-dependent. Taiwan Mandarin maximized the difference among the three discourse levels by decreasing syllable duration and increasing pause duration at the same time at higher levels, while Mainland Mandarin only utilized the latter strategy. Rhythmic difference between the two dialects was suspected to be the cause of such disparity.

This study, though focused on Mandarin, could have certain implications for prosodic universals. The heavy reliance on final lengthening in boundary demarcation as opposed to the more common mechanism of using lengthening plus pause found in the more familiar Indo-European languages seems to indicate that the choice of temporal cues in the boundary region might be to a large extent dependent on preferred linguistic rhythms. Languages that are more syllable-timed tend to rely more heavily on final lengthening, while those that are more stress-timed prefer pause. On the other hand, the role pause plays in boundary hierarchy indication seems to be more universal, as larger boundaries in Mandarin were essentially signaled by longer SOIs, the duration and the lengthening of which were mainly contributed by pause than by syllable duration, similar



to other languages in previous studies. The universal preference for this mechanism is surmised to be related to physiological and rhythmic constraints imposed on syllable elasticity as opposed to pause. However, languages do seem to have differential preferences for one type of pause over the other in temporal patterning, which might be related to rhythmic organization. Those favoring slower tempos might be more inclined to tolerate long silent pauses than those favoring faster ones. Despite various tendencies displayed by languages, one could generally conclude that discourse structure is conveyed in not only the textual content, but also the speech rhythm produced by the speaker. One of the main universal goals of rhythmic organization seems to be orchestrating the relevant temporal cues so that the length of boundary duration, contributed mainly by pause, could match the level of discourse disjuncture, and the structure could thus likely be revealed without the text.

### Acknowledgments

This study was funded by the National Science Council, Taiwan (NSC91-2411-H-003-065-), and a National Taiwan University Commission on Research and Development grant to the first author and a National Institute on Deafness and Other Communication Disorders, USA grant (NIDCD R01DC04421) to the second author. Parts of this article were presented at the 6<sup>th</sup> International Conference on Spoken Language Processing in Beijing, China, the 141<sup>st</sup> Meeting of the Acoustical Society of America in Chicago, USA, the 3<sup>rd</sup> International Symposium on Chinese Spoken Language Processing in Taipei, Taiwan, Linguistics and Phonetics 2002 in Urayasu, Japan, and the 10<sup>th</sup> International Symposium on Spontaneous Speech: Data and Analysis in Tokyo, Japan. Many thanks go to Mary Beckman and Shari Speer for the helpful comments on earlier versions. Thanks to Tsan Huang for the help on recording and transcription, Martin Jansche for structural analyses, Kuenhi Tsai, Pang-Hung Hsu, and Chao-Feng Wu for data compilation and statistical analyses, and Jim Harmon for technical support. Special thanks go to all of the subjects. The authors would like to sincerely thank the Editors, H. Timothy Bunnell and Irene Vogel, and the two reviewers, Chilin Shih and Sasha Calhoun, for the helpful comments. Naturally, all the faults are ours.

### Notes

- 1 Standing syntactic ambiguities are structures that can be disambiguated only by different ways of syntactical bracketing. For example, *the young boy and the girl went together* could be parsed either as [*the young boy*] and [*the girl*], or as *the young* [*boy and the girl*].
- 2 In this study, Mandarin romanization follows the Pinyin convention. Superscript numbers after each syllable indicate tone. Unstressed syllables are indicated by '0'.
- 3 Naturally, stress-timed and syllable-timed are relative rather than absolute terms (Dauer, 1983; Deterding, 2001; Ling, Grabe, & Nolan, 2000; Mok & Dellwo, 2008). Compared with Germanic languages, both varieties of Mandarin would be considered as more syllable-timed than stress-timed. This is because unstressed reduced syllables are not as ubiquitous, even in the Mainland variety.
- 4 About 73–80% of the population in Taiwan is ethnically Min (Chen, 1989; Cheng, 1985).
- 5 Although based on previous studies in Mandarin, initial strengthening was not found to be a significant boundary cue (Fon & Johnson, 2004; Shen, 1992), it is of course possible that the initial positions could also contain some boundary information, and thus could not serve as clean and unambiguous references. However, given the fact that what was dealt with in this study was spontaneous speech containing clauses of various lengths, we felt that the initial positions would be a practical if not the best choice for baseline references.

- 6 PART:RF is a particle that reduces forcefulness (Li & Thompson, 1981).
- 7 ACC: accusative marker.
- 8 PFV: perfective aspect.
- 9 CL: classifier.
- 10 PART:SA is a particle that solicits agreement (Li & Thompson, 1981).
- 11 Only  $\sigma_{-1}$  and  $\sigma_{-1}$ +UP were included in the  $\chi^2$  analyses since most of the less common types did not reach an expected frequency of at least 5.
- 12 Throughout the paper, Huynh-Feldt adjusted df's were used when Mauchly's test showed that the sphericity assumption of ANOVA was violated.
- 13 Bonferroni's adjustments were used throughout the paper for multiple comparisons unless otherwise mentioned.
- 14 The lengthening duration was operationally defined by subtracting syllable duration at P<sub>-3</sub> from that at P<sub>-1</sub> and P<sub>-2</sub>.
- 15 The lengthening ratio was operationally defined by dividing syllable duration at P<sub>-1</sub> and P<sub>-2</sub> by that at P<sub>-3</sub>.
- 16 Based on data reported in Shen (1992), the mean speech rate for her speakers (Mainland) was about 4.35 syll/sec and the articulation rate was about 5.15 syll/sec.
- 17 An estimation using syllables at the boundary region (Figure 2) showed that unstressed syllables only accounted for 14.28% (Taiwan) and 15.35% (Mainland) of all syllables.
- 18 This was estimated from calculation of syllables in the boundary region (Figure 2).

## References

- Beckman, M., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 9–54). New York: Oxford University Press.
- Boersma, P., & Weenink, D. (2007). Praat: Doing phonetics by computer [software]. Available from <http://www.praat.org/>
- Bolton, T. L. (1894). Rhythm. *American Journal of Psychology*, 6, 145–238.
- Butcher, A. (1980). Pause and syntactic structure. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 85–90). Berlin: Mouton.
- Campbell, W. N., & Isard, S. D. (1991). Segment durations in a syllable frame. *Journal of Phonetics*, 19, 37–47.
- Chafe, W. L. (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
- Chen, Y.-D. (1989). *Taiwan de Kejiaren* [Hakka on Taiwan]. Taipei: Taiyuan Press.
- Cheng, R. L. (1985). A comparison of Taiwanese, Taiwan Mandarin, and Peking Mandarin. *Language*, 61, 352–377.
- Clancy, P. M. (1980). Referential choice in English and Japanese narrative discourse. In W. L. Chafe (Ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (pp. 127–202). Norwood, NJ: Ablex.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51–62.
- Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, 29, 217–230.

- Duanmu, S. (2002). *The phonology of standard Chinese*. Oxford & New York: Oxford University Press.
- ERBAUGH, M. (1990). Mandarin oral narratives compared with English: The pear/guava stories. *Journal of Chinese Language Teachers' Association*, 25, 21–42.
- Feng, L. (1985). Beijinghua yuliuzhong sheng yun diao de shichang [Duration of onset, rhyme, and tone in Beijing Mandarin speech]. In L. Tao & L. Wang (Eds.), *Beijing Yuyin Shiyuan Lu* [An Experimental Record of Beijing Mandarin Phonetics] (pp. 131–195). Beijing: Beijing University.
- Fon, J., & Johnson, K. (2004). Syllable onset intervals as an indicator of discourse and syntactic boundaries in Taiwan Mandarin. *Language and Speech*, 47, 57–82.
- Fon, Y.-J. J. (1997). *What are tones really like? An acoustic-based study of Taiwan Mandarin tones*. Unpublished Thesis, National Taiwan University, Taipei.
- Frazier, L., Carlson, K., & Clifton, C., Jr. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, 10, 244–249.
- Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, 15, 103–113.
- Grosjean, F. (1980). Linguistic structures and performance structures: Studies in pause distribution. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 91–106). Berlin: Mouton.
- Grosz, B., & Hirschberg, J. (1992). Some intonational characteristics of discourse structure. *Proceedings of 2<sup>nd</sup> International Conference on Spoken Language Processing*, pp. 429–432. Banff, Canada.
- Grosz, B., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12, 175–204.
- Gussenhoven, C. (2005). Transcription of Dutch intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 118–145). New York: Oxford University Press.
- Hirschberg, J., & Nakatani, C. H. (1996). A prosodic analysis of discourse segments in direction-giving monologues. *Proceedings of 4<sup>th</sup> International Congress of Spoken Language Processing*, pp. 286–293.
- HSU, H.-J. (2006). *Some aspects of phonological leveling*. Unpublished Dissertation, National Taiwan Normal University, Taipei.
- Kärkkäinen, E. (1996). Preferred argument structure and subject role in American English conversational discourse. *Journal of Pragmatics*, 25, 675–701.
- Keller, E., Zellner, B., Werner, S., & Blanchoud, N. (1993). The prediction of prosodic timing: Rules for final syllable lengthening in French. *Proceedings of European Speech Communication Association (ESCA) Workshop on Prosody*, pp. 212–215. Lund, Sweden.
- Klatt, D. H., & Cooper, W. E. (1975). Perception of segment duration in sentence contexts. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 69–89). New York: Springer-Verlag.
- Kubler, C. C. (1985). *The development of Mandarin in Taiwan: A case study of language contact*. Taipei: Student.
- Ladd, D. R. (1988). Declination ‘reset’ and the hierarchical organization of utterances. *Journal of the Acoustical Society of America*, 84, 530–544.
- Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- Ladd, D. R., & Campbell, W. N. (1991). Theories of prosodic structure: Evidence from syllable duration. *Proceedings of 12<sup>th</sup> International Congress of Phonetic Sciences*, pp. 290–293. Aix-en-Provence, France.
- Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.
- Lehiste, I. (1975). The role of temporal factors in the establishment of linguistic units and boundaries. In W. U. Dressler & F. V. Mares (Eds.), *Phonologica 1972* (pp. 115–122). München-Salzburg, Germany: Wilhelm Fink Verlag.

- Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- Lin, T. (1985). Tanta Beijinghua qingyin xingzhi de chubu shiyan [Preliminary experiments in the exploration of the nature of Mandarin neutral tone]. *Working Papers in Experimental Phonetics*, 1–26.
- Lin, Y.-H. (1989). *Autosegmental treatment of segmental process in Chinese phonology*. Unpublished Dissertation, University of Texas at Austin, Austin.
- Ling, L. E., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43, 377–401.
- Mok, P. P. K., & Dellwo, V. (2008). Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. *Proceedings of the 4<sup>th</sup> Conference on Speech Prosody*, pp. 423–426. Campinas, Brazil.
- Oller, D. K. (1973). The duration of speech segments: The effect of position in utterance and word length. *Journal of the Acoustical Society of America*, 54, 1235–1247.
- O'malley, M. H., Kloker, D. R., & Dara-Abrams, B. (1973). Recovering parentheses from spoken algebraic expressions. *IEEE Transactions on Audio and Electroacoustics*, AU-21, 217–220.
- Pakosz, M., & Flaschner, V. (1988). Prosodic features and narrative strategies in Polish discourse. *Papers and Studies in Contrastive Linguistics*, 24, 33–46.
- Peng, S.-H., Chan, M. K. M., Tseng, C.-Y., Huang, T., Lee, O. J., & Beckman, E. M. (2005). Towards a pan-Mandarin system for prosodic transcription. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 230–270). New York: Oxford University Press.
- Scollon, R. (1985). The machine stops: Silence in the metaphor of malfunction. In D. Tannen & M. Saviile-Troike (Eds.), *Perspectives on silence* (pp. 21–30). Norwood, NJ: Ablex.
- Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America*, 71, 996–1007.
- Shen, X. S. (1992). A pilot study on the relation between the temporal and syntactic structures in Mandarin. *Journal of the International Phonetic Association*, 22, 34–43.
- Shih, C., & Ao, B. (1997). Duration study for the Bell Laboratories Mandarin text-to-speech system. In J. van Santen, R. Sproat, J. Olive, & J. Hirschberg (Eds.), *Progress in speech synthesis* (pp. 382–399). New York: Springer-Verlag.
- Shriberg, E., Stolcke, A., Hakkani-Tur, D., & Tur, G. (2000). Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication*, 32, 127–154.
- Stenström, A.-B. (1986). A study of pauses as demarcators in discourse and syntax. In J. Aarts & W. Meijs (Eds.), *Corpus linguistics II: New studies in analysis and exploration of computer corpora* (pp. 203–218). Amsterdam: Rodopi.
- Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, 30, 485–496.
- Swerts, M., & Geluykens, R. (1993). The prosody of information units in spontaneous monologue. *Phonetica: International Journal of Speech Science*, 50, 189–196.
- Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, 37, 21–43.
- Swerts, M., & Ostendorf, M. (1997). Prosodic and lexical indications of discourse structure in human-machine interactions. *Speech Communication*, 22, 25–41.
- Tannen, D. (1981). New York Jewish conversational style. *International Journal of the Sociology of Language*, 30, 133–149.
- Tinsley, H. E. A., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, 22, 358–376.

Traum, D. R., & Heeman, P. A. (1997). Utterance units in spoken dialogue. *Dialogue Processing in Spoken Language Systems*, 1236, 125–140.

Venditti, J. J. (2005). The J-ToBI model of Japanese intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 172–200). New York: Oxford University Press.

Woodrow, H. (1909). A quantitative study of rhythm. *Archives of Psychology*, 14, 1–66.

Woodrow, H. (1951). Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1224–1236). New York: Wiley.

Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (pp. 41–62). Chichester: John Wiley.

# Appendix A

Duration of recording used in this study. Numbers in parentheses indicate number of syllables

Taiwan		China	
ID	Duration (#Syllable)	ID	Duration (#Syllable)
Male speakers			
CHL	2:09.761 (397)	FBS	2:23.906 (621)
HPH	9:26.985 (2290)	LJG	2:23.623 (709)
SCH	2:28.157 (652)	XD	2:39.191 (853)
SCY	2:32.739 (599)	YYL	5:09.366 (1218)
Female speakers			
CYW	5:07.336 (1033)	CY	1:50.872 (508)
HHY	4:21.068 (1041)	LX	4:41.840 (1279)
SMI	2:18.049 (571)	XB	3:13.062 (934)
SYL	2:32.014 (567)	ZLE	1:49.957 (352)
WCF	2:47.893 (614)	ZLI	1:29.299 (364)

**Appendix B**

Number of DBIs each subject contributed

ID	DBI <sub>0</sub>	DBI <sub>1</sub>	DBI <sub>2</sub>
Taiwan speakers			
CHL	12	8	2
CYW	38	19	3
HHY	34	23	5
HPH	75	48	8
SCH	35	8	4
SCY	25	13	5
SMI	33	15	2
SYL	27	16	5
WCF	28	15	4
Mainland speakers			
CY	18	11	3
FBS	30	18	4
LJG	33	25	3
LX	63	37	4
XB	45	30	7
XD	38	26	8
YYL	48	28	10
ZLE	12	8	4
ZLI	20	8	4