Durational Cues at Discourse Boundaries in Taiwan Southern Min

Sheng-Fu Wang, Janice Fon

Graduate Institute of Linguistics, National Taiwan University, Taiwan
sftwang0416@gmail.com, jfon@ntu.edu.tw

Abstract

This study examined durational cues at discourse boundaries in spontaneous speech in Taiwan Southern Min (TSM) using a two-hour monologue corpus contributed by four elderly female speakers from the same dialectal region. Discourse labeling was done following the method of Fon et al. (2011). Two durational cues were examined: syllable duration and post-syllable pause duration. Results showed that the lengthening of final and penultimate syllable was a very robust cue for discourse boundaries. It was also found that the post-boundary syllables underwent shortening. As for boundary strength, the duration of pauses at discourse boundaries was found to be a strong indicator for boundary hierarchy.

Index Terms: boundary cues, hierarchical structure, spontaneous speech, Taiwan Southern Min

1. Introduction

Studies have shown that sentential and discourse boundaries are signaled by acoustic cues such as pitch reset, final lengthening, and increased and lengthened pauses as part of the syntactic and discourse organization in both read [1, 2, 3] and spontaneous speech [3, 4, 5, 6]. In addition to indicating the existence of a boundary, these cues have also been found to reflect the strength of the syntactic or discourse boundaries [2, 3], which has been called the "hierarchy effect" [5, 6]. Among these, durational cues such as final lengthening have been found to be fairly universal [5].

Yet, different languages or even different dialects of a language may utilize such durational cues to a different extent [5]. For example, in English, final lengthening at the boundaries of discourse units only happens at the final syllable, while penultimate lengthening is found in Mandarin Chinese. As for the hierarchy effect of discourse boundaries, it has been found that final lengthening can reflect boundary strength in Mandarin spoken in Taiwan, but the effect is not found in English, nor in the Mainland dialect of Mandarin, suggesting the difference is not only cross-linguistic but also cross-dialectal.

The present study aimed to investigate durational cues at discourse boundaries in spontaneous speech in Taiwan Southern Min (TSM), which is a tone language rooted in Taiwan hundreds of years before Mandarin immigrants settled in in the 1950s. It is the mother tongue of the Holo people, the largest ethnic group that accounts for approximately 70% of the population in Taiwan [7]. One of the most striking features of TSM is its rich tonal inventory, which consists of seven lexical tones, including two checked tones[1]. Also, TSM features an abundance of syllable-neutralization phenomena constantly described as obligatory rules [8], which suggests that such neutralization should better be considered stress patterning at the lexical level. The possible existence of lexical stress, which is absent in Taiwan Mandarin, may in turn suggest that TSM is a more-stress timed language similar to the Mainland variety of Mandarin, which also has lexical stress in its prosodic system.

As mentioned, previous studies on two dialects of Mandarin showed that the Taiwan variety differs from the Mainland dialect in terms of how discourse boundary strength is coded in final lengthening. Such a difference has been attributed to the possible influence of the prosodic, especially the rhythmic, structure of these two dialects [6]. Such a speculation implicitly pointed to the fact that Taiwan Mandarin is more syllable-timed, whereas Mainland Mandarin is more stress-timed. Analogous investigations on TSM, a language featuring more neutralized syllables and hence possibly more stress-timed, makes it possible to answer whether the difference found between Taiwan and Mainland Mandarin can be explained in terms of their prosodic structures per se, or the difference may also be attributed to the influence of geographical separation and influence of areal languages. If TSM patterns similarly with Mainland Mandarin, the similarity between these two and the difference between Mainland and Taiwan Mandarin may be explained in terms of the prosodic systems of these languages: perhaps languages that utilize stress either because of neutral tones at the lexical level, as in Mainland Mandarin, or neutralization that occurs frequently, as in TSM, would favor a certain kind of durational patterning. On the other hand, if the utilization of durational cues in TSM patterns similarly with what has been found in Taiwan Mandarin, then we have to seek other explanations concerning the difference between Taiwan and Mainland Mandarin such as geographical separation of these two varieties and Taiwan Mandarin's close contact with TSM. Of course, it is also possible that TSM employs some cues not present in both dialects of Mandarin, which will show that cross-linguistic similarities prosodic structures cannot account for all the phenomena in a given language. In other words, the result of this study may provide an important step towards a construction of typology on acoustic boundary cues, as well as probing the possibilities of how languages may vary in their utilization of prosodic cues even within a language family and with a similar prosodic structure.

2. Methods

2.1. Corpus

The corpus contained two hours of spontaneous monologues contributed by four female speakers in their sixties. All of the speakers were from the Taichung area so that dialectal difference should be at its minimum. The speech was elicited in the form of an interview in which the interviewer asked the interviewee to talk about personal experiences. The recordings were transcribed, and necessary annotations and syllable segmentation were done with the Praat [9] software.

[1] Checked tones are lexical tones on syllables with a plosive coda. For example, the tone in /liok8/ ('land') is a checked tone.

2.2. Discourse Labeling

Discourse segmentation was done with the transcribed texts of the recordings. The texts were first segmented into basic discourse units, which are clauses, defined as units that contain a main verb according to the classic proposal of Li and Thompson [10]. Next, the relationship between clauses was judged according to the level of discourse disjuncture. Adapting Fon et al.'s [6] method, four different levels of "Discourse Boundary Indices (DBI)" were distinguished in this study. The first level was DBI-0, which means that the two adjacent clauses describe the same thing or event, thus the boundary between these two is merely a clausal boundary. DBI-1 refers to the disjuncture around which the clauses talk about different but related topic. DBI-2 referred to situations where the boundary clearly differentiates two topics. DBI-3 was an additional label for handling radical shifts of themes. This kind of labeling is believed to be able to reflect the hierarchical organization of discourse units. The resulting labeling on discourse structure was subsequently annotated and aligned with recordings using Praat [9]. The distribution of discourse boundaries is shown in Table 1.

Table 1. *By-subject and Overall distribution of discourse boundaries.*

	DBI-0	DBI-1	DBI-2	DBI-3	Total
Subject1	527	70	37	14	648
Subject2	354	30	21	8	413
Subject3	512	48	28	15	603
Subject4	711	84	34	3	832
Total	2104	232	120	42	2496

2.3. Durational Cues

Two durational cues were investigated in this study, syllable duration and post-syllable pause duration. For the measurement of syllable duration, the targets were syllables before and after a discourse boundary: the syllable before a boundary was referred to as p-1, the penultimate syllable was referred to as p-2, and the antepenultimate syllable was referred to as p-3. In addition, the syllables following a discourse boundary were labeled as p+1, p+2, etc (Figure 1). The inclusion of these two categories was motivated by the possibility that post-boundary syllables may also exhibit a "post boundary effect", as is the case of English [4].

Post-syllable pauses referred to the silent intervals as perceived by the labeler. Their positions were coded in accordance with the syllables they followed.

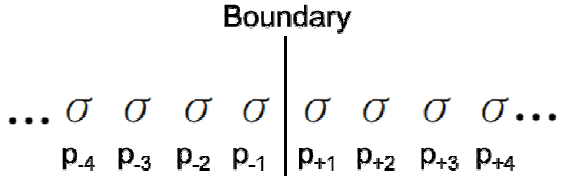

Figure 1: *The schematized representation of labeling on syllable position.*

3. Results

3.1. Syllable duration

Previous studies [6] have shown that the presence of pauses at discourse boundaries might have influence on durational cues. Thus, two groups of data were extracted from the overall dataset. Durational data of the syllables around a discourse boundary with a silent pause were categorized into the "boundary pause" subset, and the data of syllables around a boundary without a silent pause were categorized into the "no boundary pause" subset.

Two-way mixed ANOVA tests, with POSITION as a within-subject factor and HIERARCHY as a between-subject factor, were run for both sets of data. The POSITION factor, with eight levels ("p-4", "p-3", "p-2", "p-1", "p+1, "p+2", "p+3", "p+4"), referred to the positions that the syllables are in. The HIERARCHY factor referred to the level of discourse boundaries. Since the frequency of DBI-3 was too low, this category was excluded from further analyses, leaving the HIERARCHY factor containing only three levels. When the violation of sphericity happened, the degrees of freedom were adjusted using the Huynh-Feldt method unless otherwise mentioned.

The results of the "no boundary pause" subset showed an effect of POSITION, $F(5.71, 6446) = 2.68$, $p < .05$. The result is plotted in Figure 2. The tests of within-subject contrasts revealed the penultimate syllables were significantly longer than the antepenultimate syllables ($p < .001$), while the final syllables were near-significantly longer than the penultimate syllables ($p = .08$), and significantly longer than syllables on the post-boundary "p+1" position ($p < .05$)

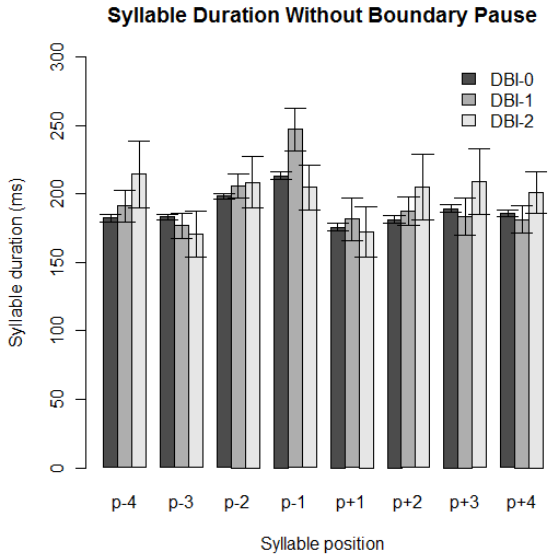

Figure 2: *The pattern of syllable duration at discourse boundaries without pause. The x-axis shows the positions relative to the boundary, and the y-axis shows syllable duration.*

The results for the "boundary pause group", presented in Figure 3, showed a main effect of POSITION [$F(2.45, 2799) = 13.56$, $p < .001$]. The tests of within-subject contrasts revealed that syllables in the penultimate position were

significantly longer than those in the antepenultimate position ($p < .05$), and the final syllables were longer than the penultimate syllables ($p < .001$). Also, syllables in the p+2 and p+4 positions were significantly longer than syllables in the preceding position ($p < .01$ for both comparisons).

There was also a POSITION × HIERARCHY interaction [$F(4.9, 2799) = 7.47$, $p < .001$]. Since homogeneity of variance was not assumed, a Games-Howell post hoc test was employed. The results showed that in the penultimate position, a syllable was longer when the following discourse boundary was indexed as the lowest level, namely DBI-0, than when it was a DBI-1 boundary ($p < .01$). Also in the final position, a syllable was longer when the following discourse boundary was labeled DBI-0, than when it was a DBI-2 boundary ($p < .05$). Both results showed that discourse boundary strength was coded by syllable duration in a reversal manner.

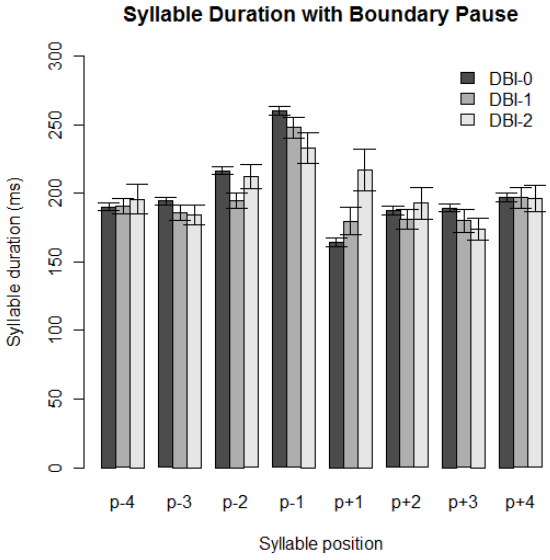

Figure 3: *The pattern of syllable duration at discourse boundaries with pause. The x-axis shows the positions relative to the boundary, and the y-axis shows syllable duration*

3.2. Post-syllable pause duration

We ran a one-way ANOVA test, with HIERARCHY as the between-subject factor, for pauses following syllables at each of the eight positions. In other words, a total of eight tests were run. Three of these tests showed a significant effect of HIERARCHY.

The test on pauses at the p-1 position revealed a significant result [$F(2,1348) = 17.63$, $p < .001$], and the LSD post hoc test revealed that pauses at the boundary were significantly longer when they were of DBI-2 and of DBI-1 than of DBI-0 ($p < .001$ in both comparisons), showing a local effect of discourse hierarchy on pauses at the boundary position.

For the p+1 position [$F(2,245) = 10.85$, $p < .001$], the LSD post hoc test showed that pauses following the post-boundary syllable were significantly longer when they were of DBI-2 and of DBI-1 than of DBI-0 ($p < .001$ for the former and $p < .05$ for the latter comparisons).

As for the p+2 position [$F(2,256) = 8.6$, $p < .001$], the LSD post hoc test revealed that pauses following the second syllable after a discourse boundary were significantly longer when they were of DBI-2 and of DBI-1 than of DBI-0 ($p < .05$ for the former and $p < .001$ for the latter comparisons).

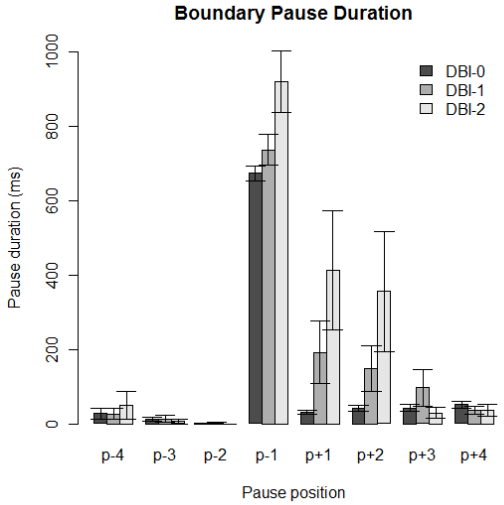

Figure 4: *The patterns of pause duration at different levels of discourse boundaries. The x-axis shows discourse boundaries of different strength, and the y-axis shows pause duration.*

4. Discussion

In the pre-boundary context, final and penultimate lengthening was shown. Final lengthening was suggested to be one of the most universal cues [5] and the discovery of it in TSM adds one more language to strengthen the case for the claim on universality. On the other hand, the discovery of penultimate lengthening showed that the range of pre-boundary lengthening in TSM is the same as in Taiwan and Mainland Mandarin [6]. Neither English, a stress language, nor Japanese, a pitch-accent language, exhibits such a two-syllable range of lengthening before discourse and syntactic boundaries [5]. Such a difference has been explained by the fact that Mandarin has a more typical syllable-timed rhythm that lacks vowel length contrasts so that durational cues are more relied upon at the pre-boundary context [6]. Since TSM is also a more syllable-timed language when compared with typical stress-timed languages such as English, its resemblance to Mandarin in term of the range of pre-boundary lengthening adds one of piece of evidence suggesting that a strong pre-boundary lengthening effect is a characteristics shared by syllable timed languages. Of course, studies on more languages with different rhythmic structure will have to be done to strengthen this typological claim.

Another boundary effect on syllable duration was found in the post-boundary context: the first syllable after a discourse boundary seemed to be shortened. This effect is not found in Mandarin, nor is it present in English, in which the first syllable, if accented, after a syntactic boundary actually exhibits a lengthening effect [5]. A possible explanation is that such a "shortening" of post-boundary syllables served to highlight the presence of pause, which was not only a cue for

discourse and syntactic boundary, but also a very robust and important cue for the hierarchical organization of discourse found in previous studies [5, 6] and in the present study.

As for the hierarchy effect, as briefly mentioned above, boundary pause duration was found to be robust, although in the present study, multiple comparisons only showed a difference between the lowest level of discourse boundary and other higher levels. These findings are in line with previous studies that found durational cues involving pauses to be better cues for reflecting the strength of discourse boundaries in production [5, 6], as well as a very robust cue for detecting boundary strength in perception [11].

A more interesting finding on pause duration may be that, in addition to pauses right at a discourse boundary, the pauses following the first and the second syllable after a discourse boundary also exhibited a hierarchical lengthening effect. It suggests that the hierarchical organization of discourse is not just manifested through local durational patterns at the boundary. However, a possible confounding factor should be noted: a substantial amount of pauses at the p+1 and p+2 positions were preceded, and likely triggered, by discourse markers such as *henn7* 'right' and *an2-ne* 'this', which initiated discourse units after boundaries. It is possible that the presence of such markers enhanced the hierarchical contrast in durational patterns at discourse boundaries. In future research, the effect of such discourse markers will have to be teased apart.

As for syllable duration and boundary strength, previous studies also showed that in Taiwan Mandarin, as well as in Japanese, the duration of syllables before a boundary may also reflect boundary, but in a reversed manner. That is, the pre-boundary syllable is longer when the boundary is of lower levels. This phenomenon was explained as a strategy to highlight the difference in boundary pause duration, which is found to be a better cue for reflecting the hierarchical organization in discourse [5, 6]. The same result was found in the final and penultimate syllables before a pause in the present study, and it should be noted that when there was not a pause at the boundary, one of the most obvious trend of the effect of boundary strength on syllable duration was the longer duration of syllables in the final position when the boundary was of DBI-1 than it was of DBI-0. In other words, when there was no pause, pre-boundary syllable duration seemed to able to encode boundary hierarchy in a positive manner, at least in terms of the contrast between DBI-0 and DBI-1. It may suggest that syllable duration may still be used as a cue for signaling boundary strength positively, although it was secondary to and less consistent than pause duration.

Overall, the results of TSM showed similar patterns with what has been found in Taiwan Mandarin. The implication could be that similar durational cue in Taiwan Mandarin and Taiwan Southern Min can be considered areal features resulting from the intensive contact of these two languages in Taiwan, and the influence of language contact is stronger than the difference of rhythmic systems, if we assume that TSM is less syllable-timed Taiwan Mandarin because of abundance of neutral and neutralized tones. Yet, the present results could also imply that languages with neutral and neutralized tones, like TSM and Mainland Mandarin, may still apply durational cues in a different manner. Further studies on the rhythmic systems of these three languages, especially on spontaneous speech, have to be done in order to see whether such a difference is mediated through the rhythmic systems, hence explaining that whether the rhythmic structure of languages may affect the utilization of durational cues in signaling discourse structure.

5. Conclusion

Investigation on durational cues in TSM spontaneous speech revealed a variety of cues that may be used to reflect discourse organization. Final lengthening at discourse boundaries was very robust, complying with the pattern found in many languages. In addition, two more universal cues found in this study are boundary pause, both of which reflect the hierarchical organization of discourse. Also, TSM discourse boundaries were found to be coded with penultimate syllable lengthening, as found in two dialects of Mandarin [5, 6], which may suggest that syllable duration is indeed a very important cue for discourse boundaries so that the domain of lengthening goes beyond the syllable at the boundary. A negative correlation between boundary hierarchy and duration of syllables before a boundary pause was also found, showing a similarity with Taiwan Mandarin. A possible shortening effect for post-boundary syllables, which is not found in Mandarin, was also found in TSM, which suggested that a language may still use some cues not present in languages that share similar prosodic structures.

6. References

[1] Grosjean, F., Grosjean, L., Lane, H. "The Patterns of Silence: Performance Structures in Sentence Production", Cognitive Psychology 11, 58-81, 1979.

[2] Smith, C. L., "Topic transitions and durational prosody in reading aloud: production and modeling", Speech Communication., 42: 247-270, 2004.

[3] Zellner, B. "Pauses and the Temporal Structure of Speech" in : Keller (E.), (ed.). Fundamentals of Speech Synthesis and Speech Recognition, Basic Concepts, State-of-the Art, and Future Challenges, 41-62, Chichester: John Wiley, 1994.

[4] Swerts, M., "Filled Pauses as markers of discourse structure", Journal of Pragmatics., 30: 485-496, 1998.

[5] Fon, Y.-J. J., A cross-linguistic study on syntactic and discourse boundary cues in spontaneous speech. Dissertation., The Ohio State University. 2002.

[6] Fon. J., Johnson, K., Chen, S. "Durational Patterning at Syntactic and Discourse Boundaries in Mandarin Spontaneous Speech", Language and Speech, 54(1): 5-32, 2011.

[7] Government Information Office, The Republic of China Yearbook 2010, Kwang Hwa Pub. Co., 2010.

[8] Ang, U.-J. Taiwan Helaoyu shengdiao yanjiu (Research on Taiwanese Tones). Taipei: Zili wanbao. 1985.

[9] Boersma, P., Weenink, D. Praat: doing phonetics by computer (Version 5.1.43) [Computer program]. Retrieved April 10, 2010, from http://www.praat.org/, 2009.

[10] Li. C. and Thompson, S., Mandarin Chinese: A functional reference grammar., University of California Press, 1981.

[11] Lin, H.-Y. & Fon, J., Perception of temporal cues at discourse boundaries. Proceedings of Interspeech 2009, 808–811, 2009.

7. Acknowledgements

This research was supported by National Science Concil, Taiwan, project number NSC99-2410-H-002-222-MY3 and by the Erasmus Mundus Action 2 program, MULTI, of the European Union, grant agreement number 2009-5259-5.