DURATION, PAUSES AND THE TEMPORAL STRUCTURE OF MANDARIN CONVERSATIONAL SPEECH

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

In this study, we investigate pauses and durational patterns in Mandarin spontaneous conversation, as well as investigate how reliably such elements can serve as boundary-marking predictors across different types of speech modes and how language activities are affected by their cognitive correlates. Our results show that pause duration is significantly correlated with specific boundary status. We show that the duration elements are a fundamental component of discourse organization in spontaneous speech and simultaneously reflect exigencies of both cognitive processes and interactive communicative exchange.

Keywords: Pauses, duration, conversation, cognition, Mandarin

1. INTRODUCTION

The organizational and cognitive aspects of pauses and duration changes are closely linked through the flow of topic ideas, and are usually produced spontaneously, without conscious effort by speakers. They thus provide a potential window to view aspects of cognitive processes that are at work during language communication. Increased knowledge of how these processes work in language is a critical element for achieving a fuller knowledge of human language and human motivations. A comprehensive characterization of pause use in language also provides important information for speech synthesis systems by enabling more natural sounding speech and better detection of speaker intention. Moreover, when combined with cross-language comparisons on pause and duration activity, we can help to discover to what degree variations in these elements can be considered as universal forms reflecting underlying cognitive organization and expression.

In this study, our goal is to investigate the functions and characteristics of pauses and syllables in Mandarin conversation and address the following questions: How do pause usage and pause duration function as indicators of phrasal and discourse organization and to what extent can pauses serve as discourse boundary marking predictors? What are the functions and distributions of pauses in spontaneous conversation and how can pause characteristics and syllable duration information help deepen our understanding of the relationship between thought and speech?

2. DATA AND METHODOLOGY

Data for this research consist of 2 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. The varied data can be expected to shed light on different speech styles and conversational modes and their effects, giving a more complete picture of the use of pauses and duration elements in speech. For this study, we followed the same general principles and methodology as in our work on the durational system of English [5, 6]. Data were segmented to the syllable level, and durational features, including syllable, word, phrase, and pause durations and distance measures, were extracted automatically. From this data, we used a subset of extended selections to focus on duration. Table 1 summarizes the speech time for each selected dataset.

Table 1: Data Summary

Name	Nature of Data	Total Time
MCS1	Spontaneous conversation	16 minutes
MCS2	Spontaneous conversation	20 minutes
MRS3	Read children's stories	7 minutes

For phrase boundary marking, a 2-level categorization scheme differentiating major and minor phrase 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 [5].

To study variations in syllable duration, syllables were classified by distance to the next occurring pause or phrase boundary, and syllable durations were automatically extracted from the segmented data. Incidence, location, and duration of pauses and syllables were then analyzed with respect to discourse activities such as interruptions, speaker turn, prompting, question and answer sequences, etc., as well as identifiable indicators of cognitive activity, including

Table 2: Pause Incidence and Pause Time (sec) Percentage for 3 Speech Corpora

File	Ph	Pauses	BP	NBP	NBP/P	Ph+P	BP/P	TT	TP	TP/TT(%)
MCS1	666	686	428	258	37.6%	64.3%	62%	900	255.9	28.4%
MCS2	695	547	384	163	29.8%	42.9%	70%	1180	168.2	14.2%
MRS3	162	216	159	54	25.0%	98.1%	75%	412	145.5	35.3%

* Ph = phrases P = pauses BP = boundary pauses NBP = non-boundary pauses   TP = total pause time TT = total time

hesitations, feedback markers of agreement and disagreement, and linguistic markers of certainty, doubt, and uncertainty. Combining the acoustic measures of incidence and duration with the analytic marking of speaker state and interactive activities provided the basis for identifying systematic patterns of cognitive aspects of language form that are mirrored in duration variations.

3. BOUNDARY VS. NON-BOUNDARY PAUSES IN MANDARIN

3.1. Factors of 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 [1, 2, 3]. *Phrases* in speech have themselves been associated with semantic/syntactic boundaries, or alternately, with idea units or thought segments [2].

To find out to what extent pauses are used in different types of speech, we calculated the total number of pauses and phrases, and their total time in absolute terms and as percentages. Table 2 shows summary statistics for the 3 sets of data we analyzed.

As shown, the total time spent on pauses as percentages of total conversation time varied from 14.25% to 35.3%. These results are comparable to the pause ratio range reported for English and other languages, e.g., [3] reported values ranging from 0.13% to 0.31% in a set of 17 narratives. In our data the results reflect the different styles of speech as well as different topics. Table 2 indicates a connection between the incidence of pause use as boundary markers and the percent of time pauses constitute of the total time.

One measure of how well pauses function as boundary markers is the percentage of phrases that end with a pause. In our data, MCS2 had the smallest percentage of phrases ending in a pause, 42.9%, and also had the smallest proportion of pauses as a percent of total time, 14.26%. MRS3 had a remarkable 98.1% of its phrases ending in a pause, and the highest percentage of pause time, 35.3%, of the 3 sets of data.

This very high percentage of phrases that end in a pause is very characteristic of read speech. MRS3 consists of read speech of Mandarin children's stories, and the reading is very endearing, but also very structured, with complete ideas encapsulated in phrases and sentences. 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 in less time spent on pauses. 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; pauses were used primarily to mark such structural boundaries and to emphasize and signal the entry of new plot characters into the story. The pauses are seen as conducive to the hearer's abilities to grasp information, necessitating longer durations, while this information needed little cognitive planning on the speaker's side.

From Table 2, we see that there is a significant difference in pause duration as a proportion of total time for MCS1 and MCS2. Detailed analysis of these conversations indicates that the nature of the conversations is quite different in these data. The MCS2 conversation centered on a very recent trip, and the events surrounding the visits, whereas the conversation of MCS1, consisted of a discussion about activities in which both speakers had an intense interest, and the level of speaker involvement was very high. Moreover, there were numerous areas of potential and actual misunderstanding that presented many cognitive difficulties for the participants. By contrast, MCS2 was a non-controversial narrative with little cognitive difficulties in topic development. This was because of the linear time-line of the MCS2 story, the non-critical nature of the subject itself, and the very recent nature of the reported events, which aided memory retrieval. These cognitive factors are reflected in the lower pause duration percentage of 14.2% in MCS2.

Another measure of how well pauses characterize phrase boundaries is the proportion of pauses that occur at phrase end. If pauses frequently occur within the phrase, they do not unambiguously mark a boundary. Conversely, if pauses are infrequent within a phrase, then the occurrence of a pause is likely to indicate a boundary. Table 2 shows clearly that in the cognitively more demanding MCS1, there is a greater proportion of non-boundary pauses (37.6%) than in the other 2 sets.

4. Pause duration and boundary status

4.1. Major, minor, and non-boundary pauses

The strength of pause as a boundary marker is not unambiguous on the basis of occurrence alone, as Table 2 indicates. But pause duration may provide additional information that helps to distinguish boundary pauses and other pauses. Are phrase boundary pauses and phrase internal pauses different enough in duration to characterize boundary and non-boundary marking? Table 3 presents the number of pauses in each category over all of the speakers in the two larger spontaneous conversations and in the read speech data. Table 3 shows that major phrases have the longest pauses on average, with minor phrases considerably shorter, at .33 seconds, and internal pauses having the shortest average duration. Figure 1 presents the average durations of pauses broken out for the 4 speakers in MCS1 and MCS2, by whether the pause occurs internally, at the end of a minor phrase, or at the end of a major phrase.

Table 3: Average Pause Durations by Type

Type	Number	Average Dur
Major Boundary	665	.49
Minor Boundary	265	.33
Non-boundary	393	.24

Interestingly, the pattern seen in Figure 1 is consistent with earlier results that we have previously reported for English [5, 6]. Phrase-internal pause have the shortest average duration, pauses at minor phrase endings are somewhat longer, and major phrase end pauses have the longest duration.

4.2. Distributions of pause duration

We analyzed the data to find if the consistent patterns of pause type average durations are adequate to characterize pause status. We found that the degree to which this is possible is dependent on the nature of the specific conversation. In Figures 2a-d, we break out the averages of Figure 1 by histograms of pause duration by type for each speaker. It is evident from these histograms that differentiation of pause boundary status is stronger in the conversation of MCS1, for both speakers: S1 and W. We postulate that this greater distinction among different pause status arises from the more emotionally involved and cognitively more complex nature of that dialogue, so that degrees of finality are more strongly marked.

The difference between internal pause and boundary pause was much greater for the read stories: the average duration for internal pauses for those readings was 0.285 secs, while the average boundary pause duration was 0.81 seconds. Our data show that in this set of reading data, there is a clear distinction in duration bet-
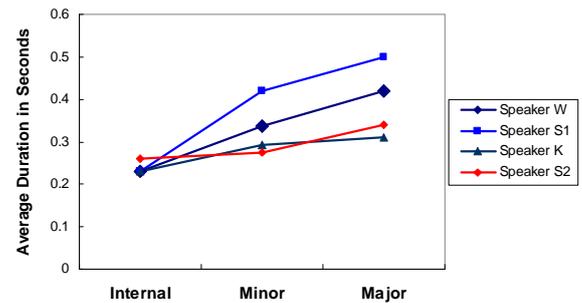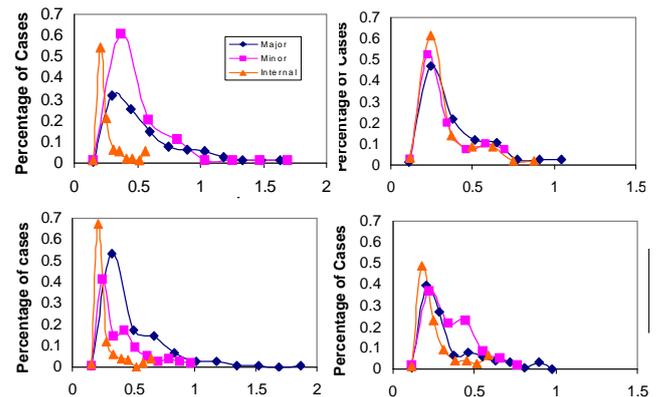

Figure 1: Comparison of Average pause duration by boundary status and by speaker.

Figures 2a-d: Pause duration distinctions by boundary status and by speaker: Phrase-internal pause, minor boundary pause, and major boundary pause. Left column MCS1: Speaker S1 and Speaker W, right column MCS2, Speakers S2 and K.

ween boundary and non-boundary pauses. Detailed analysis reveals that the great separation between pause types is due to the structurally more marked topic presentation, with internal pausing primarily used for introducing new information and for emphasis. There is virtually no overlap in the pause duration, so for this data, pause duration itself characterizes boundary status.

5. FINAL LENGTHENING IN MANDARIN

5.1. Cognitive explanations of final lengthening

Figure 3 shows syllable duration changes broken out by the speakers in MCS1 and MCS2. There is a similar pattern for all four speakers, indicating a consistent pattern of final lengthening starting from about the 5th syllables from the boundary.

The speech rates and final lengthening found here for Mandarin are very consistent with the results that we found previously for English [5, 6], including the similarities in the absolute values for syllable durations. In both our English speech data and our current Mandarin spontaneous conversational data, syllables were at approximately the 0.15 second duration level, then increased from the fifth syllable from the boundary, with the final syllable duration at about 0.25 seconds.
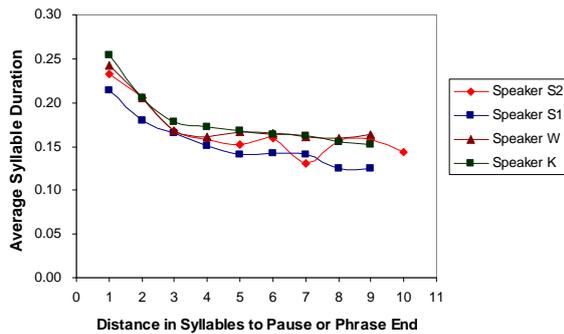
Figure 3: Average syllable duration by speaker and by distance to pause or phrase end

From the duration variations in syllables alone, we can obtain very important information on the coming approach of a boundary. As boundaries have been associated with the completion of idea units [2], we attribute the lengthening of syllables to two causes. The first is that as speakers approach the end of one active idea, there is an attempt to delay, to give time to retrieve the next idea into active, verbally expressible, memory. The effect would be similar in function to pausing, but has the advantage that a continuation of sound holds the floor for the main speaker. The second cause looks at the situation from the interactive exchange point of view: slowing of speech near the boundary enables the *hearer* to understand the segmented thought structure, and provides time to absorb the new information. We suggest both mechanisms are probably at work.

Furthermore, if we take a second look at pausing vs. syllable lengthening, and just focus on the creation of necessary delay time for new ideas to be brought into active memory, we have so far only given an *interactive* explanation of why syllables lengthen rather. If we try to focus on a more *cognitively-oriented* cause, it may be that as speakers approach the end of expressing an active unit, they cognitively begin to devote more mental energy to simultaneously search for the next idea unit, thus lowering the energy devoted to the current one and causing a slow-down in the speech production rate.

5.2. Final lengthening and pause interaction

Viewing pauses and syllable lengthening as distinct separate phenomena provides measures of how important each is in expressing cognitive state and in marking boundary status. However, we can also ask how these different phenomena interact with each other. Does the occurrence of a pause affect the lengthening of syllables or vice-versa? Figure 4 brings out these interactions clearly. From our 2 sets of spontaneous conversation, we recalculated average syllable duration separately for the cases when a phrase ends in a pause, and when a phrase does not end in a pause.
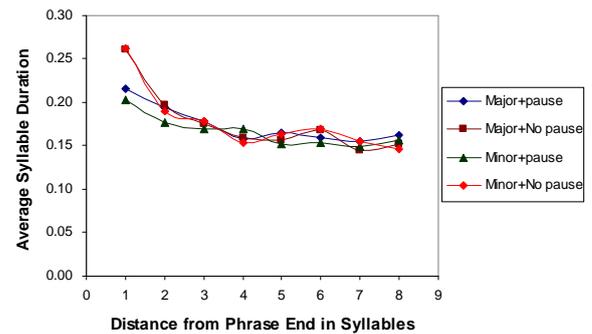

Figure 4: Phrase-end pauses and final lengthening: MCS1+2

In Figure 4 a consistent pattern is found for both minor and major phrases. When a pause ends the phrase, there is *less* lengthening of syllable duration. When there is no pause, syllable duration is clearly lengthened to a greater degree. Cognitively speaking, both pausing and lengthening extend the time, so each buys time and also acts to signal the break. This effect is especially clear on the first syllable prior to the phrase end. This data provides convincing evidence for the essential unity of language duration phenomena.

6. CONCLUSION

In this paper, we have presented evidence that duration can be characterized both at the local phrase and sentence level, as well as across entire conversations. Syntactically, we have seen that pause incidence and duration correlate fairly well with phrase and boundary marking, but that this varies in degree across corpora and speakers. We conclude that other extra-syntactical factors are at work that affect durational elements in discourse and that these cognitive, interactive and syntactic factors function together to provide coherent development of topic and expression of individual state at both global and local levels.

7. REFERENCES

[1] Butterworth, B. 1975. Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research* 4:75-87.
[2] Chafe, W. 1985. Some reasons for hesitating. In D. Tannen and M. Saville-Troike, ed., *Perspectives on Silence*, Norwood, N.J.: Ablex, 77-89.
[3] Oliveira, M. Jr. 2000. *Prosodic features in spontaneous narratives*. Ph.D. dissertation, Simon Fraser University.
[4] Shriberg, E., Stolcke, A., Hakkani_Tur, D., and Tur, G. 2000. Prosody-based automatic segmentation of speech into sentences and topics. *Speech Comm.*, 32, 127-154.
[5] Yang, L.-C. 2003. Duration and Pauses as Phrasal and Boundary-marking indicators in Speech. *Proceedings of the 15th ICPHS*. Barcelona, Spain.
[6] Yang, L.-C. 2004. Duration and pauses as cues to discourse boundaries in speech. *Proceedings of Speech Prosody 2004*. Nara, Japan.