Predictors of Pause Duration in Read-Aloud Discourse

Xiaohong YANG††, Member, Mingxing XU†, and Yufang YANG†, Nonmembers

SUMMARY The research reported in this paper is an attempt to elucidate the predictors of pause duration in read-aloud discourse. Through simple linear regression analysis and stepwise multiple linear regression, we examined how different factors (namely, syntactic structure, discourse hierarchy, topic structure, preboundary length, and postboundary length) influenced pause duration both separately and jointly. Results from simple regression analysis showed that discourse hierarchy, syntactic structure, topic structure, and postboundary length had significant impacts on boundary pause duration. However, when these factors were tested in a stepwise regression analysis, only discourse hierarchy, syntactic structure, and postboundary length were found to have significant impacts on boundary pause duration. The regression model that best predicted boundary pause duration in discourse context was the one that first included syntactic structure, and then included discourse hierarchy and postboundary length. This model could account for about 80% of the variance of pause duration. Tests of mediation models showed that the effects of topic structure and discourse hierarchy were significantly mediated by syntactic structure, which was most closely correlated with pause duration. These results support an integrated model combining the influence of several factors and can be applied to text-to-speech systems.

key words: pause duration, syntactic structure, discourse hierarchy, topic structure, phrase length

1. Introduction

In Text-To-Speech (TTS) systems, it is very important to predict prosody information for the production of natural sounding speech. Pause modeling, which is one of the most important parts of prosodic modeling, is particularly relevant to this aim. For the past decades, many publications on pause modeling have focused primarily on the predictions of phrase break [1]–[8] and pause duration between phrases [9]–[14]. Several factors have been identified to be crucial for pause modeling at sentence level: morphosyntactic features [5], syntactic structure [4], contextual length [15], textual distance [6], speaking style [12], etc. Although far less attention has been given to pause modeling at paragraph or discourse level, there have been a few promising attempts over the last few years. Smith [16] and Bruce et al. [17] incorporated the topical organization of a text as an important factor of pause duration in TTS systems. Tseng and her colleagues [18], [19] proposed a hierarchical multi-phase framework to account for the contributions of higher level discourse information on pause variation. More recently, Doukhan et al. [15] reported that including contextual length as a factor of pause duration improved performance in TTS systems. These researches have made considerable progress in improving pause modeling. However, the performances of synthesizers are still poor for the synthesis of longer texts. This should not be surprising, because when it comes to paragraph or discourse prosody, far less is known about the factors at play. The present study, by taking into account the contributions from different discourse factors, seeks to identify the major predictors of pause duration in read-aloud discourse. Findings of the present study can be useful for predicting pause duration in TTS systems.

Previous studies on discourse prosody have converged on a set of factors that influence discourse prosody. Earlier studies of discourse prosody mainly relied on the distinction between sentences and paragraphs [20]–[27]. For instance, it has been found that paragraph boundaries and sentence boundaries are virtually always marked with pauses while only some of the clause boundaries and phrase boundaries are marked with pauses [26], [27].

Other studies, however, tried to identify how the hierarchical structure of discourse modulates discourse prosody [28]–[30]. One way to build a hierarchical structure of discourse is to rely on perceived boundary strength. For example, in Swerts [28], subjects were told to draw a line between the boundaries of two paragraphs while no explicit definitions of boundary were given. Then boundary strength was defined in terms of the proportion of subjects indicating that there was a break between discourse segments. It was found that stronger boundaries marked by subjects tended to be related to longer pause duration, greater pitch reset, and lower boundary tone.

However, a more popular way to build a discourse hierarchy is to resort to sophisticated theoretical models such as the theory of Grosz & Sidner [31]–[37] and Rhetorical Structure Theory (hereafter referred to as RST) [38]–[42]. Both theories established discourse hierarchy as one of the most important aspects of discourse structure. According to these two theories, building a discourse hierarchy involves similar procedures: the analyst first segments the discourse into discourse units; then adjacent discourse units that are more closely related are linked into a larger discourse unit. This discourse unit is then related to other discourse units to create an even larger unit. This process can go on and...
on until finally the discourse is represented as a hierarchical structure. Acoustic studies have showed that the strength of prosodic realization is correlated to the position of a segment in the discourse hierarchy [38]–[40].

There are also other studies that established discourse structure on the basis of topic structure [16], [43]. Topic structure defines the type of transitions between speech turns or sentences. Transitions are classified into four categories (topic shift, topic continuation, topic elaboration, and speech-act continuation) depending on whether they reflect a shift in topic, a continuation of the same topic, and so on. It has been found that topic shift boundaries were associated with the strongest prosodic marking compared with the other categories [16], [43].

Overall, the above studies have identified several factors that affect prosodic features in discourse context. However, these studies are lacking in two areas. The first deficiency of these studies is that they only considered a simple factor and did not clarify how such factors related to other potential factors which may also affect prosody. As mentioned above, a natural discourse involves several factors that may have consequences for prosody, including discourse hierarchy [38]–[40], syntactic structure [20]–[27], and topic structure [16], [43], among others. Whether these factors interact or act independently to influence discourse prosody awaits further investigation. The second deficiency is related to the confound effects that arise when only one factor is considered and others neglected. The common practice of previous studies is to look into one factor and determine how it influences prosodic features. An intrinsic danger, therefore, is that the influence of one factor on prosody may be confounded with influences from other factors.

In order to improve on these earlier studies, we accommodate several factors in read aloud discourse and investigate how these factors affect prosodic variations both separately and jointly. We focus on pause duration, which is one of the most important parts of prosodic modeling. Even though other prosodic features such as pitch range or final lengthening are also important for prosodic modeling, we will not discuss them in this paper because different prosodic features are related to different sets of factors and a full-scale study is beyond the scope of this paper. Insofar as factors affecting pause duration of discourse boundary is concerned, we take into consideration the factors that have been found to affect pause duration in previous literature: syntactic structure [21]–[27], topic structure [16], [43], and discourse hierarchy [38]–[40]. Although the effect of sentence length in discourse context has not yet been reported in the literature, it has been shown that pauses were significantly affected by pre and postboundary phrase length in sentence strings [44] and benefits have been reported when the length factor has been considered in TTS systems [15]. Therefore, we also take into consideration the effects of another two factors: preboundary length and postboundary length. The goal of the present study is two-fold: (1) a practical concern of exploring the contributions of these factors for the prediction of pause duration in discourse context and (2), a theoretical interest of investigating whether there are some overlapping effects among these factors.

2. Method

2.1 Text Materials

The same Mandarin Chinese speech corpus used in Yang and Yang [39] was used in the present study. The corpus contained ten paragraphs (150-200 words per paragraph) of news commentaries. Each paragraph was read twice by five native speakers of Mandarin Chinese, which resulted in ten recordings for each paragraph (2 repetitions × five speakers) and a total of 19078 words for the corpus.

2.2 Labeling of the Materials

2.2.1 Labeling of Discourse Hierarchy

As mentioned earlier, both theories of Grosz & Sidner [31]–[37] and Rhetorical Structure Theory [38]–[42] can be used to produce discourse hierarchy. However, we adopted RST to model discourse hierarchy because this theory has been found to produce higher inter-analysts agreements than the theory of Grosz & Sidner [41]. Five graduate students from the Institute of Psychology, Chinese Academy of Sciences were recruited to participate in the labeling of discourse hierarchy according to the guidelines laid out by RST [42] in the RSTTool1345 software. Take the example text for instance (please see Table 1 and Fig. 1), on the basis of semantic relatedness, unit 1 and unit 2 were first related to create a

Table 1  Example paragraph used in the present study. (translated from Chinese to English)

1 At present, many middle schools open a variety of elective courses,
2 which has been well received from students and parents.
3 One middle school in Chaoyang district in Beijing has opened nearly a hundred elective courses for senior students,
4 which takes students a long time to choose,
5 and assures the parents that the school is not just aiming at exam-oriented education.
6 However, suffered from the pressure of college entrance examination,
7 most students tend to choose exam-related courses.
8 For instance, the geometric course was meant to recruit forty students,
9 but students registering for the course outnumbered one hundred,
10 after having opened one more class for the course,
11 the school had to distribute the rest of the students to other courses.
12 On the contrary, some exam-unrelated quality courses were either unpopular or neglected,
13 therefore, the school had to cancel the opening of these courses.
text unit (unit 1-2). For unit 3 to 5, unit 4 and 5 were first related to create a text unit, which was in turn related to unit 3 to create a larger unit (unit 3-5). Then unit 1-2 was related to unit 3-5 to create an even larger unit (unit 1-5). This process went on until all units were enclosed in a discourse hierarchy. For more details concerning the labeling of discourse hierarchy, please refer to Yang and Yang [39]. The resulting hierarchical scores of the corpus ranged from 3 to 9. The higher the hierarchical scores, the deeper the boundary was embedded in the discourse hierarchy.

2.2.2 Labeling of Topic Structure

Three analysts annotated the topic structure of the corpus. They were graduate students from the Institute of Psychology, Chinese Academy of Sciences. None of them had participated in the labeling of discourse hierarchy. They were instructed to annotate discourse boundaries into topic shift boundaries (labeled as 0) or topic continuation boundaries (labeled as 1). Take the text in Table 1 for instance, all the three analysts labeled the boundary between unit 4 and 5 as a topic continuation boundary and the boundary between unit 5 and 6 as a topic shift boundary. Where there was disagreement, the majority opinion was taken. Weighted kappa statistics of agreement [45] was computed to test inter-labeler reliability. Results indicated high inter-rater agreement (Kappa = 0.60, p = 0.000). We were aware that previous studies segmented discourse boundary into four kinds of topic relation: topic shift, topic continuation, topic elaboration, and speech-act continuation. The reason to abandon topic elaboration and speech-act continuation in the present study is that these two kinds of topic relation are often confounded with topic continuation, and the inclusion of these two relations would reduce inter-rater agreement (as indicated by a kappa of 0.28 in the study of Smith [16]).

2.2.3 Labeling of Syntactic Structure and Length

Syntactic structure and length of segments were labeled by the first author of the present study. Based on whether a segment was ended with a comma or a period, we categorized syntactic structure into two kinds: clause boundary or sentence boundary, which was labeled with 0 or 1 respectively. For example, the boundary between unit 4 and 5 in the example text was labeled as a clause boundary (marked with 0) and the boundary between unit 5 and 6 was labeled as a sentence boundary (marked with 1). Preboundary and postboundary lengths were labeled in terms of the number of syllables in the preboundary and postboundary segments respectively. Longer segments received higher scores.

2.3 Acoustic Measurement and Data Analysis

One recording of paragraph 1 and paragraph 2 were discarded due to speaker errors. Pause duration was defined as the silence interval between the ending of one discourse segment and the beginning of the next segment. In order to pool the data of the five speakers, the extracted duration values of the five speakers were normalized into Z scores by the following equation:

\[ Z = (X - M) / SD \] (1)

In Eq. (1), X refers to the extracted duration value of a specific syllable. M and SD refer to the mean duration value and standard deviation of all syllables in all paragraphs produced by a speaker.

For each discourse boundary, the values of pause duration were averaged over five speakers and two repetitions.

3. Results

3.1 Simple Linear Regression

To explore how pause duration varies as a function of the five factors (discourse hierarchy, syntactic structure, topic structure, preboundary length, and postboundary length), simple linear regression analysis was conducted with the five factors as independent variables one at a time and pause duration as dependent variable. The results are shown in Table 2.

As shown in Table 2, except preboundary length, all the other four factors were found to exert significant influence on pause duration. Specifically, pause duration was longer at sentence boundaries than at clause boundaries, at boundaries of higher hierarchy than at boundaries of lower hierarchy, and at boundaries of topic shift than at boundaries of topic continuation. Finally, pause duration increased as a function of postboundary length. Note that the values of adjusted \( r^2 \) were 0.763, 0.736, and 0.592 for syntactic structure, discourse hierarchy, and topic structure respectively. This suggests that syntactic structure and discourse hierarchy can account for over 70% of pause variation while topic structure can account for about 60% of pause variation. Therefore, it appears that all these three variables were strong indicators of pause duration. We hypothesized that

\(^1\)Note that in Yang and Yang [39], we did not average the values of pause duration over speakers and repetitions because we were interested in the individual differences across speakers. In the current study, however, the ten recordings were averaged across speakers and repetitions to reduce error variance caused by individual differences in the regression models.
there may be some overlapped effects in these three factors. To look at the overlapped effects of these factors, we performed stepwise multiple linear regression analysis with the four significant factors entered together as independent variables and pause duration as dependent variable. The stepwise multiple linear regression procedure allows only the variables that explain additional variance to enter the regression model and it ends when the best regression model is built.

### 3.2 Stepwise Multiple Linear Regression

Table 3 shows the results for the stepwise multiple linear regression analysis. As shown in Table 3, the best model was Model 3, which included syntactic structure, discourse hierarchy, and postboundary length. This model had a regression coefficient of 0.895 and could account for about 80% of the variance of pause duration (adjusted $r^2 = 0.796***$).

In Model 1 in which only syntactic structure was included, the model could account for about 76% (adjusted $r^2 = 0.763***$) of the pause variance. Compared with model 1, model 2 could account for 2% more of pause variance when discourse hierarchy was added into the model (adjusted $r^2 = 0.786***$). Adding both discourse hierarchy and postboundary length in Model 3 only improved about 3% of the outcome. These results suggest that syntactic structure is the strongest predictor of pause duration.

Interestingly, while topic structure and discourse hierarchy were found to have substantial impacts on pause duration when simple linear regression was conducted, topic structure was excluded from the stepwise multiple linear regression models, and the effect of discourse hierarchy was also substantially reduced. For the exclusion of topic structure, we hypothesized that the effect of topic structure was mediated by syntactic structure, discourse hierarchy, and postboundary length in such a way that when these mediators were included in the models, the effect of topic structure was greatly reduced. With regard to discourse hierarchy, we hypothesized that it was mediated by syntactic structure and postboundary length so that its effect was greatly reduced when the mediators were included in the model.

### 3.3 Tests of the Mediation Effects Caused by Syntactic Structure

To test our hypothesis of the mediation effects, we proposed two mediation models: (1) the first model testing the hypothesis that the effect of topic structure was mediated by syntactic structure, discourse hierarchy, and postboundary length (Fig. 2), and (2) the second model testing the hypothesis that the effect of discourse hierarchy was mediated by syntactic structure and postboundary length (Fig. 3).

Tests of the mediational model allow the total effect of the predictor variable on the outcome variable to be decomposed into specific indirect effects (the effect caused by the mediating variables) and direct effects (the effect caused by the predictor variable itself). If the indirect effects are significant, mediation is established when the direct effects are significantly smaller than the total effect [46]. To test the indirect effects, the method of bootstrapping was used [47].

![Fig. 2](image_url)  
**Fig. 2** Mediation model for the relation between topic structure and pause duration.

![Fig. 3](image_url)  
**Fig. 3** Mediation model for the relation between discourse hierarchy and pause duration.
tests of mediation model showed that the effects of topic structure and discourse hierarchy were mediated by the effect of syntactic structure. These results suggest that boundary pause durations are jointly influenced by a set of factors and that syntactic structure contributes the most to boundary pause duration in read-aloud speech.

4.1 Influential Factors of Pause Duration

Results for simple regression analysis showed that discourse hierarchy, syntactic structure, topic structure, and postboundary length affected pause duration significantly. Specifically, we found that pause duration was longer at sentence boundaries than at clause boundaries, at boundaries of higher hierarchy than at boundaries of lower hierarchy, and at boundaries of topic shift than at boundaries of topic continuation. Finally, pause duration increased as a function of postboundary length. This result is consistent with findings of previous studies [21]–[41], [43], [44], suggesting that boundary pause durations are jointly influenced by a set of discourse factors. These impacts of the structural factors can be driven by the look-ahead and preplanning effect of speech production. Namely, speech is encoded over structurally defined units and pauses can be longer when speakers pause longer to plan a more complex structure or a longer utterance [44].

4.2 The Mediation Effect of Syntactic Structure

Interestingly, different from the results for simple regression analysis, tests of the stepwise regression models revealed that the only salient factor affecting pause duration was syntactic structure. With further tests of mediation models, we found that the effects of topic structure and discourse hierarchy were greatly mediated by the effect of syntactic structure. These results suggest that most of the pause variation related to discourse hierarchy or topic structure can be accounted for by the effect of syntactic structure. This compares favorably with previous observations which showed that syntactic clause boundary was the strongest predictors of boundary perception in spontaneous speech [48]. Note that a strong effect of syntactic structure does not rule out the role of discourse hierarchy or topic structure, but make it clear that pause variations induced by discourse hierarchy and topic structure largely overlapped with those caused by syntactic structure. As suggested by previous literature, boundaries at higher hierarchy are related to larger discourse transitions such as topic shift boundaries [38], which are mostly sentence boundaries, but not clause boundaries. It is therefore reasonable that syntactic structure was found to be most closely correlated with pause variation.

4.3 The Effects of Preboundary and Postboundary Length

While postboundary length was found to affect pause duration, we found that preboundary length did not influence pause duration. This suggests that in read aloud speech,

Table 4 Bootstrap results for the indirect effects of the first mediation model.

|                | B     | SE   | p      | Confidence interval |
|----------------|-------|------|--------|--------------------|
| Total          | 2.10  | 0.22 | <0.001 | 1.72, 2.64         |
| Syntactic structure | 1.13  | 0.29 | <0.05  | 0.62, 1.76         |
| Discourse hierarchy | 0.88  | 0.26 | <0.05  | 0.39, 1.43         |
| Postboundary length | 0.08  | 0.05 | 0.06   | 0.00, 0.23         |

Table 5 Bootstrap results for the indirect effects of the second mediation model.

|                | B     | SE   | p      | Confidence interval |
|----------------|-------|------|--------|--------------------|
| Total          | 0.42  | 0.07 | <0.001 | 0.25, 0.60         |
| Syntactic structure | 0.40  | 0.07 | <0.05  | 0.23, 0.57         |
| Postboundary length | 0.01  | 0.01 | 0.08   | 0.00, 0.05         |

esis is partially confirmed: an overall mediation effect between topic structure and pause duration was found which was largely due to the mediations of syntactic structure and discourse hierarchy. This suggests that the overall effect of topic structure is overlapped with the combined effect of syntactic structure and discourse hierarchy. This explains why topic structure was excluded from the stepwise multiple linear regression models.

Test of the second mediation model

As predict, the effect of discourse hierarchy on pause duration was greatly reduced when the mediators were included ($B = 0.35$, $SE = 0.08$, $p < 0.001$), suggesting partial mediation. As shown in Table 5, the total indirect effect caused by syntactic structure and postboundary length was significant ($p < 0.001$). The bootstrap procedure further revealed that the indirect effects of syntactic structure were significant ($p < 0.05$), while the indirect effect of postboundary length was not significant ($p = 0.08$). Thus, again, our second hypothesis was partially confirmed: a partial mediation effect was found in such a way that the effect of discourse hierarchy on pause duration was partially mediated by syntactic structure, but not by postboundary length. This again explains why the effect of discourse hierarchy was greatly reduced when syntactic structure was included in the stepwise multiple linear regression models.

4. Discussion

Previous studies have found that factors such as discourse hierarchy, syntactic structure, topic structure, preboundary length, and postboundary length significantly affected pause duration [21]–[41], [43], [44]. Following these studies, we examined how these factors influenced pause duration both separately and jointly. Results from simple regression analysis showed that discourse hierarchy, syntactic structure, topic structure, and postboundary length had significant impacts on boundary pause duration. However, while syntactic structure remained the most salient factor in the stepwise regression models, topic structure was excluded from the stepwise regression models and the effects of discourse hierarchy and postboundary length were greatly reduced. Further
longer pauses may be more a reflection of planning upcoming information than of consolidating preceding information. With regard to the effect of pre-boundary phrase length, previous studies have yielded inconsistent results. Some researchers have found a positive correlation between pause duration and pre-boundary phrase length [44, 49, 50]. Others, however, suggest that the relationship between pause duration and pre-boundary phrase length is not so straightforward. Zvonik and Cummins [51] have found that short pauses of less than 300 ms are associated with the length of pre-boundary phrase while for pauses of longer than 300 ms, there is no predictability. The present study differs from previous studies in that the effect of pre-boundary phrase length was investigated in discourse context rather than in isolated sentences. Our data add to the literature by demonstrating that pauses were not affected by preboundary length in read aloud discourse. The results that pause duration was not affected by length of preceding segment but by postboundary length appear to signal an asymmetric influence of the right vs. left domain of the boundary on the variance of pause duration. That is, the right edge of the boundary sheds more influence than the left edge.

5. Limitations and Conclusions

The regression model that best predicts boundary pause duration in discourse context was the one that first included syntactic structure, and then included discourse hierarchy and postboundary length. This model can account for about 80% of the variance of pause duration. Clearly, these factors played a major role in determining the variance of pause duration. However, there may be other factors at play that are not yet included in this model. Due to the limitation of the present material, we have neither accounted for factors related to speaking task, nor syntactic complexity factors related to syntactic branching of the preceding and upcoming phrase. We expect significant improvement on the performance of the models when these factors are included.

To summarize, this work has two important implications. First, while the present study is limited in that only one prosodic variable, i.e., pause duration, was investigated, our results support an integrated model combining the influence of several factors and we believe that this applies to the modeling of other prosodic features as well. Second, instead of attempting to determine the prosodic correlates of a simple factor, we took a different approach, which is distinctive in its attempt to accommodate multiple factors. Although prior studies have found robust effects of structural factors on prosody, our results advance our understanding about how these effects arise. Thus, we believe that the approach used in the present study can lead to a better understanding of why and how prosodic features vary greatly in read aloud discourse.

References

[1] S. Lee and Y.-H. Oh, “Tree-based modeling of prosodic phrasing and segmental duration for Korean TTS systems,” Speech Commun., vol.28, pp.283–300, 1999.
[2] P. Taylor and A.W. Black, “Assigning phrase breaks from part-of-speech sequences,” Computer Speech & Language, vol.12, pp.99–117, 1998.
[3] J. Hirschberg and P. Prieto, “Training intonational phrasing rules automatically for English and Spanish text-to-speech,” Speech Commun., vol.18, pp.281–290, 1996.
[4] P. Koehn, S. Abney, J. Hirschberg, and M. Collins, “Improving intonational phrasing with syntactic information,” Proc. ICASSP-00, pp.1289–1290, Istanbul, 2000.
[5] F.C. Díaz, J. van Santen, and E.R. Banga, “Integrating phrasing and intonation modelling using syntactic and morphosyntactic information,” Speech Commun., vol.51, pp.452–465, 2009.
[6] K. Yoon, “A prosodic phrasing model for a Korean text-to-speech synthesis system,” Computer Speech & Language, vol.20, pp.69–79, 2006.
[7] M. Chu and Y. Qian, “Locating boundaries for prosodic constituents in unrestricted Mandarin texts,” Computational linguistics and Chinese language processing, vol.6, pp.61–82, 2001.
[8] E. Navas, I. Hernández, and N. Ezeiza, “Assigning phrase breaks using CARTs for Basque TTS,” Speech Prosody Aix-en-Provence, France, 2002.
[9] S. Lee and Y.-H. Oh, “Tree-based modeling of prosodic phrasing and segmental duration for Korean TTS systems,” Speech Commun., vol.28, pp.283–300, 1999.
[10] S.D. Mandal, A. Saha, T. Basu, K. Hirose, and H. Fujisaki, “Modeling of sentence-medial pauses in bangla readout speech: occurrence and duration,” INTERSPEECH-2010, pp.1764–1767, Makuhari, Chiba, 2010.
[11] S. Acharya and S.K.D. Mandal, “Prosody modeling: A review report on Indian language,” Mining Intelligence and Knowledge Exploration, pp.831–842, Springer, 2013.
[12] A. Parlikar and A.W. Black, “Modeling pause-duration for style-specific speech synthesis,” ARCTIC, vol.56, p.130, 2012.
[13] Sudipta Acharya and S.K.D. Mandal, “Occurrence and duration modeling of sentence medial pause for Bangla text reading at different speech rate,” O-COCOSDA 2012, pp.62–67, 2012.
[14] N. Kaiki and Y. Sagisaka, “Pause characteristics and local phrase-dependency structure in Japanese,” Second International Conference on Spoken Language Processing, 1992.
[15] D. Doukhan, A. Rilliard, S. Rosset, and C. d’Alessandro, “Modelling pause duration as a function of contextual length,” INTERSPEECH, 2012.
[16] C.L. Smith, “Topic transitions and durational prosody in reading aloud: production and modeling,” Speech Commun., vol.42, pp.247–270, April 2004.
[17] G. Bruce, M. Filipsson, J. Frid, B. Granström, K. Gustafson, M. Horne, and D. House, “Modelling of Swedish text and discourse intonation in a speech synthesis framework,” Intonation: Analysis, modelling and technology, pp.291–320, 2000.
[18] C. Tseng and C.H. Chang, “Pause or no pause?: Prosodic phrase boundaries revisited,” Tsinghua Science & Technology, vol.13, pp.500–509, 2008.
[19] C.Y. Tseng, S.H. Pin, Y. Lee, H.M. Wang, and Y.C. Chen, “Fluent speech prosody: Framework and modeling,” Speech Commun., vol.46, pp.284–309, July 2005.
[20] I. Lehiste, “The phonetic structure of paragraph,” in Structure and Process in Speech Perception, ed. A. Cohenand and S. Nooteboom, pp.195–206, Berlin, Germany: Springer-Verlag, 1975.
[21] G. Bruce, “Textual aspects of prosody in Swedish,” Phonetica, vol.39, pp.274–287, 1982.
[22] D.R. Ladd, “Declination “reset” and the hierarchical organization of utterances,” J. Acoustical Society of America, vol.84, pp.530–544, 1988.
[23] A.M.C. Sluiter and J.M.B. Terken, “Beyond sentence prosody: Paragraph intonation in Dutch,” Phonetica, vol.50, pp.180–188.
YANG et al.: PREDICTORS OF PAUSE DURATION IN READ-ALOUD DISCOURSE

1993.

[24] N.G. Thorsen, “Intonation and text in standard danish,” J. Acoustical Society of America, vol.77, pp.1205–1215, 1985.

[25] N.G. Thorsen, “Sentence intonation in textual context—Supplementary data,” J. Acoustical Society of America, vol.80, pp.1041–1047, 1986.

[26] E. Strangert, “Where do pauses occur in texts read aloud?”, Twelfth Scandinavian Conference of Linguistics, pp.403–414, 1990.

[27] E. Bringmann, “The distribution of Dutch reading pauses: a preliminary investigation on the influence of prosodic phrasing and punctuation on pause duration,” Doctoral paper in manuscript Utrecht University, 1990.

[28] M. Swerts, “Prosodic features at discourse boundaries of different strength,” J. Acoustical Society of America, vol.101, pp.514–521, Jan. 1997.

[29] M.E. Van Donzel and F.J. Koopmans, “Perception of discourse boundaries and prominence in spontaneous Dutch speech,” in Working Papers-Lund University Department of Linguistics, pp.5–24, 1997.

[30] M. Swerts, “Filled pauses as markers of discourse structure,” J. Pragmatics, vol.30, pp.485–496, 1998.

[31] B.J. Grosz and C.L. Sidner, “Attention, intentions, and the structure of discourse,” Computational Linguistics, vol.12, pp.175–204, 1986.

[32] R. Herman, “Phonetic markers of global discourse structures in English,” J. Phonetics, vol.28, pp.466–493, 2000.

[33] B.J. Grosz and J. Hirschberg, “Some intonational characteristics of discourse structure,” Proc. 2nd International Conference on Spoken Language Processing, pp.429–432, Banff, Canada, 1993.

[34] J. Hirschberg, “Communication and prosody: Functional aspects of prosody,” Speech Commun., vol.36, pp.31–43, Jan. 2002.

[35] J. Hirschberg and B. Grosz, “Intonational features of local and global discourse structure,” Proc. Speech and Natural Language Work-shop, pp.441–446, Harriman, New York, 1992.

[36] J. Hirschberg and C.H. Nakatani, “A prosodic analysis of discourse segments in direction-giving monologues,” Proc. 34th Annu. Meeting Assoc. for Comput. Linguist., Santa Cruz, pp.286–293, 1996.

[37] J. Hirschberg and J. Pierrehumbert, “The intonational structuring of discourse,” Proc. 24th annual meeting of the association for computational linguistics, pp.136–144, San Fransisco, CA, 1986.

[38] H. den Ouden, L. Noordman, and J. Terken, “Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports,” Speech Commun., vol.51, pp.116–129, Feb. 2009.

[39] X. Yang and Y. Yang, “Prosodic realization of rhetorical structure in Chinese discourse,” IEEE Trans. Audio Speech Language Process., vol.20, pp.1196–1206, 2012.

[40] L. Noordman, I. Dassen, M. Swerts, and J. Terken, “Prosodic markers of text structure,” in Amsterdam Studies in the Theory and History of Linguistic Science Series 4, pp.133–148.

[41] H. den Ouden, Prosodic realizations of text structure, Unpublished Ph.D. dissertation, University of Tilburg, Tilburg, The Netherlands, 2004.

[42] W.C. Mann and S.A. Thompson, “Rhetorical structure theory: Toward a functional theory of text organization,” Text, vol.8, pp.243–281, 1988.

[43] S. Nakajima and J.F. Allen, “A study on prosody and discourse structure in cooperative dialogues,” Phonetica, vol.50, pp.197–210, 1993.

[44] J. Krivokapic, “Prosodic planning: Effects of phrasal length and complexity on pause duration,” J. Phonetics, vol.35, pp.162–179, 2007.

[45] J. Cohen, “Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit,” Psychological Bulletin, vol.70, p.213, 1968.

[46] J.C. Becker and S.C. Wright, “Yet another dark side of chivalry: Benevolent sexism undermines and hostile sexism motivates collective action for social change,” J. Personality and Social Psychology, vol.101, pp.1–16, 2011.

[47] K.J. Preacher and A.F. Hayes, “Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple media-