BOUNDARY AND LENGTHENING
— ON RELATIVE PHONETIC INFORMATION

TSENG Chiu-yu, SU Zhao-yu

ABSTRACT

The aim of the present study is to better understand the temporal structure of discourse prosody through relative phonetic information, in particular, phrase final lengthening and discourse boundary discrimination. Using data of fluent Mandarin narrative speech, we tested two assumptions to compare their contributions to discourse boundary discrimination, namely, independent/single vs. integrated/paired acoustic cues. Single factors included five acoustic features to see whether the identities of discourse boundaries can be discriminated, namely, (1) boundary pause (BP), (2) pre-boundary duration (PrDu), (3) pre-boundary intensity (PrIn), (4) duration contrast (DuCon) and (5) syllable intensity contrast (InCon). Relative factors were ten paired combinations from the above five features to test their respective contributions to discourse boundary discrimination as well. The results demonstrated that single discrete cues were not as discriminative as paired ones, suggesting that boundary information is related to relative combined cues. Among the paired combinations, the combined cue of pre-boundary syllable duration and the following pause, PreDu+BP, is most discriminative. We further examined pre-boundary lengthening in relation to discourse organization by three discourse prosody units: the syllable, the prosodic word (PW) and the prosodic phrase (PPh), and found pre-boundary global lengthening by the PPh is systematically related to higher-level discourse specifications. The results suggest that discourse constrained tempo modulation across speech flow is default within the same speaking rate. Therefore, we argue that temporal planning is constrained by higher-level discourse planning; higher level planning induces overall lengthening; global lengthening reflects cognitive load. In addition to phone- and syllable-contributed factors, discourse temporal organization is also constrained by discourse unit.

Keywords: discourse boundary, boundary discrimination, final lengthening, relative phonetic information.

1. INTRODUCTION

In previous work on narrative prosody, we have established a hierarchical discourse framework the HPG (Hierarchy of Prosodic Phrase Group) through corpus analysis [8]. The discourse perspective allows examination of fluent speech prosody from top-down; it also makes possible clarifications of terms used for phonetic, phonological and prosody investigations. While the segments are phonetic terms [6][5], terms like syllable, prosodic word, intonation phrase (IP) are phonological terms [4][7] often used interchangeably as phonetic units or prosodic units as well [3][1]. However, in a discourse prosody framework pitch (perceived relative F_0), rhythm and tempo (temporal structure and distribution), loudness (perceived relative strength), boundary lengthening (rather than phrase final lengthening) and boundary pause are all but relative prosodic phenomena in relation to discourse organization. Therefore, by default the syllable (Syl), prosodic word (PW),

prosodic phrase (PPh), breath-group (BG or compulsory change of breath during fluent speech) and prosodic phrase group (multiple-phrase speech paragraph) are all discourse prosodic units that can be analyzed and compared for relative prosodic properties. In particular, an IP is no longer an ultimate intonation unit (IU) but a discourse sub-unit; phrase intonations are no longer examined in isolation; intonations no longer examined for trajectories only but also for overall relative height and tempo. Further, the HPG framework specifies how prosodic units are constrained and governed by prosodic layers, and how theses units and layers contribute systematically and cumulatively to global output prosody, and why fluent speech prosody must include unit-dependent information in relation to discourse organization, rather than simply phone-and-syllable dependent information plus phrase/sentence intonation [8][9][10]. As a result, three major characteristics of the HPG distinguish it from other prosody studies: (1) it emphasizes the relative cross-unit prosodic association contained in fluent speech and specifies how such relative phonetic in the supra-segmental domain can be accounted for. (2) Boundary breaks across fluent speech are treated as discourse units and bear discourse identities. (3) An intonation phrase is a discourse unit subject to HPG specifications.

The HPG prosodic units from the bottom layer upward are the syllable (SYL), the prosodic word (PW), the prosodic phrase (PPh), the breath group (BG) and the multiple phrase group (PG). Two prosodic layers are higher than PPh in the hierarchy. The immediate higher node of the BG is the PG, which is the highest node in the HPG hierarchy and refers to breathing limit which corresponds to a compulsory physiological constraint. The highest node PG refers to a complete multiple-phrase speech paragraph and corresponds to the obligatory and ultimate cognitive constraint of speech. The hierarchical relationships among these nodes are SYL<PW<PPh<BG<PG. In correlation to the HPG units are respective discourse boundaries B1, B2, B3 B4 and B5 which bare the same hierarchical relationships and function as prosodic units. Hence, the relationships among the discourse boundaries are B1<B2<B3<B4<B5. The identities of these discourse boundaries are outcome of perceptual annotation by trained transcribers. The intra- and inter-transcriber consistency was over 93% [19]. Furthermore, by specifications of the HPG an intonation phrase is a discourse subunit PPh, and by default not an ultimate prosody unit. The discourse identity of a PPh is subject to three PG specifications the PG-initial, -medial or -final. As a result, output global discourse prosody must contain higher level PG information accordingly. In short, our previous work has shown that in fluent continuous speech, additional prosodic information exists in addition to tones, stress and phrase intonation in the supra-segments, and no prosodic information should be studied as discrete units; and relative associations must be accounted for.

In the following sections, we will present a study on the timing structure of Mandarin discourse prosody through boundaries and lengthening to show how higher level temporal allocation is organized by discourse units and represents discourse-relative phonetic information, and how the phrase also functions as a temporal unit in fluent speech.

2. Experiments

2.1 Speech material

Two types of Mandarin speech corpus in different speaking rates were used. Read speech of (1) plain text of 26 discourse pieces (CNA, approximately 6700 syllables) by one male M051 and one female F051, and (2) three rhyme formats of Chinese Classics (CL approximately 1600 syllables) by one male M056 and one female F054. The speech data were recorded in sound proof chambers. Pre-analysis annotation included automatically labeled segmental identities in the SAMPA-T notation using the HTK toolkit, and subsequent manual tagging by trained transcribers of perceived boundary breaks using the Sinica COSPRO Toolkit [11]. Annotated segments were spot-checked

by professional transcribers for identities and alignments. Table1 summarizes the speech material by corpus type, speaker, and the number of the HPG prosodic units and boundaries.

Table 1: Summary of speech data by corpus type, speaker, and the HPG prosodic units and boundaries. The HPG prosodic units are the syllable (SYL), prosodic word (PW), prosodic phrase (PPh), breath group (BG) and phrase group (PG).Corresponding HPG boundaries following each of the prosodic units are B1, B2, B3 B4 and B5 respectively.

corpus	speaker	SYL/B1	PW/B2	PPh/B3	BG/B4	PG/B5
CNA	F051	6583	3468	1092	297	151
	M051	6661	3332	1207	270	129
CL	F054	1444	599	290	135	58
	M056	1551	619	318	142	47

The mean syllable durations for speakers F051 and M052 are 199ms and 189 ms; the mean syllable durations for speakers F054 and M056 are 265ms and 202ms. Taken as a reference to speaking rate, we found a positive correlation by speech material than by speaker. The above materials were used for all three experiments in the present study.

2.2 Experiment 1

We have stated in Section 1 that discourse prosody is mainly about semantic cohesion through relative associative information manifested in the supra-segmental domain. The rationale implies that using relative acoustic information would result in better generalized pattern and discrimination of discourse information than discrete acoustic information. Given that boundaries are important discourse information, three discourse boundaries, PPh boundary B3, BG boundary B4 and PG boundary B5, were selected as the categories of generalization and discrimination. Three discrete acoustic variables were chosen to test the generalization and discrimination. They are (1) boundary pause (BP), (2) pre-boundary syllable duration (PrDu) and (3) pre-boundary syllable intensity (PrIn). The following two steps were employed to examine patterns of generalization and boundary discrimination.

Procedure 1. Whether a single acoustic factor is sufficient to generalize and discriminate discourse boundary identities

The procedure involved testing whether generalization and discrimination could be achieved by any single acoustic factor. The average values of specified acoustic feature for B3, B4 and B5 were derived from the speech materials by speaker and by speech type. These derived mean values across B3, B4 and B5 were plotted as reference that denotes the tendency among boundaries by speech data type and speaker. We then compared the trajectories among different speech data to look for whether the best single acoustic factor with most generalized pattern could be identified. We also tested whether discrimination of discourse boundary identities could be attributed to any one of these single discrete factors.

Procedure 2. Whether a relative acoustic factor is sufficient to generalize and discriminate discourse boundary identities

The same rationale from Procedure 1 was utilized to test boundary generalization and discrimination, but using one relative acoustic factor at a time. Between-boundary duration contrast (BwDuCon) and between-boundary intensity contrasts (BwInCon) were calculated and used as the contributing factors. Between-boundary duration contrasts were defined by subtracted outcome of cross-boundary syllables. The same subtraction was applied to derive the between-boundary intensity contrasts as well. Both duration and intensity contrasts specify cross-unit as well as cross-boundary relative acoustic information. The same averaging and comparison methods used in

Procedure 1 were employed to see if any single relative factor is sufficient to discriminate the identities of discourse boundaries.

2.3 Experiment 2

We further hypothesized that pairing of single factors would result in better generalization and discrimination than results from Experiment 1, and the discrimination varies by pair, and thus specified pairing would result in better discrimination than single factors to discriminate the three discourse boundaries B3, B4 and B4.

The five acoustic features generated from Experiment 1, namely, (1) boundary pause (BP), (2) pre-boundary duration (PrDu), (3) pre-boundary intensity (PrIn), (4) duration contrast (DuCon) and (5) syllable intensity contrast (InCon), were used as feature candidates to generate paired-combinations and as variables for ANOVA. These five features were first normalized then paired. A total of ten paired combinations were selected. These 10 paired variables were calculated by ANOVA for discriminating categories B3, B4 and B5 from each other.

2.4 Experiment 3

We have previously established that temporal templates for each prosodic unit can be derived by the HPG framework [8,9], suggesting that default temporal patterns exists in each prosodic layer in addition to phone and syllable durations. This rationale made possible a hypothesis that final lengthening is discourse unit/boundary dependent and must be addressed with boundary pause information. In other words, boundary discrimination must include both boundary pause information as well as pre-boundary duration patterns to signal discourse information. Since in fluent continuous speech, an intonation phrase is a sub-unit of discourse unit, pre-boundary lengthening is not simply constrained by phrase information alone. To test the hypothesis, we calculated pre-boundary duration patterns by the HPG prosodic units, namely, the syllable, the PW and the PPh, and compared their respective patterns to the speech data.

3. Results

3.1 Experiment 1

Results from Procedure1 reveal that none of the three single factors (1) pause duration, (2) pre-boundary syllable duration and (3) pre-boundary syllable intensity is discriminative of discourse boundaries except the factor of pause duration. We noted that the factor of pause duration discriminated boundaries B3 from B4 and B5; yet at the same time it was not discriminative of boundaries B4 and B5. In other words, discrimination by pause duration is limited. Moreover, no identities of discourse boundary can be discriminated by either the pre-boundary syllable duration or the pre-boundary intensity as shown in Figure 1. The results suggest that boundary discrimination cannot be attributed to any single factor. Furthermore, pre-boundary lengthening is not a boundary feature by itself. The results thus motivated further examination of the role of final lengthening in relation to discourse information in subsequent investigations.

Results from Procedure 2 reveal that between paired single-factor contrasts, namely, the contrasts between pre- and post- PPh boundary duration and intensity by one syllable, no significant discrimination of discourse boundaries could be achieved either, as shown in Figure 2. However, we note that paired duration factor, the between-PPh-duration contrast, is discriminative of text type CAN and CL, as shown in the upper panel of Figure 2; while single duration factor, the pre-boundary duration is not, as shown in the middle panel of Figure 1. In addition, cross-speaker consistency is found by text type, thus indicating text type may be inherent to rate of reading whereas speaker difference is not, as reported in Sec. 2.1. Regarding intensity patterns, the between-boundary intensity contrasts by only one syllable also provide better discrimination than

pre-boundary intensity alone. In other words, although using paired single factors to bring out minimum relative information does not produce better discrimination across the board; paired factors still perform better than single features.

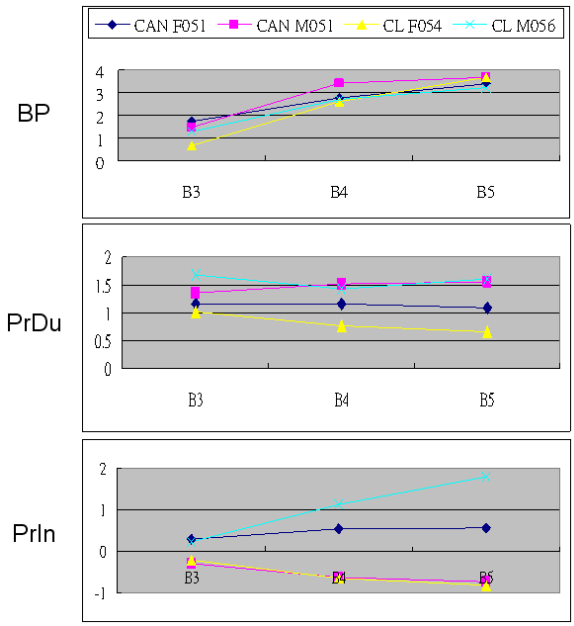

Figure 1: Cross boundary discrimination by single acoustic features. Each panel denotes one specific acoustic feature. The horizontal axis represents the prosodic boundary indexes B3, B4 and B5. The vertical axis represents the coefficient of normalized values of boundary pause (BP), pre-boundary duration (PrDu) and pre-boundary intensity (PrIn), respectively. Zero at the vertical axis is defined as the mean of syllable duration.

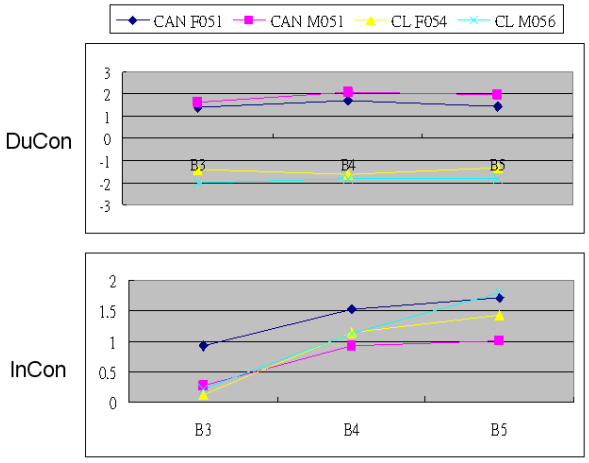

Figure 2: Cross boundary discrimination by single contrastive factors. Each panel denotes one specific contrastive feature. The horizontal axis represents the prosodic boundary indexes B3 to B5. The vertical axis represents the coefficient of normalized values between boundary duration contrasts (DuCon) and between boundary intensity contrasts (InCon). Zero at the vertical axis is defined as the mean of syllable duration and intensity.

3.2 Experiment 2

The results of Experiment 2 are summarized in Table 2. The terms Within and Between are two evaluation indicators for discourse boundary discrimination. We defined Within by the population variance of distribution of sample means; and Between the distance between the sample means. The F-ratio (Between/Within) indicates distinctions among B3, B4 and B5.

The obtained results indicated that among the ten paired combinations, significance of boundary distinction was only found among two pairs, namely, PrDu+BP and PrIn+BP, where $F(2, 40) = 0.28387$, $P<0.05$. That is, the PrDu+BP pair contributes most to discourse boundary discrimination, followed by the PrIn+BP pair. It is therefore obvious that when boundary pause is combined with either pre-boundary duration or intensity, identities of discourse boundaries can be discriminated. In addition, the Within is minimal for pairs PrIn+BP and PrDu+BP, indicating that when pre-boundary duration is combined with between-boundary intensity contrast, boundary discrimination is best, followed by the combination of boundary pause and pre-boundary duration. The above results further suggest that the PrDu+BP combination, namely, the PPh-final syllable duration plus the following pause, is the most discriminative relative cue of discourse boundary identities.

Table 2: Summary of speech data by corpus type, speaker, and the HPG prosodic units and boundaries. The HPG prosodic units are the syllable (SYL), prosodic word (PW), prosodic phrase (PPh), breath group (BG) and phrase group (PG).Corresponding HPG boundaries following each of the prosodic units are B1, B2, B3 B4 and B5 respectively.

Pairs of Acoustic features	**PrIn+BP**	**PrDu+BP**	**BP+InCon**
Between	2.394360811	2.117735421	0.930326811
Within	0.714117065	0.479096215	1.116294559
F-ratio	3.352896784	4.420271653	0.833406204

Pairs of Acoustic features	**BP+DuCon**	**PrIn+InCon**	**PrDu+PrIn**
Between	0.070391796	1.297194809	0.120075103
Within	1.810655131	0.912193354	0.875052214
F-ratio	0.038876424	1.422061237	0.137220501

Pairs of Acoustic features	**PrIn+DuCon**	**PrDu+InCon**	**PrDu+DuCon**
Between	0.353532872	1.020569418	0.076907482
Within	1.652913517	0.374550542	1.763574503
F-ratio	0.213884676	2.72478425	0.043608865

Pairs of Acoustic features	**DuCon+InCon**
Between	1.254027187
Within	1.736954223
F-ratio	0.721969048

Table 3 summarizes the averaged sum of PPh-final syllable duration and boundary pause duration in seconds, where constant pattern across boundaries can be observed.

Table 3: A list of average sum of final syllable duration and pause (sec by speech data type and speaker)

corpus	speaker	B3	B4	B5
CNA	F051	0.499738	0.607713	0.684998
	M051	0.519527	0.800465	0.880004
CL	F054	0.52102	0.833563	1.007355
	M056	0.456447	0.679508	0.774484

3.3 Experiment 3

Figure 3 shows the phrase final duration patterns by HPG prosodic units--the syllable, the PW and the PPh across speech data and speaker. We noted that by analyzing the pre-boundary duration pattern of the final syllable alone, it revealed that consistent and patterned lengthening occurs before the B3 boundary, but not before higher boundaries B4 and B5. This result does not explain why discourse boundary identities could be consistently perceived across listeners. However, if the same inconsistency was found across all boundaries, which happened to patterns found for the PW, then lengthening may not be a reliable boundary cue and is related to the lower level phrase boundary only. However, when pre-boundary duration patterns were examined by PPh, similar patterns are found across speaker and data-type, as shown in the lower panel of Figure 3. Furthermore, the lengthening patterns of pre-boundary PPh are also consistent with discourse boundary type. Such consistency can be seen as evidence of higher level overall slowing down or lengthening by phrase, thus suggesting calculation of tempo, rhythm and speaking rate merits more sophisticated considerations.

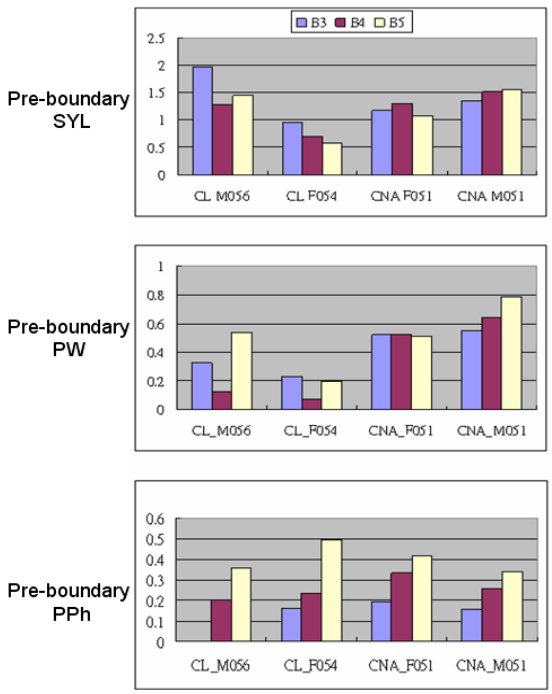

Figure 3: Cross boundary comparison of duration patterns by prosodic units--the syllable (SYL), the PW and the PPh. The horizontal axis represents indices of the speech data and speaker. The vertical axis denotes normalized average duration of prosodic units.

Figure 4 shows the results of average duration patterns by discourse boundary identities B3, B4 and B5. We found that discourse boundaries can be discriminated by pre-boundary duration patterns of higher level discourse unit (the PPh) across speaker and speech data type, as shown in the lower panel of Figure 4; but not by pre-boundary duration patterns of lower level discourse units (the SYL and PW), as shown in the top and middle panels in the same figure. In other words, the identities of discourse boundaries B3, B4 and B5 are only consistent with the pre-boundary duration patterns of the PPh.

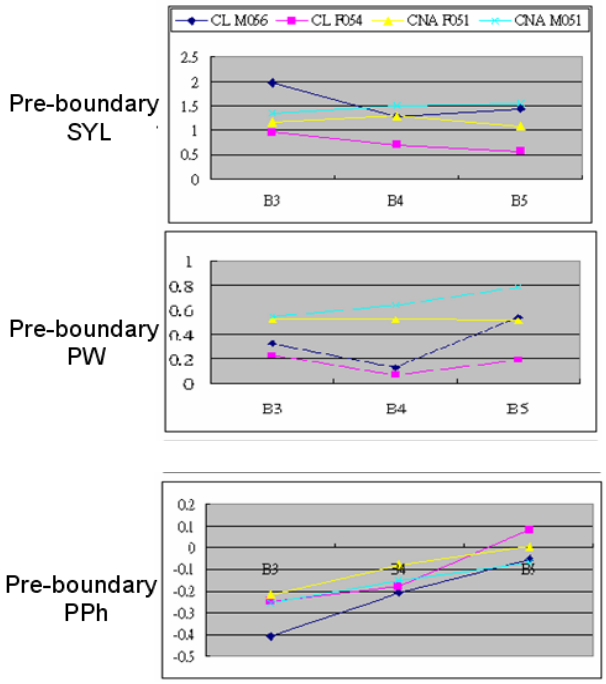

Figure 4: Cross-boundary duration patterns by boundary breaks. The panel denotes result of specific prosodic unit. Each curve denotes one of speech data. The horizontal axis represents prosodic boundary index. The vertical axis denotes the normalized average duration for specific prosodic unit.

4. Discussion

One of the most difficult tasks of phonetic analysis is how to qualitatively and quantitatively account for relative phonetic information. Both the pitch and temporal features in phonological structures are presented in relative abstract organization such as high vs. low or fast vs. slow, while realization in speech signals is often presented in absolute units such as Hz and msec. Although timing structure of the Mandarin segments has been studied as in phonetic terms to provide phone-dependent information, the Mandarin syllable has been investigated more as a phonological unit instead of a rhythmic unit, and Mandarin speech rhythm hardly investigated above the syllable. In addition, little reference has been made of speaking rate and speech rhythm with reference of global duration patterns by prosodic units.

For example, one of the most well known previous acoustic studies on Mandarin duration patterns is how stress is related to temporal modulation instead of F_0 contours [16], referring to segmental and syllable-duration modifications at the lexical level only. However, much less attention has been paid to relative phonetic information at the discourse level and prosody units

above segments and the syllable. Studies on boundary and lengthening were no exception. For example, a comprehensive investigation on Mandarin segment lengthening made important observations of how prosodic boundaries and pause occur between prosodic units instead of within units, and manifestation of pre-boundary lengthening bear prosodic functions to the phrase [14]. More recent studies reported on the role of lengthening with reference to prosodic boundary and its perceptual significance in continuous speech using the pre-boundary syllable [20][13], thus inadvertently suggested the syllable as the default unit of lengthening. Another study reported that lengthening is complemented by pause duration at prosodic boundaries, but the units did not go beyond the intonation phrase [17]. Perceptual studies reported that although the pause duration at sentence-final positions is significantly longer than that of phrase-final ones, the syllable duration at sentence- and prosodic-phrase-final positions was not significantly different [18]. Furthermore, although more recent studies reported how the degree of final lengthening is modulated by boundary types [2] and how segmental strengthening is relative to prosodic functions [15], little discussion with reference to discourse units and structure is discussed. In short, almost all of the previous studies have focused on modulation of segmental duration at the syllabic level. We noted also that even when the discourse factor was considered, there has been less reported account in relation to discourse organization. In particular, the relative aspect of timing structure with respect to boundary features and boundary identities in discourse prosody has been overlooked. We think that one reason of the oversight could be due to taking the phonetic or phonological unit IP (intonation phrase) as an independent prosodic unit, whereas by HPG account, an IP corresponds to a PPh which is a discourse and speech-paragraph sub-unit and therefore requires further discourse specifications by default.

Interestingly, in one of the recent studies on the PPh boundary B3, we studied the much varied B3 pause duration not by the duration of pauses, but with respect to cross-unit contrastive patterns in the acoustic signals, and discovered that within PG phrase boundaries can be accounted for by boundary immediate contrastive patterns of duration and intensity without any pause information [12]. The findings thus explained why the within-PG phrase boundary B3 was consistently perceived across listeners irrespective of pause duration, even when there was no pause at the boundary. We discussed when processing fluent speech on-line, the listener makes use of crucial relative phonetic information related to prosody organization. In short, both neighboring and cross-over prosodic references provide cues to global prosody; higher-level discourse organization is reflected in global units and boundaries in relation to each other. Therefore, discourse prosodic units should not be taken as discrete ones, nor should they be investigated when discourse context is removed and discourse organization/information absent.

The results from Experiments 1 show that single factors are not discriminative of discourse boundaries. The results of Experiment 2 show that the identities of discourse boundaries can be discriminated when pre-boundary syllabic duration or intensity is combined with the following boundary pause. In other words, pre-boundary syllabic information by itself is not sufficient for the discrimination of boundary identities, but when coupled with the following pause, the combined feature proved to be adequate. The results suggest that when prosodic context is limited, a little extra relative information goes a long way. We believe that more high-level relative information is utilized by the listener to facilitate faster and easier on-line top-down processing.

The results from Experiment 3 are most interesting because it provided evidence of how global lengthening could be represented and what its discourse function is. The results thus make direct reference to how overall timing modulation can be represented quantitatively and how lengthening occurs to the entire pre-discourse-boundary phrase. The evidences also show why in fluent speech lengthening is applied by prosody unit instead of by the syllable, and how global temporal planning is manifested. Consistent perceptual identification of discourse boundary identities echoes the finding, because listeners must make use of global relative information to facilitate on-line processing. Alternatively, the same results also imply that overall modulation of temporal

allocation is regulated by discourse prosodic organization, and interact with fixed or changed speaking rate. We believe that the implications of global lengthening have shed new lights to how speakers plan and process the temporal features across fluent speech, and default discourse temporal templates could be derived and modeled.

5. Conclusion

We have shown that (1) overall temporal modulation within a fixed speaking rate involves the timing structure and temporal arrangement at the discourse level and result in overall lengthening of the pre-boundary phrase, (2) how lengthening is in fact an integral part of boundary information by discourse units, and when coupled with boundary pause facilitates boundary identities to emerge, (3) lengthening is relative and should be addressed as relative phonetic information, and (4) global lengthening related to overall modulation of speaking rate shows how the timing structure of discourse prosody is subject to discourse organization and discourse association. In summary, we hope to show that relative phonetic information that exists in the speech events, though usually not accounted for in phonological investigations, contributes significantly to the production and processing of fluent continuous speech, the most natural and used form of speech communication. Such relative information would not emerge, unless we adopt a discourse perspective of investigation and make use of methodological innovations.

6. REFERENCES

[1] Crystal, D. 1969. *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

[2] Fon, J., Johnson, K. 2004. Syllable onset intervals as an indicator of discourse and syntactic boundaries in Taiwan Mandarin. *Language and Speech* 47(1), 57-82.

[3] Halliday, M. A. K. 1967. *Intonation and Grammar in British English.* The Hague: Mouton.

[4] Pierrehumbert, J. B. 1980. *The phonology and phonetics of English intonation*. PhD dissertation, MIT, Cambridge

[5] Pike, K. L. 1943. *Phonetics: a Critical Analysis of Phonetic Theory and a Technic for the Practical Description of Sounds*. Ann Arbor: University of Michigan Press.

[6] Scott, N. C. 1941. Broad Transcriptions. *Le Maitre Phonetique* 76: 48-51.

[7] Selkirk, E. O. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge: MIT Press.

[8] Tseng, C. Pin, S. Lee, Y. 2004. Speech prosody: Issues, approaches and implications. *From Traditional Phonology to Modern Speech Processing* (语音学与言语处理前沿), In: Fant, G., Fujisaki, H., Cao, J. and Xu, Y.(eds.), Foreign Language Teaching and Research Press (外语教学与研究出版社), 417-437, Beijing, China.

[9] Tseng, C. Pin, S. Lee, Y. Wang, H., Chen, Y. 2005. Fluent Speech Prosody: Framework and Modeling. *Speech Communication (Special Issue on Quantitative Prosody Modeling for Natural Speech Description and Generation)*, Vol. 46:3-4, 284-309.

[10] Tseng, C. 2006. "Prosody Analysis" in *Advances in Chinese Spoken Language Processing*. In: Chin-Hui Lee, Haizhou Li, Lin-shan Lee, Ren-Hua Wang, Qiang Huo (eds.), *World Scientific Publishing*, 57-76, Singapore.

[11] Tseng, C, Cheng, Y, Chang, C. 2005. Sinica COSPRO and Toolkit--Corpora and Platform of Mandarin Chinese Fluent Speech. *Oriental COCOSDA 2005*, (Dec. 6-8, 2005), Jakarata, Indonesia.

[12] Tseng, C., Chang, C. 2008. "Pause or No Pause? –Prosodic Phrase Boundaries Revisited." *Tsinghua Science and Technology*, 13.4: 500-509.

[13] Zu, Y., Chen, X., 1999. Segmental duration and lengthened syllables." *ICPh99*`, San Francisco, 277-280.

[14] 曹剑芬 （1998） 汉语普通话语音节奏的初步研究。中国社会科学院语言研究所语音研究报告1998。北京：中国社会科学院语言研究所。

[15] 曹剑芬 （2005） 音段延长的不同类型及其韵律价值。中国社会科学院语言研究所语音研究报告2005。北京：中国社会科学院语言研究所，5-12。

[16] 林焘（1983）探讨北京话轻音性质的初步实验。语言学论丛， 第 10 辑。北京：商务印书馆。
[17] 钱瑶、初敏、潘悟云 （2001）普通话韵律单元边界的声学分析。第五届全国现代语音学会议。中国：北京，70-74。
[18] 王蓓、杨玉芳、吕士楠 （2001）汉语韵律层级边界结构的声学相关物。第五届全国现代语音学会议。中国：北京，161-165。
[19] 郑秋豫 （2001） 语流中韵律结构的主要征信。 第 6 届全国语音通讯学术会议 (NCMMSC-6)，(Nov. 19-24, 2001)。中国：深圳，169-172。
[20] 祖漪清、陈肖霞 （1999） 连续语流中的音节延长及其作用。第四届全国现代语音学学术会议。中国：北京，58-63。