Duration and Pauses as Boundary-markers In Speech: A Cross-linguistic Study

Li-chiung Yang

Spoken Language Research, Department of Foreign Languages and Literature
Tunghai University, Taiwan

lyang@sprynet.com or yang_lc@thu.edu.tw

Abstract

Duration is an important feature both to achieve high-quality synthesis and as a marker of phrasal organization in speech. In this study, we investigate pauses and durational patterns in English and Mandarin spontaneous conversation, as well as how reliably such elements can serve as boundary-marking predictors cross-linguistically. Results show that pause duration is significantly correlated with specific boundary status and that syllable duration is inversely correlated with distance to phrase end in both English and Chinese. Our findings support the view that duration features are highly consistent and that it is useful to integrate such knowledge to enhance performance in interactive systems.

Index Terms: duration, pauses, Mandarin, English, final-lengtheing

1. Introduction

Recent research has focused on how pauses and syllable duration serve as markers of discourse organization, and to what degree they are reliable indicators of phrase boundaries in speech [1] [2] [3] and [4]. As a primary phonological variant, duration has been approached as both a local segmental and syllable level phenomenon and as a feature exhibiting more global influences. Previous work on duration has found that duration relationships in natural speech are complex and are affected by a number of contextual factors such as phoneme identity, stress, number of syllables, position in utterance, and focus. Moreover, attention has also focused on the influence of the speech environment itself as conditioning the incidence and functional effect of duration within a specific speech venue. Conversational speech occurs on a spectrum of spontaneity, including narration, interactive give-and-take, and in more formal or structured settings, and each of these exhibits a range of discourse functions of duration. In addition, duration also plays a role in mirroring individual speaker characteristics and underlying cognitive processes, adding additional layers of complexity. Each of the elements contributing to this complexity may have a greater or lesser degree of language-specific incidence and form. For this reason, an increased knowledge of how duration functions in different languages is a critical element for understanding the universality of duration phenomena and achieving a fuller characterization of how structure and coherence are communicated in language through duration.

2. Duration and Pauses: Data

2.1. Goals

In this study, we present the results of our comparative investigation into duration structure in English and Mandarin.

We address the following questions: How do pause usage and pause duration in English and Mandarin function as indicators of phrasal organization and to what extent can pauses serve as boundary marking predictors? How do pause duration and syllable duration compare in English and Chinese within similar discourse functions, and how does pause distribution differ across different speech types and in specific conversational venues? What are the specific functions and forms of pause and syllable duration in narrative speech? How universal are durational phenomena across languages, and how do these phenomena bring structure and coherence to narrative and conversational speech?

2.2. Database

For this study, data were collected from both English and Mandarin conversational speech, and a wide variety of speech settings were chosen to shed light on different speech styles and conversational modes and their effects, especially with regard to narrative speech. English language data for this research consist of broadcast speech from a variety of settings, including 2 short TV interviews, each about 4-5 minutes, 1 longer interview from a TV news magazine of about 15 minutes in length, and 1 single speaker radio story of about 17 minutes. Mandarin data consist of subsections taken from 6 hours of spontaneous Mandarin dialogues in informal conversational settings, as well as 4 children's stories read by a native adult speaker of Mandarin. For both English and Chinese, data were segmented to the syllable level, and durational features, including syllable, word, phrase, pause durations, and distance measures were extracted automatically. For phrase boundary marking, a 2-level categorization scheme differentiating major and minor phrases was adopted, resulting in 3 types of labels to account for these boundary pauses as well as internal non-boundary pauses. Phrases were segmented as minor or major corresponding to whether the phrase is a subsidiary or tangential part of a larger idea unit. Major phrases correspond roughly to sentences, while minor phrases are clauses and phrases like PP, NP, VP, and fragments.

3. Distribution and Frequency of Pauses

Pauses have been associated with syntactic phrasing as well as with such cognitively associated activities as hesitation, confusion, non-understanding, and reluctance. Syntactically, pauses have been considered as boundary markers of phrases, with the specific duration associated with different hierarchical levels of phrase structure and with finality of topic or paragraph, as well as with delineation of semantic/syntactic phrase structure segmented into idea units [5]. Pauses at phrase junctures can play a communicative role in emphasizing finality of idea or topic and pause duration

may also directly the differential difficulties in activating successive idea units into expressible speech.

To find out to what extent are pauses used in different types of speech, for both English and Chinese we calculated the total number of pauses and phrases, and their total time in absolute terms and as percentages. Table 1 shows summary statistics for the 4 sets of English language data we analyzed, with the total number of phrases (Ph), number of pauses, number of boundary pauses (BP) and non-boundary pauses (NBP), and Table 2 shows similar statistics for the 3 sets of Chinese data.

Table 1. *Distribution of Pauses by Type, English data.*

File	Ph	Pauses	BP	NBP	BP/Ph	BP/Pauses
DS1	177	70	62	8	35.0%	88.6%
DS2	187	110	67	43	35.8%	60.9%
DS3	435	326	288	38	66.2%	88.3%
DS4	337	385	254	131	75.3%	66.0%

Table 2. *Distribution of Pauses by Type, Chinese data.*

File	Ph	Pauses	BP	NBP	BP/Ph	BP/Pauses
MD1	666	686	428	258	64.3%	62.0%
MD2	695	547	384	163	55.3%	70.0%
MD3	162	213	159	54	98.1%	74.6%

Tables 1 and 2 show, for both languages, that pauses correlate relatively well with phrase boundaries and that this is consistent across all of the speech venues represented, whether TV interview, extended narrative, or read speech. Over both languages, the percent of pauses that are boundary pauses (BP/Pauses) is consistently high, varying from 60% to 88.6%, and both languages are comfortably within that range, leaving roughly one-third of pauses as internal, non-boundary pauses. Thus, over 60% of pauses indicate a boundary status, and depending on the type of speech data, this can approach 90%.

How well pauses can serve as boundary markers also depends on how consistently phrases are marked by a pause, i.e., the ratio of boundary pauses to total number of phrases (BP/Ph). From the data we can see that there is a very large variation in how consistently phrases are marked by pauses; DS1 and DS2 of the English data had the smallest percentages of pause-marked phrases, at about 35%. By contrast, the Chinese read speech data of MD3 had a remarkable 98.1% of its phrases ending in a pause. For both languages, there was a wide range of approximately 40% in the percentages of phrases marked by pauses, although the percentages are higher in Chinese than in English for these corpora.

Our analysis of the data shows that the narrative structure and associated differing cognitive nature of the conversations contribute to the differences in phrase-marking by pauses. In particular, the very high percentage of phases that end in a pause is very characteristic of read speech, and MD3 consists of read speech of a children's story; the reading is entertaining and considerate of his listener, but also very structured, with complete ideas encapsulated in phrases and sentences. Cognitively, read speech involves little mental strain and few, if any, unexpected situations that call for memory search or mental hesitation. Planning for topic development is built into the written record rather than having to be done by the speaker. In normal speech, we would expect that this would result less time spent on pauses [6]. In the special case of children's reading, however, clarity of story line and focus on providing the child-hearers with easily grasped organizational structure may be especially emphasized; in this case pauses are seen as conducive to the hearer's abilities to grasp information. On the other hand, the read story needed little *cognitive planning* on the speaker's side, resulting in fewer phrase-internal pauses, consisting of about 25% of pauses.

The long English narrative of DS4 also matched this pattern: it has the highest proportion of pause-marked phrases of the English data, at 75.3%, and although it is not read speech, it is similar to the children's story in that it is an extended story with a structured plot line and many distinct subtopics. The speaker's narration has a relatively greater number of non-boundary pauses, about 25%, than the children's story, as the narrative is longer, spontaneously delivered, and consequently required greater cognitive search and organization in its presentation. There are many internal pauses used for emphasis, rhythmic effects, and hesitation, with phrase-to-phrase development also systematically marked by boundary pauses.

Spontaneous and interactive speech are likely to exhibit phase-marking by pauses to a lesser degree, as in the low 35% marking seen in DS1 and DS2. DS1 and DS2 are short interviews involving two speakers, males in DS1 and females in DS2. The style in each interview is informal, with considerable freedom for interaction and topic development. Thus, coherence and looser narrative structure and development may rely more on interactive cues and interruptions than on phrasal marking by pauses.

This influence of spontaneity on pause distribution is also seen when we compare MD1 and MD2 of the Chinese data. MD1 consisted of a discussion about activities of intense interest to both speakers, with frequent points of potential miscommunication. By contrast, MD2 was a more descriptive linear narrative with little cognitive difficulty in topic development. This is reflected in the greater use of internal pauses for MD1 (about 38%) than for MD2 (30%), and less pause-marked phrases for MD2 (55%) than for MD1.

4. Pause Duration and Boundary Structure

Narrative structure is distinct in that it consists of hierarchies of ideas, topics, and sub-topics, and a natural segmentation into minor and major phrases occurs in speech that together move a communicative exchange forward. What role does pause duration play in marking phrase boundaries, and can pause duration help to signal this hierarchy of narrative flow? We categorized the pauses in both our English and Chinese corpora as either phrase-internal pauses, minor phrase-final pauses, or major phrase-final pauses, the differences in duration by speaker for each type of pause can be seen in Figures 1 and 2.

Our data indicate that across languages, pause duration functions very similarly to delineate minor from major phrases, and pause duration also acted in a similar manner to distinguish non-phrase boundary pauses from boundary pauses. In both the English and the Chinese data, average pause durations for major phrase endings are from about .3 seconds to .5 seconds; for minor phrase endings somewhat lower, and for non-boundary pauses, average pause length was between .2 and .3 seconds for all speakers in both languages. These patterns indicate clear consistency in distinguishing phrase internal pauses, and minor and major phrase boundary pauses.
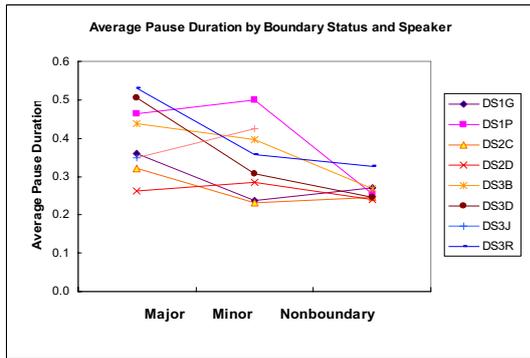
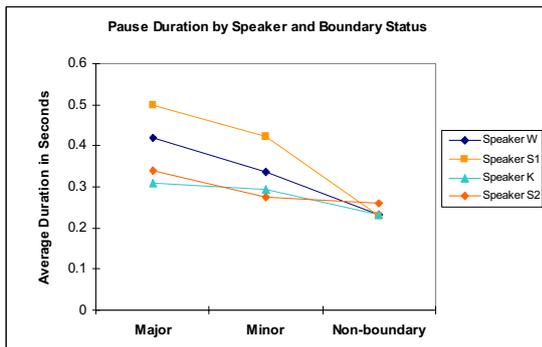Figure 1: *Average pause duration by boundary status and by speaker: English Data.*

Figure 2: *Average pause duration by boundary status and by speaker: Chinese Data.*

Figure 3: *Histogram of pause duration for MD1, speaker S.*

Figure 4: *Histogram of pause duration for MD3, read children's stories.*

To see how well we can rely on pause duration to give unambiguous differentiations between these different narrative events, we need to look beyond the averages and consider the overlapping in their respective distributions. Histograms of pause duration shown in Figures 3-4 for different speakers illustrate that the nature of the conversation and of the speakers themselves is a critical element in how well pause duration distinguishes between boundary and non-boundary, as well as between minor and major phrases.

It is clear from the histograms that differences in speaking style have a great difference in the scale of the pause durations, as well as in the predictive capability to distinguish boundary and non-boundary pauses. In particular, the slower and more structured story presentation has longer pauses, and more differentiation between the different types of pauses, while the more interactive talks have shorter pauses, and relatively more overlap. However, pause duration by itself is not enough to unambiguously distinguish boundary and non-boundary pauses. (| =minor boundary, || = major boundary)

Example 1: Read children's story data, from MD3

1 p<1.053> It's snowing |
2 p<0.368> Susu and Tongtong were holding their little hands together |
3 p<0.522> jumping up and down into the compound ||
4 p<0.748> they saw the old man Grandpa Zhang p<0.319> had bought a full carload of cabbages |
5 p<0.340> piled up high next to the staircase ||
6 p<0.977> Grandpa Zhang picks up a big cabbage |
7 p<0.504> hunching his back p <0.080> and walking upstairs slowly and unsteadily |
8 p<0.550> stopping every few steps |
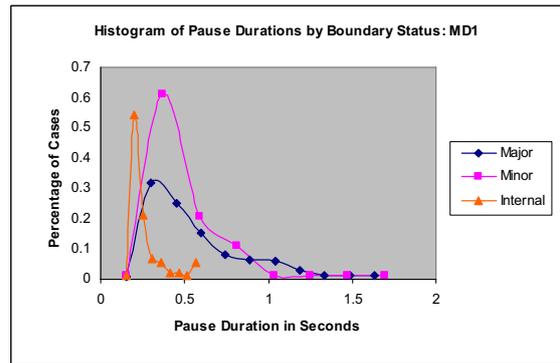9 p<0.071> taking a few deep breaths ||

10 p<1.029> come on |
11 p<0.203> let's go help Grandpa Zhang carrying the cabbages upstairs ||
12 p<0.730> Tongtong is a little boy |
13 p<0.392> he runs over there immediately ||
14 p<0.862> Susu follows him with her pigtails swinging in the wind |
15 p<0.449> wait for me ||
16 p<1.006> the two kids put down their book bags |
17 p<0.491> each of them picks up a big head of cabbage |
18 p<0.637> walking up the stairs hastily ||

Example 2: Spontaneous Mandarin conversation from MD1

253 W: p<0.383> you yige jiao shenme Julia Tao - |
254   p<0.383> wo buxiaode ta zhongwen jiao shenme |
255 S: zhong (mingzi) zhongguo ren? ||
   W: p<0.095> zhongguo ren ||
256 S: p<0.064> m umhum ||
257 W: <1.000> ta de nege machine translation shi nege |
258   p<0.473> eh |
259   p<0.501> Chinese character p<0.069> input |
260 S: umhum ||
261 W: p<0.689> ranhou shi p<0.149> English output ||
263 S: p<0.194> umhum ||
264 W: p<0.036> shi paraphrase |
265 S: p<0.326> umhum ||
   W: p<0.100> shi zheyangzi de ||
266   p<0.912> wo shi ting tamen shuo |
267   shuo haoxiang |
268   p<0.180> input de shi yingwen de |
269   ranhou output de shi p<0.334> zhongwen de nege Chinese character output |
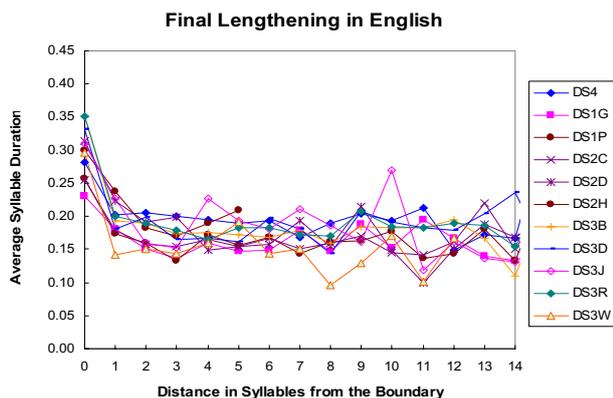270   zhege shi bijiao rongyi de ||

Figure 5: *Average syllable duration by speaker as a function of distance to phrase end in syllables for English data.*

Figure 6: *Average syllable duration by speaker as a function of distance to phrase end in syllables for Chinese data.*

5. Final Lengthening and Distance to Phrase End

Final lengthening of the syllable at or near the phrase boundary has been one of the key research findings linking duration and phrase boundaries, and there has been some debate on whether final lengthening is confined only to the last syllable [3] [4] [7] and [8]. For both our English and Chinese corpora, we calculated average syllable durations by speaker and by closeness to phrase end, as shown in Figures 5-6. It is clear from these figures that lengthening in syllable duration is not evident when the distance to phrase end is greater than 4 or 5 syllables, but with syllables closer to phrase end, there is a *progressive* lengthening, with the final syllable before the boundary, at distance 0, having the longest duration. It is interesting that the upper limits for English in our data are slightly higher than the Chinese at about .25 seconds. Moreover, both syllable duration patterns flatten out at about .15 seconds.

Our previous research has suggested that Chinese tones have an effect on final lengthening, but it is not yet clear whether there is a more language-specific structural reason for the difference here. The results seen here are consistent across all of the speakers in our data and provide convincing evidence for final lengthening in spontaneous discourse, and it further shows that this effect is not confined *solely* to the final syllable but spread over several preceding syllables.

6. Conclusion

In this paper we have compared English and Mandarin pause structure and syllable duration under a variety of speech venues and conversational contexts. We have demonstrated that these contextual and individual speaker factors have systematic effects on how pauses and duration are used to structure narrative and conversational discourse, We have shown that pauses correlated fairly well for phrase and boundary marking, but the strength of boundary-marking through duration varies across corpora, depending upon the degree of spontaneity, as well as cognitive and communicative effort in conversational speech. We have found that the duration of the pause is also significantly correlated with specific boundary status and that syllable duration is inversely correlated with distance to phrase end.
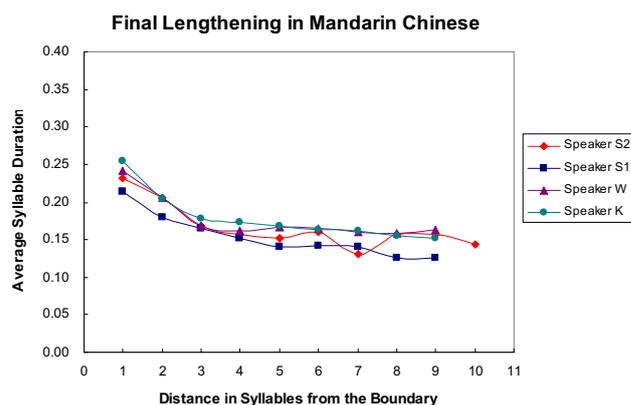
The similar findings for both English and Mandarin Chinese suggest that these durational elements are reflections of deeper universal aspects of language and are rooted in the physiological, cognitive, and social bases of human speech communication. Our findings demonstrate that duration features are a valuable information source on the universal and fundamental elements of narrative and interactive conversational speech across languages.

7. Acknowledgements

8. References

[1] Campbell, W.N., "Synthesizing spontaneous speech," in Computing Prosody, Y. Sagisaka, W.N. Campbell, and H. Norio, eds. Springer-Verlag, 165-186, 1996.

[2] Kompe, R., Prosody in Speech Understanding Systems, Lecture Notes in Artificial Intelligence, Springer. 1997.

[3] Shriberg, E., Stolcke, A., Hakkani_Tur, D., and Tur, G., "Prosody-based automatic segmentation of speech into sentences and topics," Speech Communication, 32(1-2), 127-154, 2000.

[4] van Santen, J.P.H. "Contextual effects on vowel duration," Speech Communication, 11:513-546, 1992.

[5] Chafe, W., "Some reasons for hesitating", in D. Tannen and M. Saville-Troike, ed., Perspectives on Silence, Norwood, N.J.: Ablex, 77-89, 1985.

[6] Oliveira, M. Jr., Prosodic features in spontaneous narratives. Ph.D. dissertation, Simon Fraser University, 2000.

[7] Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P.J., "Segmental durations in the vicinity of prosodic phrase boundaries". JASA91: 1707-1717, 1992.

[8] Yang, L.-C., "Contextual effects on syllable duration," in Advances in Speech Synthesis, Proceedings of the 3rd ESCA Workshop on Speech Synthesis, ed. By Nick Campbell, 1998.

[9] Yang, L.-C., Duration and Pauses as Phrasal and Boundary-marking indicators in Speech. Proceedings of the 15th ICPHS. Barcelona, Spain, 2003.