Duration and Pauses as Phrasal and Boundary Marking Indicators In Speech

Li-chiung Yang
† Spoken Language Research Institute and Georgetown University
Washington, D.C., U.S.A.
E-mail: lyang@sprynet.com

ABSTRACT

Duration is a primary factor both to achieve more natural-sounding synthesis and as an indicator of phrasal organization in speech recognition. In this study, we investigate pauses and durational patterns in spontaneous conversation, as well as how reliably such elements can serve as boundary-marking predictors across different types of speech corpora. Our results show that pause duration is significantly correlated with specific boundary status and that syllable duration is inversely correlated with distance to phrase end, suggesting that syllable duration is very significant in predicting phrase boundary status. Our findings show that duration features are highly informative and that it is crucial to integrate such knowledge to enhance performance in spoken language systems.

1. INTRODUCTION

Recent research has focused on providing more precise determinations of pitch, amplitude, and durational features in speech, both to achieve more intelligible and natural-sounding synthesis and as indicators of phrasal organization and intention-marking in speech recognition [1] [3] [4] and [5]. In particular, duration has been approached as both a local segmental and syllable level phenomenon and as a feature exhibiting more global influences. Both researchers in synthesis and recognition have pointed out that duration plays an important role in prosody. Previous work on duration modeling 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. Spontaneous conversational speech can be expected to be even more complex. Researchers in recognition have also identified a variety of relevant features for boundary detection, including pause, duration, final lengthening and laryngealization [7]. As boundary detection is critical to speech understanding, it is important to investigate the role of prosody and duration in natural discourse to isolate the contributing factors.

2. DURATION AND PAUSES: DATA

In this study, our goal is to investigate the timing structure in natural speech and address the following questions: How do pause usage and pause duration function as indicators of phrasal organization and to what extent can pauses serve as boundary marking predictors? How do the functions and distributions of pauses compare across different speech types such as are typical in natural speech settings? And how can pause and syllable duration information be optimally utilized in speech understanding tasks?

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 news magazine, about 15 minutes, and 1 single speaker radio story of about 17 minutes. 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. We used a combination of different criteria to come up with a working categorization scheme for minor or major phrases. 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. BOUNDARY VS. NON-BOUNDARY PAUSES

3.1 Distribution and Frequency of Pauses

Results from our corpora show that pauses correlate fairly well with phrase boundary and that this result is consistent across all data types. Table 1 presents the distribution of pause location and duration across the four different corpora. Over our entire corpus, the percent of pauses that are boundary pauses is consistently high, varying between 60.9% to 88.6%, so that well over 60% of pauses indicate a boundary status. How well pauses can serve as boundary markers is also dependent on how consistently phrases are marked by a pause. The data in Table 1 show that there is considerable variation among the different corpora on this point. In particular, the percent of phrases with a boundary pause is about 35% for both the DS1 and DS2 data, but is

ISBN 1-876346-48-5

closer to double that value for the DS4 and DS3 data. These results suggest that the strength of boundary marking may depend upon a number of other factors such as speaker, gender, and speech style (i.e. degree of spontaneity).

Table 1: Distribution of Pauses by Type

File	Ph	Pause	BP	NBP	Percent	BP/P
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%

The highest proportion of phrases with a boundary pause, 73.3%, occurred with the narrative, a well-developed story with many distinct subtopics. The speaker's narration is marked with a relatively large number of both non-boundary and boundary pauses, and this corresponds to the more structured nature of the narration. There are many internal pauses used for emphasis, rhythmic effects, and hesitation, with phrase-to-phrase development also systematically marked by boundary pauses.

By contrast, 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, the topics are less structured than the story data, and in addition, topic structure and development may rely more on interactive cues and interruptions (for clarification, for example) rather than on phrasal marking by pauses, and these factors might have lead to the lower percentage of phrases marked by pauses in such corpora.

The phrases in DS3 are also well-marked by pauses, with 66.2% of phrases ending in a pause, and 88.3% of pauses are boundary pauses, in other words, when there was a pause, it was almost always a boundary pause; however, about 33.8% of phrases were not marked by pauses at all. In our observation, this is related to the more professional and organized news magazine speech style. The presentation of topics by the main speaker is more structured, and his speech is more grammatically well-formed, with few of the hallmarks of more interactive speech, and not as much rhythmic effect as the story in our corpus.

In our data there is a close relationship between speech rate and pause occurrence in that the relatively fast speakers have proportionally fewer internal pauses. For example, the slower average speech rates for DS4, about .25 seconds per syllable, and DS3 (about .20 sec/syl) contrasted with the faster average speech rate in DS1 (about .17 sec/syl) and DS2 (about .19 sec/syl), and this slower speech rate was also associated with a much greater use of phrase-internal pausing. A reasonable speculation is that in more structured or more formal speech, boundaries are more well-formed and marked by pauses; in more spontaneous or casual speech, boundaries will be less marked by pauses, while speech rate may be a significant factor in the frequency of phrase-internal pauses.

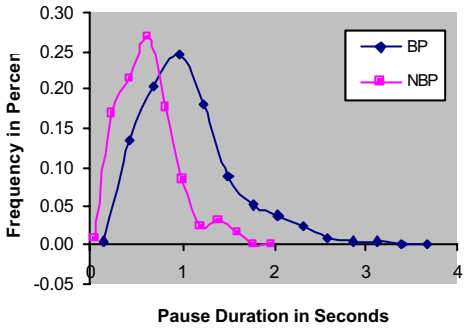

Figure 1: Histogram of boundary and non-boundary pause durations for DS4.

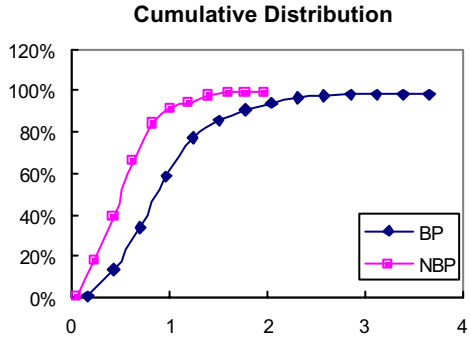

Figure 2: Cumulative distribution for DS4 showing pause duration differences by boundary status for DS4.

3.2 Duration of pauses

What role does pause duration play in marking phrase boundaries? In Figure 1, we show the distribution of pause duration for boundary vs. non-boundary pauses for one corpus – the 17-minute extended narrative DS4. This histogram shows pause duration in seconds by frequency counts of boundary and non-boundary pauses as percentages (P=385, BP=254, NBP=131). The overall greater average length of boundary pauses is evident in the histogram, with the highest frequency occurring at about .5 seconds for non-boundary pauses and at about 1.0 seconds for boundary pauses. The histogram shows that the longer the pause is, the greater the chance that it is a boundary pause. The overlap in duration between boundary and non-boundary pauses can be seen as well, and implies that if the pause has a reasonable duration, it is harder to tell whether it is a boundary pause or a non-boundary pause. For example, 22% of non-boundary pauses have duration of about .3 seconds, but about 13% of boundary pauses are about .3 seconds long as well. The associated cumulative distribution shown in Figure 2 shows this result clearly as well. While over 60% of boundary pauses have less than about .5 seconds duration, only about 20% of boundary pauses are less than .5 seconds in duration. Thus, longer duration, especially above 1 second in length, will tend to signal a phrase boundary.

ISBN 1-876346-48-5

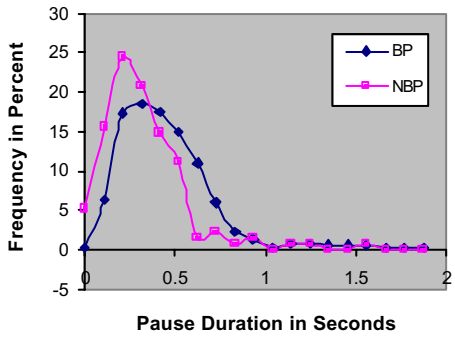

Figure 3: Histogram of boundary and non-boundary pause durations for DS1-DS3.

When we look at the data for other speakers in our corpus, we find that speaking style has a significant effect on pause duration. In Figure 3 we show the percentage histogram of pause duration for other speakers in our corpora (P=524, BP=389, NBP=135). The histogram exhibits a similar pattern: the longer the duration, the greater chance that a pause is a boundary pause. But for this data, durations for both types of pauses are much less than for the story. Non-boundary pause durations here virtually end at about .6 seconds, rather than at about 1.0 seconds for the story, and there are very few boundary pauses greater than .8 seconds. The peak percents for both occur at correspondingly lower durations as well, at .21 seconds and .31 seconds, respectively. It is clear that the 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. In both cases, longer durations are more likely to occur at a boundary. However, pause duration by itself is not enough to unambiguously distinguish boundary and non-boundary pauses.

4. BOUNDARY STATUS – MAJOR, MINOR AND NON-BOUNDARY PAUSES

To further investigate the role of pause duration in structuring discourse, we calculated pause durations corresponding to minor phrase boundary, major phrase boundary, and phrase-internal pauses. Table 2 presents average duration in each category for 3 dialogues of our corpus.

Table 2: Average Pause Durations by Type

TYPE	NUMBER	AVERAGE DUR
Major Boundary	206	.461908 sec
Minor Boundary	183	.354539 sec
Non-boundary	135	.277070 sec

Our data show that the duration of the pause is also significantly correlated with specific boundary status in that the longest pauses occur on major phrase boundaries, while shorter pauses accompany minor phrase boundaries, and non-boundary pauses have the shortest durations on average. For example, on average major phrase break pauses are longer than minor boundary pauses, at .46 seconds and .35 seconds respectively. Non-boundary pauses are the shortest, at .28 seconds.

When we break out these results by speaker, we can clear see that most speakers follow this pattern consistently, however there are also individual differences in the relative durations of major phrase pauses, minor phrase pauses, and non-boundary pauses. Figure 4 shows the average duration by major phrase pause, minor phrase pause, and internal pause by speaker for three datasets in our corpus. For 5 of the 8 speakers the expected pattern holds. Unexpectedly, for 3 of 8 speakers, however, the average duration for minor phrases actually went up. The non-proportional sample from several speakers and the occurrence of disfluencies are possible reasons to account for this result.

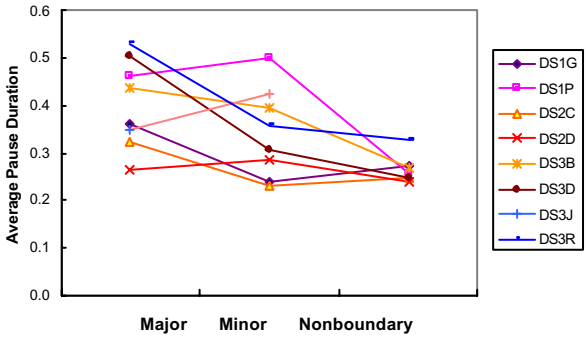

Figure 4: Comparison of average pause duration by boundary status and by speaker.

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, particularly with read speech, and there has also been some debate on whether final lengthening is confined only to the last syllable and where within the syllable lengthening occurs [3] [5] [7] and [8]. Inspired by the importance of finding reliable prosodic cues to phrase boundary status, we looked into alternative duration measures that may function as boundary markers in spontaneous speech. We calculated average syllable durations by speaker and by closeness to phrase end for the entire corpus. Our data show that the final syllable before the phrase boundary has the longest duration, and that syllable duration is *inversely* correlated with distance to phrase end. This suggests that syllable duration is very significant in predicting phrase boundary status in speech.

Figure 5 shows average syllable duration by distance to phrase end in syllables for all eleven speakers in our corpora. The effects of final lengthening can be seen clearly by the rise in syllable duration the closer the syllable is to the end. It is clear from the Figure 5 that lengthening in

ISBN 1-876346-48-5

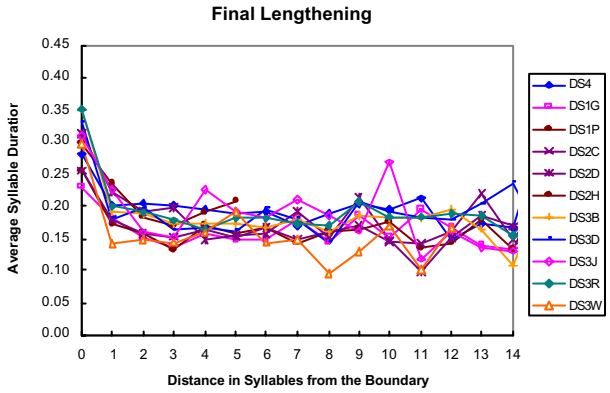

Figure 5: Average syllable duration by speaker as a function of distance to phrase end in syllables.

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. This result is consistent across all of the speakers in our data and provides 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. DISCOURSE FUNCTIONS OF PAUSES: WHY DO PAUSES OCCUR IN SPEECH

Boundary pauses constitute a large proportion of the pauses in our corpus, and if their use for boundary marking were fully consistent and complete, prediction of phrase boundaries would be a relatively simple task. However, as seen from the data, non-boundary pauses constitute about 1/3 of all pauses in the corpus, and in addition, phrase endings are not always marked by a pause. This complication is a direct manifestation of the greater functionality that pauses have in spontaneous conversation. Pauses in conversation function as interactive signals for turn-taking and suggested topic direction, and are also used as expressive elements in discourse, especially for emphasis or dramatic effect and for building up tension and climax. This effect was particularly strong for the storyteller in his narrative, and often acted as a punctuated sequence of emphasized points, where the use of pauses for rhythmic effect is particularly prominent.

The organization of discourse in spontaneous speech occurs not only through semantic relationships among phrases, but also through exigencies of both cognitive constraints and interactive negotiations. The on-line topic redirection and memory search requirements frequently require time to coordinate, and pauses are frequently used to hesitate in these circumstances or in situations of uncertainty or doubt [2]. Such environments often persist for a time in a conversation, and the extended time domain over which the associated cognitive state persists often causes pauses of a given type, whether boundary or non-boundary, to occur in clusters, a phenomenon that was frequently observed in our data. For this same reason, disfluencies, discourse markers, and pauses have a natural affinity for occurring together. In our corpus, pauses frequently co-occurred with discourse markers such as 'and', 'but', 'then', 'so', you know' as well as interjections and particles. In such instances, both the discourse marker and the pause may act to provide time for cognitive refocusing at key transition points of topic development.

7. CONCLUSIONS

In this paper, 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 constrainedness and the rhythmic structure of the specific speech modality. 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. Our results are consistent with previous research in descriptive work as well as with recent findings in speech segmentation, pointing to the importance of duration in speech recognition. Our findings demonstrate that duration features are a valuable knowledge source and that it is crucial to integrate such knowledge to enhance performance in spoken language systems.

REFERENCES

[1] W.N. Campbell, "Synthesizing spontaneous speech," in *Computing Prosody: Computational Models for Processing Spontaneous Speech*, Y. Sagisaka, W.N. Campbell, and H. Norio, eds. Springer-Verlag, 165-186, 1996.

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

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

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

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

[6] J.P.H. van Santen. "Assignment of segmental duration in text-to-speech synthesis," *Computer Speech and Language*, 8: 95-128, 1994.

[7] C.W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, and P.J. Price, "Segmental durations in the vicinity of prosodic phrase boundaries". Jou*rnal of the Acoustical Society of America,* 91: 1707-1717, 1992.

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