Available online at www.sciencedirect.com

SciVerse ScienceDirect

Lingua 133 (2013) 101–126

Lingua

www.elsevier.com/locate/lingua

Prosodic correlates of discourse boundaries and hierarchy in discourse production

Joseph Tyler *

University of Michigan, 440 Lorch Hall, 611 Tappan Street, Ann Arbor, MI 48109-1220, USA

Received 3 August 2011; received in revised form 7 April 2013; accepted 10 April 2013
Available online

Abstract

A well-formed discourse is more than just a series of well-formed sentences. While often left implicit, this structure to discourse is sometimes overtly cued. And though most attention in this area has focused on lexicalized cues like discourse markers, prosody can also convey information about the structure of discourse. This paper presents the results of a production study examining prosodic correlates of discourse structure in readings of a newspaper article. Prosodic measures of pause duration, pitch, intensity and speech rate were found to significantly correlate with discourse structural measures of boundary size, discourse coordination/subordination, and their interaction. This interaction effect shows that the effect of boundary size on an utterance's prosody often depends on whether that utterance is coordinated or subordinated, and vice versa. These results expand our understanding of how prosody correlates with discourse structure, setting the stage for follow-up perception studies of what prosodic variation listeners use in discourse interpretation.

Keywords: Discourse; Prosody; Intonation; Coordination; Subordination; Segmented discourse representation theory

1. Introduction

Language is clearly structured in many different ways. Established areas of linguistics have for decades studied the systematic organization of sounds (phonology) and parts of a sentence (syntax). Similarly, the sentences of a discourse are structured, and a well-formed discourse is more than just a series of well-formed sentences. One way to reveal this structure is to remove it, perhaps by re-ordering the sentences of a discourse. For instance, if you were to read the sentences of this paragraph from last to first, the resulting discourse would be quite hard to follow. Even the two possible orderings of two sentences can lead to different interpretations of the events narrated.

(1) John banged his head. He fell over.
(2) John fell over. He banged his head.

A natural interpretation of the discourse in (1) is that John's banging his head happened before his falling over, while a natural interpretation of (2) is that John first fell over and then banged his head. In addition to the temporal ordering contrast, these two discourses likely also have different causal relationships. In (1), the banging of his head seems likely to have caused John to fall over. In (2), John's falling over seems likely to have led to him to bang his head.

* Present address: P.O. Box 2713, Department of English Literature and Linguistics, Qatar University, Doha, Qatar. Tel.: +974 734 761 5549; fax: +974 734 936 3406.
E-mail addresses: jctyler@umich.edu, jctyler5@gmail.com.

0024-3841/$ – see front matter
http://dx.doi.org/10.1016/j.lingua.2013.04.005

While it seems clear there is structure in discourse, it is less clear exactly what that structure is. Sometimes aspects of discourse structure are explicitly cued, while other times a speaker leaves the structure implicit, leaving listeners to fill in the gaps with their own reasoning. Most work that has analyzed explicit cues to discourse structure has focused on *lexical* cues, e.g. discourse markers. If (1) was instead produced as (3), with the addition of the explicit marker of temporal succession *then*, the temporal relationship between the two sentences would be explicit.

(3) John banged his head. Then he fell over.

In (3), it is explicit that John banged his head and subsequently fell over. An alternative, though dispreferred, interpretation of (1) could have been that it described two separate, independent events with no information about when each happened. In this interpretation, (1) would describe two independent events that happened to John, banging his head and falling over. With the addition of the discourse marker *then* in (3), the temporal ordering is explicitly encoded and this alternative is ruled out. Thus, the addition of a lexical item like a discourse marker can make explicit how the sentences of a discourse are related.

One feature of discourse identified by many theorists (Grosz and Sidner, 1986; Hobbs, 1985; Mann and Thompson, 1988; Polanyi, 1988; Van Kuppevelt, 1995) is that it is hierarchically structured. Asher and Vieu (2005) discuss the intuitions motivating hierarchical structure in the context of Segmented Discourse Representation Theory (SDRT) (Asher and Lascarides, 2003). They mention paragraph structure as an orthographic manifestation of discourse hierarchy, where paragraph-initial sentences are in some sense higher-order than paragraph-medial sentences. A paragraph-medial sentence likely provides more detail about whatever was introduced by the paragraph-initial sentence. They also argue that temporal structure motivates a hierarchical conception of discourse. If one sentence introduces an event and a second sentence describes something occurring at the same time as that first event, the second is likely providing more detail about the first event. By contrast, if a second sentence describes an event at a different time, the two events likely have equal status.

Like most theories of discourse structure, SDRT analyzes the structure of discourse by segmenting the discourse, identifying relations that hold between segments, and constructing a hierarchy from the segments and relations. SDRT focuses on both semantic and pragmatic information for all stages of analysis (segmentation, relation identification, hierarchy). SDRT also provides an inventory of discourse relations (e.g. Elaboration, Background, Result) that are claimed to hold between the segments of a discourse. But most importantly here, SDRT builds hierarchy in discourse by classifying all discourse relations as either coordinating or subordinating. Coordinating relations link discourse segments at an equal hierarchical level while subordinating relations link a discourse segment with another segment one hierarchical level lower.

Rhetorical Structure Theory (RST) (Mann and Thompson, 1988), like SDRT, analyzes a discourse into segments, identifies relations between segments, and constructs the discourse into a hierarchical structure. RST also has a local hierarchical structure contrast in its nucleus-satellite distinction. In RST, all discourse segments are considered to be either a nucleus or a satellite. The distinction between the two is defined in terms of a segment's relative importance to the coherence of the discourse. One diagnostic test is that satellites can be deleted without harming the overall message of the discourse, while deleting a nucleus would disrupt the discourse's coherence. This test reveals one of RST's applications: automatic text summarization. If all satellites in a text were deleted, the result would be a stripped down summary of the discourse.

While RST's nuclearity principle has been compared to SDRT's coordinating/subordinating contrast (Danlos, 2010), there are points of contrast. In RST, nuclearity is a feature of a discourse segment. This means that every discourse segment is either a nucleus or a satellite. In SDRT, coordinating and subordinating relations are theorized to hold between discourse segments, but are not strictly features of the segments themselves. This means that any one segment in an SDRT analysis could be coordinated to one segment and subordinated to another. Another difference between RST's nuclearity and SDRT's coordinating/subordinating contrast is in terms of how an analyst identifies a segment's nuclearity or CoordSubord status. In RST, a central criterion for satellite status is that a discourse segment be expendable: if it can be deleted without harming the discourse's coherence, it is a satellite. In SDRT, the main point of contrast between coordination and subordination is in terms of the level of detail. So, RST and SDRT both supply theoretical constructs that account for local hierarchical contrasts, but the nature of those local hierarchical constructs is not exactly the same.

Another influential theory of discourse that analyzes discourse into segments, relations between segments, and hierarchy is the Grosz and Sidner model (1986). Unlike SDRT and RST, which focus on the propositional content of utterances as the basis of their analyses, the Grosz & Sidner model analyzes discourse using speaker purposes, goals and intentions. In this theory, a speaker may have one overall purpose to their discourse, e.g. to give directions on how to replace a car battery. Then, this overall purpose may be subdivided into a series of subgoals, e.g. how to identify the battery, how to remove the old battery, and how to install the new battery. Grosz & Sidner propose two structural relations that organize these discourse purposes into a hierarchical structure: dominance and satisfaction-precedence. The higher-order purpose of replacing a car battery is said to dominate the three subgoals. And since the removal of the old battery needs to be complete before the

installation of the new battery begins, the purpose of the battery removal portion of the discourse is said to satisfaction-precede the purpose of the battery installation portion of the discourse. These two relations (dominance and satisfaction-precedence) therefore create contrasting hierarchical structure. Dominance relations link segments at different hierarchical levels while satisfaction-precedence relations link segments at the same hierarchical level.

The Grosz & Sidner model, RST and SDRT are all capturing ways in which discourse is segmented, how the segments are related, and how the whole is hierarchically structured. The more cues we can draw on in the speech signal, the better we can understand what that structure is and how speakers and listeners communicate it to each other. When there are no overt cues to discourse structure, listeners must draw on more general reasoning about how the sentences are likely to fit together. This was the case with (1) above, where a plausible interpretation involves the banging of his head causing John's falling over, even though this causal information was not explicitly asserted.

And while most work on cues to discourse structure has focused on lexical cues, there is a body of research that has also identified systematic correlates between aspects of discourses' structure and prosodic measures of pitch, pause duration, intensity and speech rate (den Ouden et al., 2009; Hirschberg and Grosz, 1992; Lehiste, 1975). This indicates that the prosody of speech can carry cues to the structure of discourse. One common feature of discourse with which prosody is correlated in this work is the size of a discourse boundary. This work uses diverse criteria to identify boundaries of different sizes. These criteria include orthographic markers of paragraph boundaries (Lehiste, 1975, 1982) and intuitive analyses of breaks in the discourse, either by the experimenter (Yule, 1980) or the participants themselves (Swerts, 1997). Other work creates measures of boundary size using a specific theory of discourse structure, e.g. den Ouden et al. (2009) use Rhetorical Structure Theory (Mann and Thompson, 1988) and Hirschberg and Grosz (1992) use the Grosz & Sidner model (Grosz and Sidner, 1986). These studies use different terms to describe similar phenomena, but for consistency I will use the term boundary size. Compared with boundary size, less is known about prosodic correlates of local hierarchical relationships in discourse like coordination and subordination, though see the results for prosodic correlates of nuclearity in den Ouden et al. (2009). And very little is known about their interaction, i.e. how the effects of boundary size and coordination/subordination may depend on each other.

The prosodic measures most commonly found to correlate with discourse have been pause duration and pitch maxima, though others have been explored as well. Pause durations have tended to be longer at larger discourse boundaries (den Ouden et al., 2009; Lehiste, 1982). Pitch maxima, characterized variously as pitch range (Hirschberg and Grosz, 1992), pitch reset (Auran and Hirst, 2004), and high onset pitch (Couper-Kuhlen, 2001; Yule, 1980), tend to be higher following larger boundaries in the discourse. den Ouden et al. (2009) found a correlation between the nucleus/satellite contrast and articulation rate, but no correlation with pause duration or maximum pitch. And while they have gotten less attention, discourse has been found to correlate with other prosodic measures like amplitude (Herman, 2000; Hirschberg and Grosz, 1992) and rhythm (Müller, 1996).

This paper presents the results of a discourse prosody production study testing for prosodic correlates of boundary size and coordination/subordination, as well as their interaction. In addition to some of the more traditional prosodic measures, it includes measures of how far through a discourse segment pitch and intensity maxima occur. These measures can help illuminate if discourse structure also correlates with pitch or intensity peaks along a temporal dimension. This study also includes the results of two correlation analyses, one correlating the predictor variables and the other correlating the prosodic measures. These correlation analyses reveal how independent the variables are, informing which ones to exclude and how to interpret those that remain.

The spoken data were elicited by having participants read aloud a newspaper article, and then correlating prosodic features of those productions with features of the article's discourse structure. As discussed by Smith (2004:249), a benefit of using read speech instead of spontaneous speech is that the discourse annotation is not based on prosodic information (for more discussion of this circularity concern, see Swerts, 1997). But because read speech sometimes differs from spontaneous speech (Laan, 1997), further research would be needed to see if results from this study extend to more spontaneous forms of speech production.

Results from this study can inform how prosody correlates with discourse structure in speech production and set the stage for follow-up perception studies. The results could also inform and improve the development of speech synthesis and recognition systems by better accounting for the prosodic variation those systems need to take into account.

1.1. Discourse structure

The discourse structure variables in this study are derived from a discourse representation constructed using *Segmented Discourse Representation Theory* (SDRT) (Asher and Lascarides, 2003). As mentioned above, SDRT identifies the structure of a discourse by dividing it into segments, inferring rhetorical relations that hold between those segments (e.g. ELABORATION) and assembling them into a hierarchical structure. To set the stage for the explanation of how these variables were created, it will be helpful first to exemplify how SDRT performs each of these processes. We can work through these processes with the excerpt in (4), drawn from the newspaper article used in this study. The article and

all SDRT annotations come from the DISCOR annotated corpus (Reese et al., 2007), a research project that used SDRT to determine the discourse structure of natural language texts and identify dependencies between anaphors and their antecedents. The goal of the current study is not to test the value of an SDRT representation against other theories' representations, but to take the SDRT analysis as a good discourse representation from which to test relationships between prosody and discourse structure.

(4) The White House will try to assuage at least some opponents' concerns as Congress undertakes to reconcile the Senate bill with a much different House measure. Justice Department officials, who were criticized for not visibly exerting influence over the Senate bill last year, will play a more overt role in removing or modifying the more extreme provisions this year. Deputy Attorney General Philip Heymann plans to testify at House crime legislation hearings, and Mr. Clinton himself held out the carrot of help to endangered youth in his speech to Congress.

The first step in analyzing the discourse structure of (4) is segmentation. Sentence boundaries were all treated as segment boundaries, and sub-sentential portions were treated as discourse segments if they served "a discernible discourse function" (Reese et al., 2007:3).[1] For ease of representation, the resulting segments are each assigned a number corresponding to their sequential position in the text:

(5) [27 The White House will try to assuage at least some opponents' concerns] [28 as Congress undertakes to reconcile the Senate bill with a much different House measure.] [29 Justice Department officials, who were criticized for not visibly exerting influence over the Senate bill last year, will play a more overt role in removing or modifying the more extreme provisions this year.] [30 Deputy Attorney General Philip Heymann plans to testify at House crime legislation hearings,] [31 and Mr. Clinton himself held out the carrot of help to endangered youth in his speech to Congress.]

The brackets indicate segment boundaries and the numbers are a shorthand way to refer to the discourse's segments.

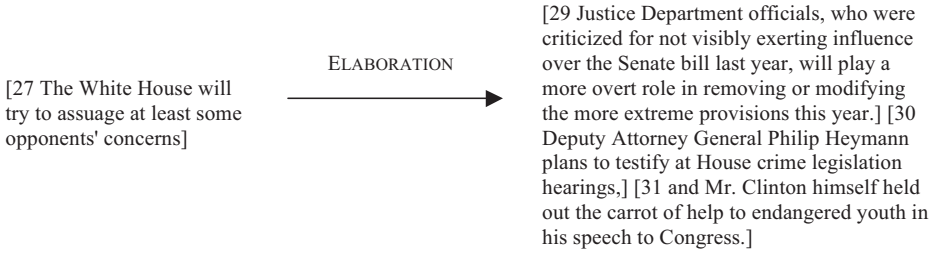

Fig. 1. Schematic representation of an ELABORATION relation.

After segmentation, rhetorical relations are identified that are inferred to hold between those segments. Written as RELATION(ARG1,ARG2), a relation is said to hold between its two arguments. The DISCOR annotators identified the following relations in the excerpt:

(6) ELABORATION(27,[29,30,31])
BACKGROUND(27,28)
CONTINUATION(29,30)
CONTINUATION(30,31)

Here, ELABORATION(27,[29,30,31]) captures a rhetorical relation of elaboration where the text corresponding to segment 27 is elaborated by the text corresponding to segments 29, 30 and 31 (Fig. 1).

The DISCOR annotation manual explains that "ELABORATION(α,β) holds when β provides further information about the eventuality introduced in α" (Reese et al., 2007:7). In this example, the first argument of the elaboration relation introduces the proposition of the White House assuaging opponents' concerns, about which the second argument provides further information in the form of the Justice Department's more overt role (segment 29), Heymann's testimony (segment 30) and

[1] Full details of the SDRT annotations are available in the DISCOR annotation manual (Reese et al., 2007).

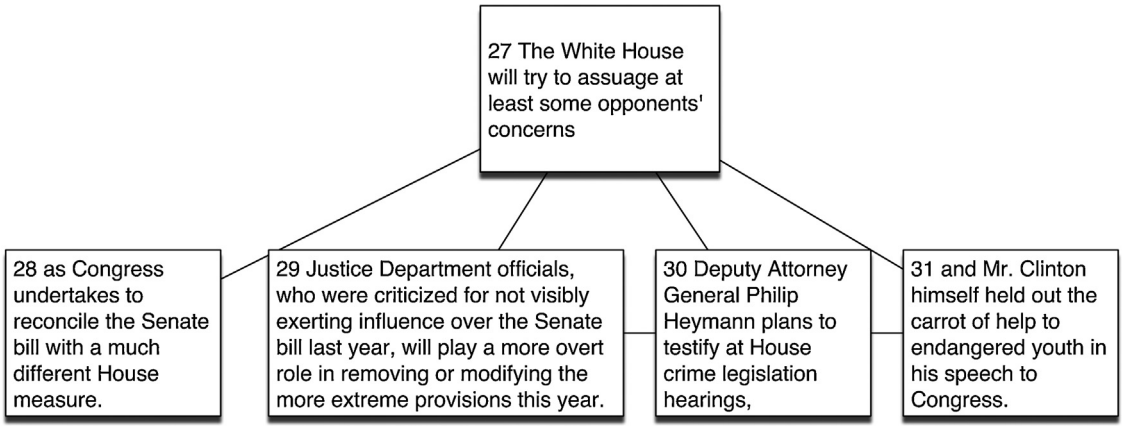

Fig. 2. Graphical representation of SDRT analysis of segments 27–31.

Clinton's reaching out (segment 31). In this example, the elaboration relation's second argument is composed of three discourse segments while the other argument is composed of a single discourse segment. SDRT arguments can be simple (made up of a single segment) or complex (made up of multiple segments). Along the same lines, the relations BACKGROUND (27,28), CONTINUATION(29,30) and CONTINUATION(30,31) indicate one background and two continuation relations between their first and second arguments, respectively. A discourse is said to be coherent if rhetorical relations connect all of its segments.

In SDRT, hierarchical structure is captured by categorizing all relations as either coordinating or subordinating. The information captured in the coordinating/subordinating contrast is the 'granularity' or level of detail being given in the discourse'' (Asher and Lascarides, 2003:8). The second argument of a subordinating relation provides more detail than the first argument, while the second argument of a coordinating relation provides a similar level of detail as the first argument. The relation ELABORATION, for example, is a subordinating relation, meaning the second argument provides more detailed information than the first, and as a result is a level below the first. CONTINUATION is a coordinating relation, meaning the second argument provides a similar level of detail as the first and are at the same level. This hierarchical view contrasts with conceptions of discourse as a set of propositions or possible worlds, as well as with ''the dynamic semantic view of text information as a sequence of information updates'' (Asher and Vieu, 2005:591). SDRT's hierarchical structure also achieves empirical gains in its ability to account for phenomena like anaphoric reference and temporal structure in ways a non-hierarchical theory cannot (Asher and Lascarides, 2003). The SDRT annotations in the DISCOR corpus are the result of three annotators identifying the segments of the discourses and the relations that hold between them. Final annotations were decided by consensus between annotators.

In the graphical representation in Fig. 2, vertical lines are used to indicate subordinating relations and horizontal lines to indicate coordinating relations. The arguments of those relations are represented as the boxes at the end of the lines.

This graph shows segments 29, 30 and 31 subordinated to, i.e. a level below, segment 27 while coordinated to each other. In terms of propositional content, segment 27 introduces the proposition of the White House assuaging opponents' concerns, and segments 29–31 elaborate on that proposition. Segment 28 gives background information about 27.

If a similar analysis is applied to the whole article, we get the graph in Fig. 3.

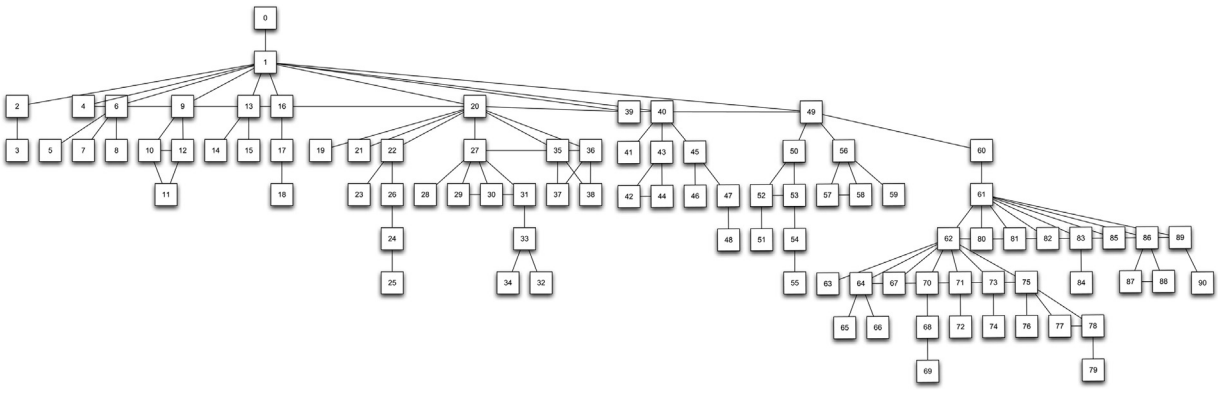

Fig. 3. Graphical representation of SDRT analysis of entire article.

The graph in Fig. 3 captures the segmentation, discourse relation and hierarchical organization information that constitutes an SDRT discourse structure representation.

2. Methods

The production data in this study were elicited by having participants read, analyze and then read aloud a newspaper article. The article's structure, modeled with SDRT, was converted into two predictor variables corresponding to boundary size (Bsize) and the contrast between coordinating and subordinating relations (CoordSubord). Prosodic measures were taken from the readings and tested for correlations with those predictor variables.

2.1. Participants

Ten students from the University of Michigan participated in this study in exchange for ten dollars. Eight speakers were female and two were male. All reported English as their native language and English as their major field of study. English majors were chosen because they were expected to be particularly capable of identifying the discourse structure of a news article, a necessary first step to test the larger question of how prosody correlates with discourse structure. With this population, non-significant results are less likely to be due to speakers not identifying the discourse structure and more likely to be due to how prosody correlates with discourse structure. And because the goal of this study is to gain the greatest insight into how prosody correlates with discourse, not to identify the average person's discourse prosody, this non-random selection best fulfilled this goal.

2.2. Materials

Participants were asked to read aloud the 1994 Wall Street Journal newspaper article titled *Blacks' increasing vocal opposition to violence is matched by strong opposition to crime bill* (Davidson, 1994).[2] The article comes from the DISCOR corpus of news articles annotated within SDRT (Reese et al., 2007). The article addresses new crime legislation proposed during Bill Clinton's presidency, the reaction to it among black leaders and the various political factions involved. This article was chosen because it was sufficiently long and diverse in terms of features necessary to test the research questions. Having all participants read the same article, instead of each one reading a different article, meant that variability between speakers could not be due to idiosyncratic differences between texts. This study's discourse structure variables were derived from the article's discourse structure as characterized in the SDRT annotations from DISCOR.

2.2.1. Design

This study's overarching goal is to determine how information about a discourse segment's position in the discourse can help predict that segment's prosody. It will be tested by asking whether speakers indicate with their prosody (a) the size of a boundary between discourse segments (Bsize), (b) whether a segment is coordinated or subordinated to the most recent segment to which it is attached (CoordSubord), and (c) whether the effect of either (a) or (b) is mitigated by the other. These sub-questions will be addressed using predictor variables for Bsize and CoordSubord that were converted from the above SDRT representation and then testing them for significant correlations with prosodic measures.

2.2.2. Boundary size (Bsize)

The Bsize variable captures the amount of structure intervening between sequential segments of a discourse, e.g. segments 47 and 48. The Bsize variable's values are calculated as the number of nodes in the discourse structure intervening between two sequential segments. In practice, this involves identifying the shortest path between two segments and counting how many other segments must be traveled through to reach the next one.

The excerpt in (7) below can help exemplify how the values for Bsize were calculated.

(7) Excerpt from news article used in study, with discourse segments numbered in sequential order

40. But the mainstream civil-rights leadership generally avoided the rhetoric of ''law and order,''
41. regarding it as a code for keeping blacks back.
42. Law and order didn't mean justice,
43. Mr. Jackson used to say,

[2] The full text of the article as presented to participants is in Appendix A.

44. but ‘‘just us.’’
45. In the past, many were hesitant to speak about crime in public
46. because ‘‘the larger community would talk about ‘lock them up and throw the key away’ and hide behind black leaders in doing it,’’
47. explains Rep. Craig Washington,
48. the Houston Democrat who led the caucus hearing.
49. Now there is escalating discourse within the black community about what it can and must do to stop crime.
50. Just after the new year, Mr. Jackson held the first of several conferences focusing on just that.
51. ‘‘The premier civil-rights issue of this day is youth violence in general and black-on-black violence in particular,’’
52. he has said.

The text in (7) has a discourse structure of the form shown in Fig. 4. In Fig. 4, the pair 47–48 has no intervening segments, and so has a boundary size of 0. By contrast, the pair 48–49 has three intervening nodes (segments 47, 45 and 40) and so has a boundary size of 3.

Boundary size was calculated in the same way for all sequential pairs, resulting in the following distribution (Table 1).

As the table makes clear, there are decreasing numbers of segments as boundary size increases. While a more even distribution would be preferable for the statistical analysis, it is the nature of this discourse, and perhaps all discourses, that boundaries between adjacent segments are more often small than large. In addition, the analysis requires enough observations at each level for the statistical model to be able to compare them. Because there was only one level 4 boundary, level 4 was merged with level 3. Combining the data from levels 3 and 4 led to sufficient data at every level for the statistical model to have the power necessary to test for Bsize as a predictor of prosody.

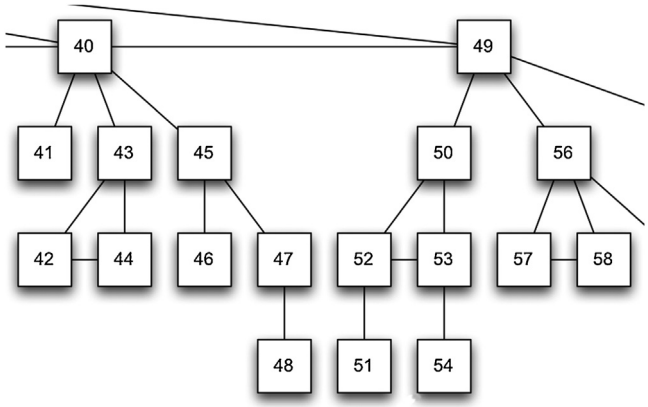

Fig. 4. Graphical representation of SDRT analysis of segments 40–58.

Table 1
Distribution of discourse segment frequencies and average number of words at each level of boundary size.

Boundary Size	Level 0	Level 1	Level 2	Level 3	Level 4
Discourse segments ($n = 90$)	54 (60%)	24 (27%)	7 (8%)	4 (4%)	1 (1%)
Average # words	11.47	16.25	15.86	21.25	4

2.2.3. Coordinating vs. subordinating relations (CoordSubord)

The CoordSubord variable is designed to test whether local hierarchical relationships between discourse segments result in different prosody in speech production. The variable captures three different ways a discourse segment could be connected to the larger discourse structure. The Coord part of CoordSubord is named for discourse segments whose most recent attachment is via a coordinating relation, e.g. segment 58 in the graph in Fig. 4. The Subord part of CoordSubord is named for discourse segments whose most recent connection is via a subordinating relation, e.g. segment 57 in the graph in Fig. 4. The variable only considers the most recent discourse relation for each discourse segment, regardless of what connections that discourse segment may have to earlier relations in the discourse. The decision to code based on the most recent relation was due to the overall distribution of relations. All discourse segments except the first were subordinated to at least one other segment, while only 27 segments were coordinated. And all of

Table 2
The set of acoustic features extracted from each discourse segment. For the pitch and intensity peak location measures, the unit captures what proportion through a discourse segment the peak occurred.

Acoustic feature	Units	Description
Pause duration	ms	Duration of silence preceding discourse segment
Pitch maximum (f0max)	Hz	Maximum F0 across entire discourse segment
Pitch minimum (f0min)	Hz	Minimum F0 across entire discourse segment
Mean pitch	Hz	Mean F0 across entire discourse segment
Initial pitch	Hz	Mean F0 for initial 5% of discourse segment
Pitch peak location	0–1	How far through the discourse segment the highest pitch point occurs
Max intensity	dB	Maximum decibel level in discourse segment
Mean intensity	dB	Mean decibel level in discourse segment
Min intensity	dB	Minimum decibel level in discourse segment
Intensity peak location	0–1	How far through the discourse segment the highest intensity point occurs
Speech rate	Words/duration	Words divided by duration of discourse segment

those coordinated segments were more recently coordinated than subordinated. Therefore, the contrast between coordination and subordination appears in terms of which relation is most recent.

There is a third group of discourse segments that are not connected to any earlier discourse segments but only to those that come later in the discourse. These discourse segments are produced before it is clear to which segments they are rhetorically related. Of the 90 total discourse segments in the text, 56 were coded as subordinated, 27 as coordinated, and 7 were related to an upcoming segment. The data for the third level of CoordSubord were excluded from the analyses; this was done to better isolate the purpose of the variable, namely to compare coordinated vs. subordinated discourse segments. And because this resulted in a total of 83 out of 90 segments remaining, there was still enough data to address the question about effects of coordination vs. subordination on prosody.

2.2.4. Prosodic measures

This study focused on a broad set of prosodic measures, including pause duration, pitch, intensity, and speech rate (Table 2).

The measures were extracted automatically using a script in the acoustic analysis software program Praat (Boersma and Weenink, 2009). For women, the pitch window was set at 100–500 Hz, while for men the window was 75–300 Hz. For the automatic measurement, it was necessary to adjust Praat's default pitch settings to be more conservative about what it accepts as pitch in order to reduce errors. For f0max, it sufficed to raise the voicing threshold from the default 0.45 to 0.6, as performed elsewhere (Ljolje, 2002). The resulting output was reliable enough that the few remaining errors could be spotted and fixed by hand. The f0min measurements, however, required different and more conservative pitch settings. F0min can be chaotic to measure because speakers sometimes enter creaky voice as they descend in pitch. When using the f0max pitch settings, f0min had a binomial distribution, with a cluster of measurements around 100 Hz all in creaky voice. In order to filter out these creaky voice measurements, the Praat settings were made more conservative by increasing the voicing threshold, the octave cost and the voicing/voiceless cost,[3] resulting in a more normal distribution. Remaining outliers were checked individually.

The measures for pitch peak location and intensity peak location are measured as how far through a discourse segment the peak is produced. If the peak occurred at the very beginning, the measure would equal zero; if the peak occurred at the very end, the measure would equal 1. And if the peak occurred 20% of the way through the discourse segment, the measure would equal 0.2. The goal of these peak location measures was to explore variation along the temporal dimension, i.e. where in the segment prosodic phenomena occurred. Given that high onset pitch has been found to occur after large boundaries in discourse (Auran and Hirst, 2004; Wichmann, 2000), pitch peaks were expected to occur earlier following larger discourse boundaries and on coordinated discourse segments. Though little is known about its behavior, intensity peak location was included to be able to compare results for pitch and intensity.

Excluded from the analysis were discourse segments with disfluent speech production. Previous research has described disfluencies as ''fillers like uh or um, unfilled pauses, repeated words, repaired words, or even disfluent-sounding prosody'' (Arnold, 2008:508). Because the focus of this study is on prosodic production, this study treats lexically anomalous production as disfluent, but not ''disfluent-sounding prosody'' like intuitively unexpected lengthening or

[3] For details, see Appendix C.

awkward pauses. Disfluency was defined as when a speaker repeated a word, said a word that was not present in the text, did not utter a word that was present, or had some extra-verbal interruption like coughing or sneezing. 153 out of 910 total segments (17%) were excluded from the analysis due to disfluency.

Each speaker read the text of the newspaper article aloud twice. For nine of the ten speakers, prosody measures were taken from their second reading. For the tenth speaker, the first reading was used. The second reading was chosen for most speakers because it was thought they would have gotten more familiar with the text and the task, and so have produced more fluent speech. The one speaker whose first reading was used was flustered when asked to read it again. I analyzed her first reading because on the first reading she appeared less distracted from the task.

2.2.5. Procedure

The recordings were done in the noise-controlled sound lab at the University of Michigan, using Praat and an AKG C 4000 B microphone. Participants first read the article silently to themselves, then paraphrased it out loud, and finally read the entire article out loud twice. Participants were directed to read the text aloud in such a way as to most clearly communicate the article to a listener. All of the article's paragraphing was removed, but sentence-level punctuation was left in. There were no subheadings. As a result, participants had no information about paragraph structure, and so overt paragraphs themselves could not account for prosodic variation.

The motivation for having participants paraphrase the article out loud before reading it aloud was both to get participants to think about the text in a structural way and to have a record of what they saw as the text's most important points. The paraphrases provide a check on whether the SDRT representations are capturing features of the text that participants found important. A comparison between the paraphrases and SDRT appears in the Discussion section. Participants were encouraged to study the article for as long as necessary prior to reading aloud in order to understand its structure as well as possible.

3. Results

Potential effects of discourse structure on prosody were tested by fitting a linear mixed model to the data. Each model used contained the predictor variable(s) of interest (Bsize, CoordSubord), control variables, and the dependent measures of prosody. The control variables were included to help rule out explanations other than discourse structure for the prosodic variation. One potential confound may be that speakers change their prosody over the 5–10 min participants took to read the text, perhaps due to factors like fatigue or wandering attention. A variable was included in the model that indicated how far along in the discourse the segment was uttered to try to control for these potential location effects (*Number*, for discourse segment number). Another confound could be whether material in the text was in quotes or not. A variable was added that indicated whether the discourse segment was wholly, partially, or not at all in quotes (*Quot*). Of the text's 90 discourse segments, 71 had no quoted material, 9 had some, and 11 had all quoted material. Additionally, some discourse segments began sentences while others began in the middle of sentences; a sentence-initiality variable was added to capture this information (*Sentinit*). Of the text's 90 discourse segments, 46 are sentence-initial and 44 are non-initial. And finally, the length of a discourse segment may affect how extreme a prosodic measure becomes; to capture this information, one variable was included that indicated the number of words in the discourse segment (*Words*) and another that indicated the duration of the segment in seconds (*Duration*).

Before analyzing the results of discourse structure predicting prosody, it will be useful to analyze correlations among the predictor variables (predictors of interest as well as controls) and then analyze correlations among the prosodic outcomes. The correlation analyses can show which variables pattern together, revealing if some are redundant and can be excluded. The correlations can also help in the interpretation of the variables that remain, by showing how independent each variable is from the others.

In Table 3, all predictor variables are laid out in a matrix of Pearson correlations. The higher the correlation values, the more closely the variables pattern together. As the value gets closer to zero, the variables are more independent from each other. And as the correlation gets closer to negative one, the more the variables pattern in opposite directions.

It is clear the variables *Words* and *Duration* are highly correlated with each other (0.904) and independent of the other variables. These two variables seem to be capturing the same information. This makes sense because the number of words in a segment is likely to affect how long it takes to say that segment. Because of the high correlation, one of either *Words* or *Duration* should be removed. And because *Duration* is less correlated with *Bsize* than words, *Duration* will be left in the model and *Words* will be removed. A weaker correlation shows up between *Bsize* and *Sentinit* (0.471), indicating that segments after larger boundaries are more likely to be sentence-initial. *CoordSubord* is largely independent of the other variables, and notably has almost no correlation with *Bsize*.

Correlation results for the prosodic measures are laid out in Table 4. The largest cluster of high correlation values are among the measures for maximum, minimum, mean and initial pitch. Because they pattern together, only one measure is needed to capture this effect. While any of the correlated pitch measures could be used, pitch maximum was retained and

Table 3
Correlation matrix for predictor and control variables.

	Sentence-initiality	Boundary size	Coord/Subord	Quoted	Discourse segment number	Words	Segment duration
Sentence-initiality	1	0.471	−0.086	0.011	−0.038	0.263	0.248
Boundary size	0.471	1	0.022	0.017	−0.042	0.261	0.194
Coord/Subord	−0.086	0.022	1	0.085	−0.073	−0.03	−0.009
Quoted	0.011	0.017	0.085	1	0.005	0.018	−0.02
Discourse segment number	−0.038	−0.042	−0.073	0.005	1	0.022	0.097
Words	0.263	0.261	−0.03	0.018	0.022	1	0.904
Segment duration	0.248	0.194	−0.009	−0.02	0.097	0.904	1

Table 4
Correlation matrix for prosodic measures.

	Pause duration	Pitch maximum	Pitch minimum	Mean pitch	Initial pitch	Pitch peak location	Maximum intensity	Mean intensity	Minimum intensity	Intensity peak location	Speech rate (words/s)
Pause duration	1	0.212	−0.111	0.05	0.2	−0.262	0.262	0.099	−0.177	0.258	−0.052
Pitch maximum	0.212	1	0.654	0.88	0.846	−0.134	0.276	0.2	0.187	0.233	−0.082
Pitch minimum	−0.111	0.654	1	0.876	0.675	−0.003	0.047	0.146	0.383	0.09	−0.063
Mean pitch	0.05	0.88	0.876	1	0.843	−0.071	0.181	0.226	0.34	0.16	−0.072
Initial pitch	0.2	0.846	0.675	0.843	1	−0.191	0.212	0.167	0.26	0.232	−0.07
Pitch peak location	−0.262	−0.134	−0.003	−0.071	−0.191	1	−0.216	−0.163	0.042	−0.211	−0.102
Maximum intensity	0.262	0.276	0.047	0.181	0.212	−0.216	1	0.707	−0.105	0.116	−0.116
Mean intensity	0.099	0.2	0.146	0.226	0.167	−0.163	0.707	1	0.317	0.082	0.188
Minimum intensity	−0.177	0.187	0.383	0.34	0.26	0.042	−0.105	0.317	1	0.051	0.309
Intensity peak location	0.258	0.233	0.09	0.16	0.232	−0.211	0.116	0.082	0.051	1	0.143
Speech rate (words/s)	−0.052	−0.082	−0.063	−0.072	−0.07	−0.102	−0.116	0.188	0.309	0.143	1

pitch mean, pitch minimum and initial pitch were excluded. This decision was made because pitch maximum is a common measure in other discourse prosody research (den Ouden et al., 2009; Hirschberg and Grosz, 1992). Also, maximum and minimum intensity are both correlated with mean intensity, though not with each other. Therefore maximum and minimum intensity were retained but mean intensity was removed. The remaining measures are largely uncorrelated.

The final set of predictor, control and dependent variables are listed in Table 5. There is one control variable in this table that was not included in the correlation analyses. This variable captures the prosody in the previous discourse segment, testing how much a prosodic measure's value in a current discourse segment is predictable from that same measure in the prior segment. For example, saying the previous segment loudly might lead to the subsequent segment being louder. As a result, a speaker's maximum intensity in one segment may be highly related to their maximum intensity in the prior segment. This *ProsPrev* variable can help separate variation in a prosodic measure that is due to the previous segment's prosody and not to the discourse structure. The *ProsPrev* variable was not included in the correlation analysis because it varies from dependent variable to dependent variable.

The ability of each predictor variable to predict each dependent variable was tested for significance with a linear mixed model (LMM). Because each subject is providing many data points, the observations are not fully independent, an

Table 5
Full list predictor, control and dependent variables used in the linear mixed model.

Predictor and control variables	Dependent variables (prosody)
Boundary size (Bsize)	Pause duration
Coordination vs. subordination (CoordSubord)	Pitch maximum (f0max)
Sentence-initiality (Sentinit)	Pitch peak location
Segment duration (Duration)	Maximum intensity
Quoted material (Quot)	Minimum intensity
Discourse segment number (Number), i.e. how far through the discourse the segment occurs.	Intensity peak location
Previous segment's value for same prosodic measure (ProsPrev).	Speech rate

assumption in statistical models like ANOVA. We may be better able to predict a prosodic outcome by taking into account who produced it. We can take these subject effects into account by including a random intercept for subjects in a linear mixed model.

Linear mixed models offer a range of benefits over other repeated measures models like repeated measures ANOVA (Quené and van den Bergh, 2004, 2008). Quené & van den Bergh have run two studies demonstrating the benefits of mixed models over ANOVA, first with normally distributed data (2004) and then with binary data (2008). In both cases, mixed models serve as better fits of the data. Mixed models benefit from being able to accommodate missing data and unequal cell sizes, two concerns in this data set. Another benefit is the ability to avoid making false assumptions about the independence of the observations by taking into account repeated measures. Mixed models also afford higher statistical power and so are more able to accurately identify effects in the data. All statistical modeling was performed with SPSS.

3.1. Boundary size

To identify overall patterns of boundary size on prosodic outcomes, a linear mixed model was fitted that contained Bsize but not CoordSubord. This model tells us what effect a change in boundary size has on each prosodic outcome. Boundary size was entered as a continuous variable. In Table 6, we see the results for each prosodic measure by row. The intercept indicates the model's predicted value for each prosodic measure when boundary size is 0. The coefficient indicates the model's predicted slope, i.e. the amount of change in that prosodic measure for every level increase in boundary size.

Results for boundary size indicate that pause duration, max pitch, max intensity, and speech rate all increase as a discourse segment's preceding boundary increases in size. Furthermore, a discourse segment's intensity peak occurs later in the segment as the preceding boundary gets larger. These effects are all highly significant ($p < 0.01$). The graphs in Fig. 5 plot each prosodic measure on the *y*-axis for each level of boundary size on the *x*-axis. All prosodic measures other than speech rate show an increase from level 0 to 1 to 2, with a plateau between 2 and 3. Speech rate drops from level 0 to 1 before rising to 2 and 3. Results for the control variables are presented in Table 7. There are many significant effects of the control variables on the prosodic outcomes, demonstrating the importance of including them in the model.

Sentence-initiality (*Sentinit*) was a strong predictor of all measures except intensity peak location and speech rate. So, while a segment's preceding pause duration was dramatically affected by whether that segment was sentence-initial or not, that segment's speech rate was not.

The position of the discourse segment in the overall discourse (*Number*) was a significant predictor for max pitch, max intensity and speech rate, but not for pause duration, pitch peak location or intensity peak location. This suggests that over time speakers changed how high, loud and fast they spoke, but did not change pause durations or the relative position of the pitch and intensity extremes. Whether a discourse segment contained quoted material predicted maximum intensity and speech rate.

A discourse segment's duration in seconds (*Duration*) was a significant predictor of all prosodic measures except the relative pitch and intensity peak measures. This suggests that where in a discourse segment a pitch or intensity peak occurs is not dependent on how long it takes to say the segment.

And finally, the prosody of the previous segment (*ProsPrev*) significantly predicted pause duration, pitch peak location, max and min intensity, and speech rate. It seems for these measures, speakers may get into periods of using the prosody

Table 6
Results for boundary size as a predictor of prosody, collapsing across CoordSubord. The intercept indicates the model's predicted value for each prosodic measure when boundary size is 0. The coefficient indicates the model's predicted slope, i.e. the amount of change in that prosodic measure for every level increase in boundary size.

Boundary size	Intercept	Coefficient	*F*-statistic	*p*-Value
Pause	665.659	78.490	25.806	0.000**
Maximum pitch	215.205	4.392	11.449	0.001**
Pitch peak location	0.176	−0.022	2.432	0.119
Maximum intensity	70.820	0.612	18.243	0.000**
Minimum intensity	33.139	−0.080	0.140	0.709
Intensity peak location	0.722	0.052	17.313	0.000**
Speech rate	3.948	0.124	10.734	0.001**

* $p < 0.05$
** $p < 0.01$.

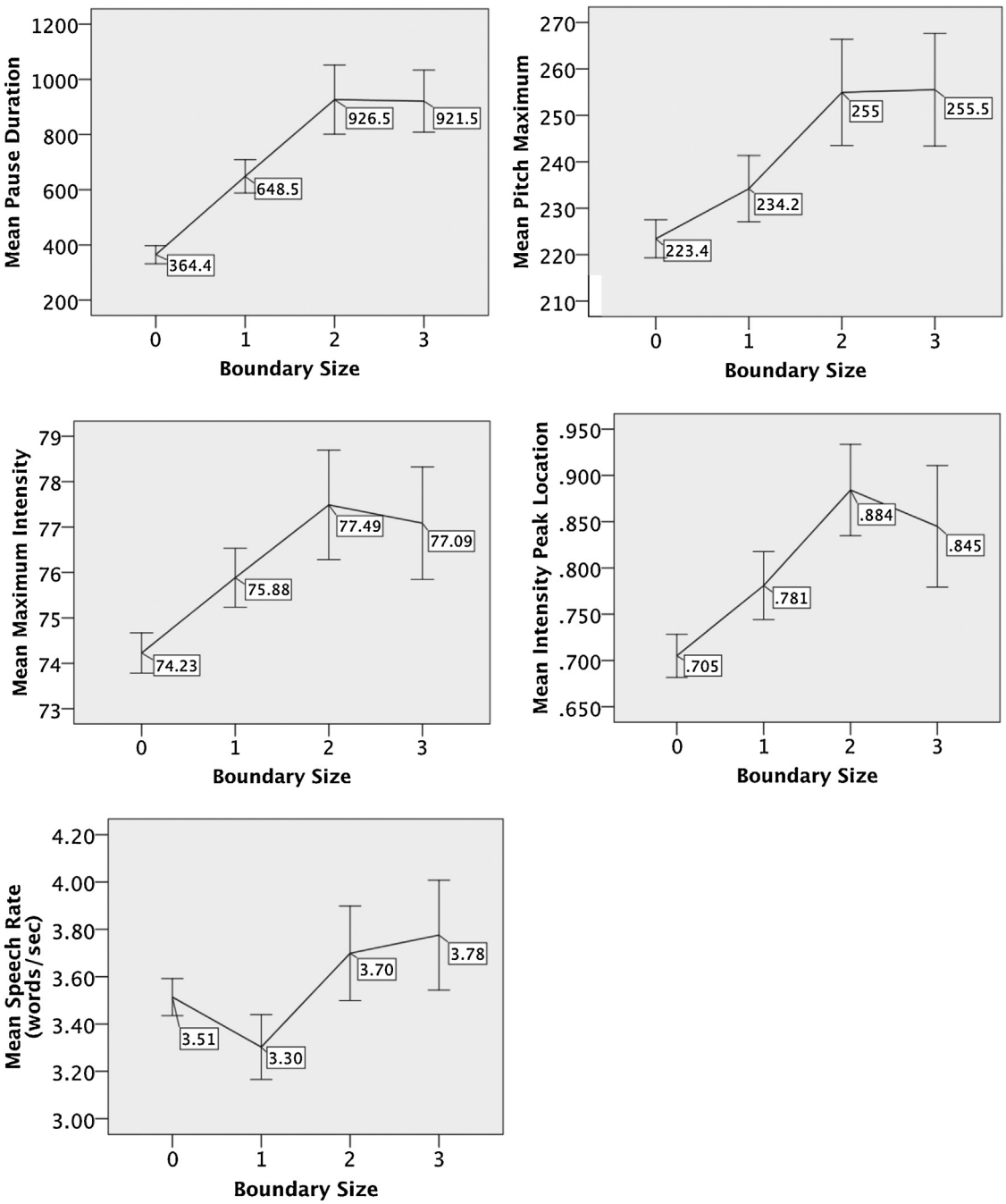

Fig. 5. Line graphs with boundary size on *x*-axis and pause duration, f0max, max intensity, intensity peak location and speech rate on the *y*-axis. Error bars indicate 95% confidence intervals.

in one way that spans multiple discourse segments and is independent of the discourse structure. For example, speakers may get into periods of longer or shorter pauses.

In sum, measures of preceding pause duration, pitch maximum and intensity maximum all increased as preceding boundary size increased. Moreover, intensity peaks occurred later in segment after larger boundaries. And while Fig. 5 indicates speech rate drops from boundary size 0 to 1 before increasing at levels 2 and 3, overall it was the case that speech rate increased following larger boundaries.

Table 7
Results for all independent variables in the model with boundary size as the only predictor variable of interest. Prosodic measures are in the left column, and predictor variables are along the top row.

	Intercept	Sentinit	Number	quot	Duration	Prosprev	BSize
Pause duration							
F	88.521	295.646	0.914	1.058	20.965	21.427	25.806
p	0.000**	0.000**	0.340	0.348	0.000**	0.000**	0.000**
Maximum pitch							
F	215.540	72.485	6.065	1.706	87.167	0.454	11.449
p	0.000**	0.000**	0.014*	0.182	0.000**	0.501	0.001**
Pitch peak location							
F	61.690	11.079	0.732	1.680	1.525	8.093	2.432
p	0.000**	0.001**	0.393	0.187	0.217	0.005**	0.119
Maximum intensity							
F	644.139	49.337	10.406	9.283	71.205	4.176	18.243
p	0.000**	0.000**	0.001**	0.000**	0.000**	0.041*	0.000**
Minimum intensity							
F	237.309	5.090	0.181	1.357	177.497	22.302	0.140
p	0.000**	0.024*	0.670	0.258	0.000**	0.000**	0.709
Intensity peak location							
F	281.138	3.777	0.278	1.761	2.444	2.715	17.313
p	0.000**	0.052	0.598	0.173	0.118	0.100	0.000**
Speech rate							
F	399.602	1.025	28.866	7.810	38.547	4.826	10.734
p	0.000**	0.312	0.000**	0.000**	0.000**	0.028*	0.001**

* $p < 0.05$.

** $p < 0.01$.

3.2. CoordSubord

To identify overall patterns of CoordSubord on prosodic outcomes, a model was fitted that contained CoordSubord as a predictor but not boundary size. This model tells us what effect a change in a segment being subordinated or coordinated has on each prosodic outcome. CoordSubord was entered as a binary categorical variable. In Table 8, we see the results for each prosodic measure by row. The intercept indicates the model's predicted value for each prosodic measure when a discourse segment is subordinated. The coefficient indicates the model's predicted slope, i.e. the amount of change in that prosodic measure when a segment, instead of being subordinated, is coordinated.

Results for CoordSubord indicate that a discourse segment's preceding pause duration, max pitch and max intensity all increase when a discourse segment is coordinated instead of subordinated. Furthermore, a coordinated discourse segment's pitch peak occurs earlier while its intensity peak occurs later relative to subordinated discourse segments. The graphs in Fig. 6 plot each prosodic measure on the *y*-axis for both Coord and Subord on the *x*-axis. Results for the control variables are presented in Table 9.

Table 8
Results for CoordSubord as a predictor of prosody, collapsing across Boundary Size. The intercept indicates the model's predicted value for each prosodic measure when a segment is subordinated (the reference value). The coefficient indicates the model's predicted slope, i.e. the amount of change in that prosodic measure by being coordinated instead of subordinated.

CoordSubord	Intercept	Coefficient	*F*-statistic	*p*-Value
Pause	695.116	123.877	25.544	0.000**
Maximum pitch	221.491	10.496	25.604	0.000**
Pitch peak location	0.181	−0.054	5.900	0.015*
Max intensity	70.869	1.194	25.827	0.000**
Minimum intensity	33.170	−0.148	0.173	0.678
Intensity peak location	0.758	0.054	6.996	0.008**
Speech rate	4.107	0.065	1.063	0.303

* $p < 0.05$

** $p < 0.01$.

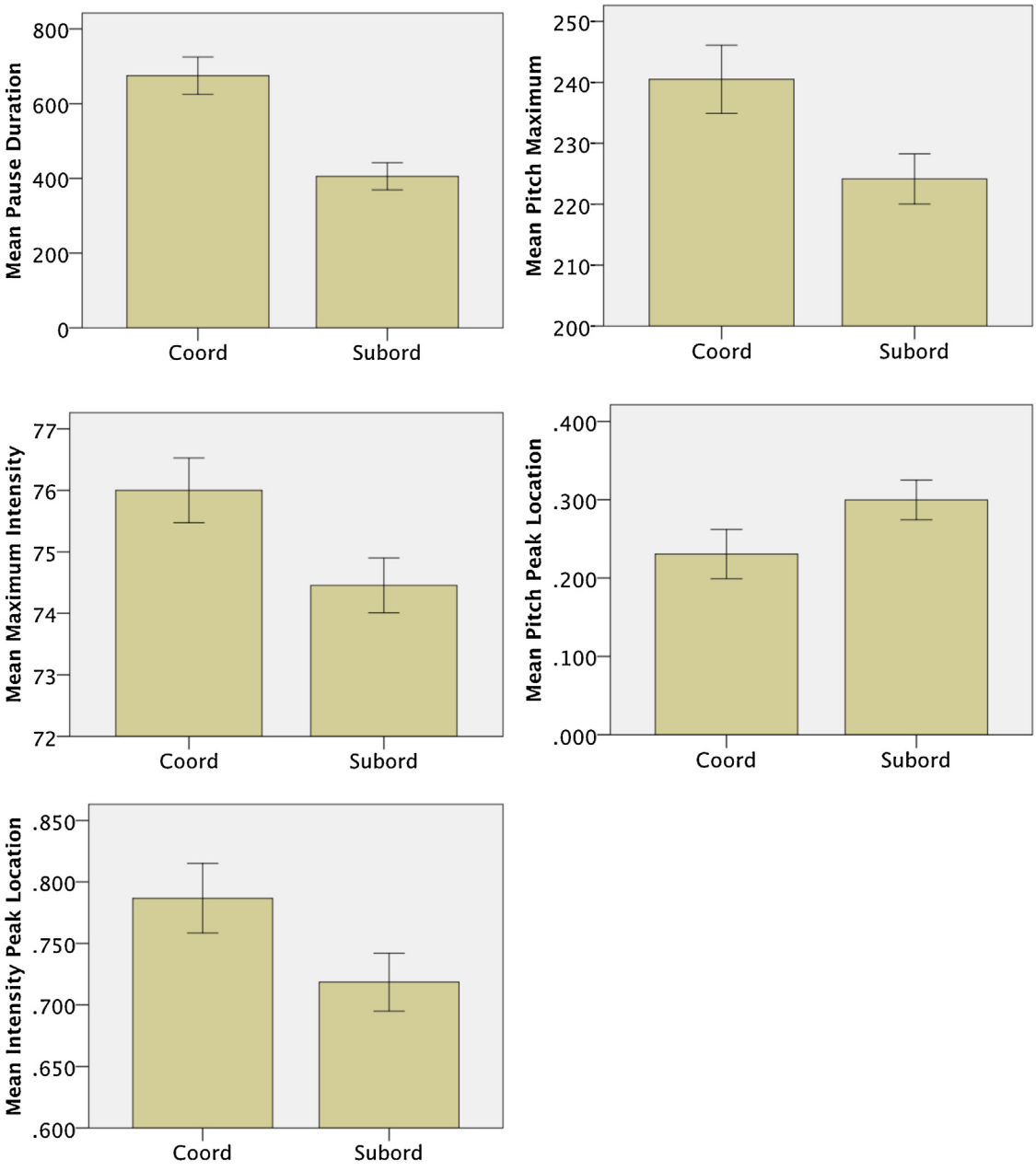

Fig. 6. Bar graphs with CoordSubord on *x*-axis and pause duration, f0max, max intensity and speech rate on the *y*-axis.

Like in the model with boundary size, there are many significant effects of the control variables, demonstrating the importance of having them in the model. Only two results are different when CoordSubord is in the model instead of Bsize. First, sentence-initiality becomes a strong predictor of intensity peak location. This is perhaps not that surprising if we recall that sentence-initiality and Bsize are somewhat correlated ($r = 0.471$). When Bsize is not accounting for some of the variation in a prosodic outcome, sentence-initiality fills some of that absence. And second, the speech rate of the previous segment no longer predicts speech rate of the current segment. This suggests that part of why a previous segment's speech rate was predictive of the current segment's speech rate is due to whether those segments are linked to the larger discourse via coordination or subordination. So even though CoordSubord does not predict speech rate, it seems to have

Table 9
Results for all independent variables in the model with CoordSubord as the only predictor variable of interest. Prosodic measures are in the left column, and predictor variables are along the top row.

	Intercept	sentinit	Number	quot	Duration	Prosprev	CS
Pause duration							
F	122.865	351.059	1.777	0.666	28.632	35.860	25.544
p	0.000**	0.000**	0.183	0.514	0.000**	0.000**	0.000**
Maximum pitch							
F	236.632	81.453	4.340	2.495	103.376	0.042	25.604
p	0.000**	0.000**	0.038*	0.083	0.000**	0.838	0.000**
Pitch peak location							
F	54.040	12.188	1.094	2.160	2.432	6.885	5.900
p	0.000**	0.001**	0.296	0.116	0.119	0.009**	0.015*
Maximum intensity							
F	662.713	60.715	12.946	11.942	85.234	4.251	25.827
p	0.000**	0.000**	0.000**	0.000**	0.000**	0.040*	0.000**
Minimum intensity							
F	237.196	5.439	0.203	1.415	176.765	21.780	0.173
p	0.000**	0.020*	0.652	0.244	0.000**	0.000**	0.678
Intensity peak location							
F	338.275	9.579	0.527	2.099	4.107	1.171	6.996
p	0.000**	0.002**	0.468	0.123	0.043*	0.280	0.008**
Speech rate							
F	428.699	0.015	29.782	8.378	35.582	2.835	1.063
p	0.000**	0.903	0.000**	0.000**	0.000**	0.093	0.303

* $p < 0.05$.

** $p < 0.01$.

an effect on other factors. We will explore the complexity of the relationship between CoordSubord and speech rate in more detail in the next section.

3.3. Interaction of boundary size and CoordSubord

In addition to modeling Bsize and CoordSubord independently as predictors of prosodic outcomes, we want to see if the effect of one variable depends on the value of the other. For example, the CoordSubord contrast may only be relevant at some levels of Bsize. We can test this by modeling both predictors together in the same model, including each as a main effect as well as their interaction. If the interaction is significant, then the slope for coordinated segments across the levels of boundary size differs from the slope for subordinated segments across the levels of boundary size. That is, a significant interaction would tell us that the effect of a segment being coordinated vs. subordinated would depend on the size of the preceding boundary.

A linear mixed model was fitted to the data with Bsize, CoordSubord and their interaction as predictors, along with the controls listed in Table 5, for each prosodic outcome. Results are shown in Table 10. Table 10 shows a significant interaction between Bsize and CoordSubord for pause duration, max pitch, max intensity, and speech rate. This means that for each of these prosodic measures, the effect of CoordSubord depends on Bsize. The nature of this interaction is visible in the graphs in Fig. 7. These four graphs plot boundary size along the *x*-axis, with the relevant prosodic measure on the *y*-axis. The blue dashed line corresponds to Coord and the solid green line corresponds to Subord. For three of the four measures (pause duration, max pitch and max intensity), we see separation between Coord and Subord when Bsize = 0. In these three cases, Coord has a higher value (longer pause duration, higher max pitch and max intensity). As boundary size gets bigger, the Coord and Subord lines get closer, meaning that the differences between Coord and Subord get smaller. Speech rate behaves differently. For speech rate, Subord is higher at Bsize = 0, the lines cross and separation increases with Coord higher at higher levels of boundary size.

For all four of these prosodic measures, results in Table 10 show us that the difference between Coord and Subord is significant when Bsize = 0 (the reference value for Bsize). We can test whether the difference between Coord and Subord is significant at the other levels of Bsize by making each level the reference value for Bsize (Table 11). The results in Table 11 show that the difference between coordinated and subordinated segments is significant for pause duration, max pitch and max intensity when Bsize = 0 or 1, but not when Bsize > 1. By contrast,

Table 10
Table of results for linear mixed model with Bsize, CoordSubord and the interaction Bsize × CoordSubord. The CoordSubord result indicates whether prosodic outcomes are significantly different between Coord and Subord when Bsize = 0. The Bsize result indicates whether the prosodic outcomes for Subord (the reference value of CoordSubord) change as Bsize changes.

	Bsize	CoordSubord	Interaction
Pause			
F	25.698	30.857	8.795
p	0.000**	0.000**	0.003**
Coefficient	119.936	165.961	−84.734
Maximum pitch			
F	9.110	25.147	4.111
p	0.003**	0.000**	0.043*
Coefficient	6.347	12.782	−4.908
Pitch peak location			
F	1.641	2.571	0.185
p	0.201	0.109	0.667
Coefficient	−0.012	−0.044	−0.011
Max intensity			
F	16.171	27.501	5.788
p	0.000**	0.000**	0.016*
Coefficient	0.899	1.507	−0.657
Minimum intensity			
F	0.118	0.165	0.029
p	0.731	0.685	0.865
Coefficient	−0.110	−0.180	0.071
Intensity peak location			
F	14.996	2.689	0.098
p	0.000**	0.101	0.754
Coefficient	0.045	0.041	0.008
Speech rate			
F	7.500	7.533	33.086
p	0.006**	0.006**	0.000**
Coefficient	−0.104	−0.206	0.413

* $p < 0.05$.

** $p < 0.01$.

speech rate shows a significant contrast between Coord and Subord at every level of boundary size. In Fig. 7, we see that speech rate increases for Coord segments and decreases for Subord segments as Bsize increases. Moreover, the two lines cross. At Bsize = 0, Coord segments are spoken significantly slower than Subord segments. But at all levels of Bsize > 0, Coord segments are spoken significantly *faster* than Subord segments. We saw there was a main effect of Bsize on speech rate (see Table 6) but no main effect of CoordSubord on speech rate (see Table 8). While one may have interpreted the lack of main effect of CoordSubord on speech rate as meaning CoordSubord didn't matter for speech rate, the interaction results show it just matters in a more complex way. By analyzing the interaction effect, we see that the impact of boundary size on speech rate is mediated by whether a discourse segment is coordinated or subordinated.

Table 11
Results testing whether the prosodic outcomes are significantly different between Coord and Subord measurements at each level of boundary size.

CoordSubord	at Bsize = 0	at Bsize = 1	at Bsize = 2	at Bsize = 3
Pause duration	$F = 30.857$; $p < 0.001$**	$F = 9.431$; $p = 0.002$**	$F = 0.006$; $p = 0.940$	$F = 0.006$; $p = 0.940$
Max pitch	$F = 25.147$; $p < 0.001$**	$F = 11.954$; $p = 0.001$**	$F = 0.564$; $p = 0.453$	$F = 0.100$; $p = 0.752$
Max intensity	$F = 27.501$; $p < 0.001$**	$F = 11.058$; $p = 0.001$**	$F = 0.190$; $p = 0.663$	$F = 0.448$; $p = 0.503$
Speech rate	$F = 7.533$; $p = 0.006$**	$F = 9.184$; $p = 0.003$**	$F = 27.563$; $p < 0.001$**	$F = 31.805$; $p < 0.001$**

* $p < 0.05$.

** $p < 0.01$.

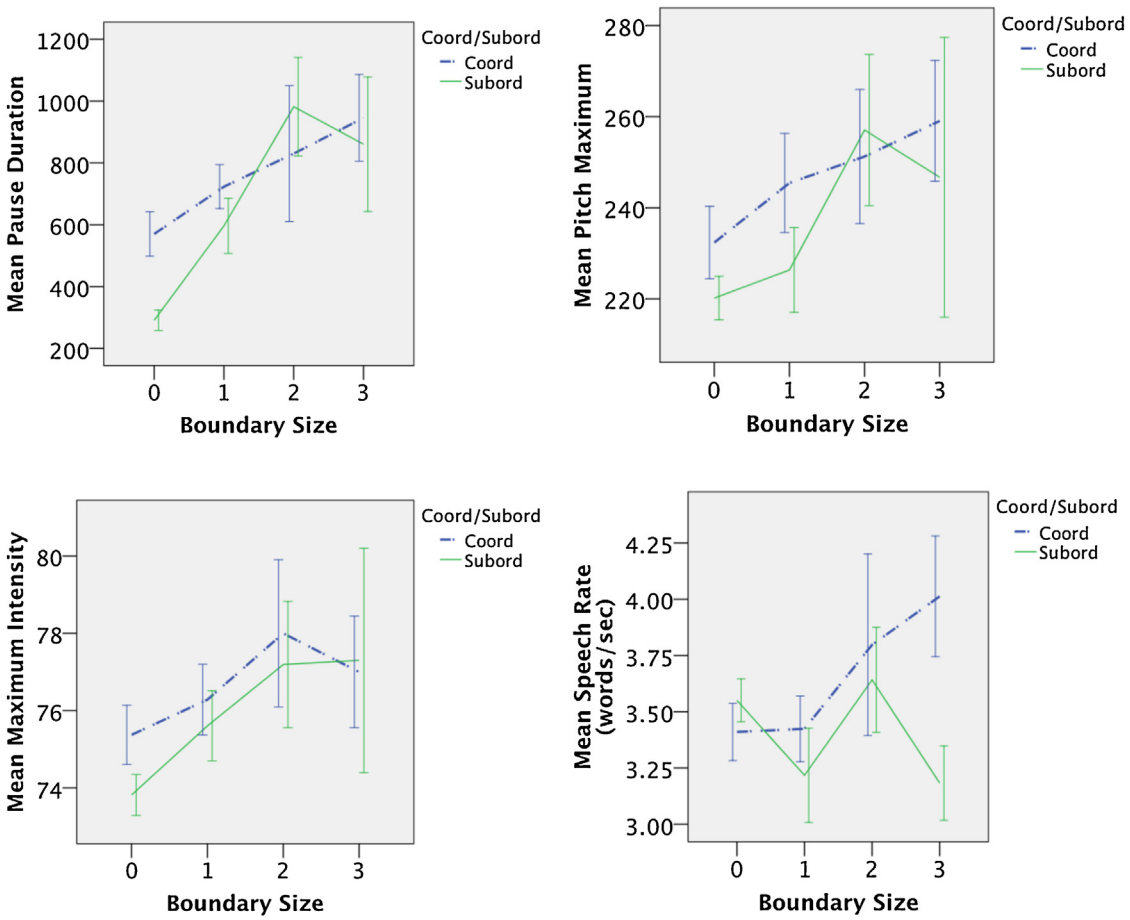

Fig. 7. Line graphs with boundary size on *x*-axis, a dashed blue line for Coord and a solid green line for Subord. The graphs are tiled by prosodic measure, with the relevant scale for each on their *y*-axis (pause duration, f0max, max intensity and speech rate).

4. Discussion

This study has found evidence of prosodic correlates of both boundary size (Bsize) and coordination vs. subordination (CoordSubord). Moreover, it identified significant interactions between Bsize and CoordSubord, showing that prosody's relationship to CoordSubord is integrally related to its relationship to Bsize and vice versa. The interaction of Bsize and CoordSubord was significant for pause duration, max pitch, max intensity, and speech rate. This means that for these measures the effect of Bsize or CoordSubord depends on the value of the other.

The results for Bsize show increasing values for pause duration, max pitch and max intensity as Bsize increased. This is in line with existing research on prosodic correlates of discourse structure, which Smith (2004) summarizes as suggesting "greater prominence at the beginning of a discourse or immediately after a major boundary" (p. 250). This greater prominence is indicated in this study by longer pauses and higher pitch and intensity. Speech rate is more complicated, showing an overall trend of increasing speech rate with increasing boundary size. But the picture is actually more complex, where speech rate actually drops from level 0 to 1, and then increases substantially for levels 2 and 3. Furthermore, there is a significant interaction between Bsize and Coord in the prediction of speech rate. Subordinated segments actually show a mild slowing in speech rate as Bsize increases, while Coord segments get produced faster. And finally, intensity peaks occur later in a discourse segment as Bsize increases. This suggests that later intensity peaks, by patterning with the other prosodic measures, may work in tandem with pause duration, max pitch and max intensity to indicate greater prominence.

Results for CoordSubord showed higher values of pause duration, max pitch and max intensity for coordinated segments than subordinated segments. This provides prosodic evidence that coordination is in a more prominent position, because the same measures that conveyed prominence for Bsize do so for CoordSubord as well. But Bsize and

CoordSubord are not correlated with each other ($r = 0.022$, see Table 3), so this prosodic prominence is conveying different information about the discourse structure. It also makes sense to think of coordination as more prominent than subordination, as by definition coordinated segments are hierarchically higher than subordinated segments, all else being equal.

More surprising are the results for the two proportional measures for pitch and intensity peak locations, which showed that coordinated segments had earlier pitch peaks but later intensity peaks. Given that pitch and intensity peak values pattern together for Bsize, it is remarkable that they pattern in opposite directions in terms of how far through the discourse segment those peaks occur. The result for earlier pitch peaks is consistent with research on pitch reset and claims that high onset pitch occurs at topic onsets (Auran, 2007; Yule, 1980). Less is known about the behavior of intensity peaks, raising questions about how these two measures are able to operate independently and in opposite directions.

We also know relatively little about prosodic correlates of coordination and subordination in discourse, though den Ouden et al. (2009) test something similar in their study of prosodic correlates of the RST distinction between nuclei and satellites. In RST, all discourse segments are classed as either nuclei or satellites, where satellites are those segments that are less important and can more easily be removed without disturbing the larger coherence of the discourse. Danlos (2010) compares RST and SDRT in terms of their theoretical underpinnings and ability to account for the felicity and infelicity of discourses. He concludes that the two theories roughly rely on the same set of discourse relations and give them the ''same type,'' i.e. coordinating/subordinating or nucleus/satellite. He seems to be treating the two binaries coordinating/subordinating and nucleus/satellite as comparable and in some sense equivalent. It is noteworthy then that den Ouden et al. (2009) found different results for prosodic correlates of the nucleus/satellite distinction than this study found for CoordSubord. den Ouden et al. found no correlation between pause duration or max pitch with nuclei and satellites, but did find that nuclei were produced with a slower articulation rate than satellites. In contrast, this study found coordinated segments were produced with longer preceding pause durations and higher maximum pitch, but no difference in speech rate. There are a few possible interpretations for these contrasting results. First, den Ouden et al. (2009) used Dutch texts and Dutch participants while this study was conducted entirely in American English. Perhaps the languages themselves can account for the different prosodic correlates. Second, it could be due to a difference in coding, where the way the nucleus/satellite distinction in den Ouden et al. (2009) and CoordSubord in this study were coded differed. Third, it is possible that RST's nuclearity and SDRT's coordination/subordination contrast are not actually equivalent. One piece of evidence for nuclearity and CoordSubord being different is in their relation to measures of boundary size. In den Ouden et al. (2009), the measures for nuclearity and boundary size, which they call ''Hierarchy'', are highly correlated ($p < 0.001$) (p. 125). In the study described in this paper, Bsize and CoordSubord are not correlated ($r = 0.022$). This suggests the nuclearity variable in den Ouden et al. (2009) is capturing much of the same information as boundary size, while CoordSubord in this study reflects different features of the discourse than boundary size.

And while the independent variables Bsize and CoordSubord are not correlated, results show a significant interaction between them as predictive of pause duration, max pitch, max intensity, and speech rate. For pause duration, pitch max and intensity max, there is a significant difference between coordinated and subordinated discourse segments at levels 0 and 1 of Bsize, but this CoordSubord effect disappears when Bsize is 2 or 3. This indicates that when the boundary between segments is smaller, information about whether the new segment is coordinated or subordinated matters. But when boundary size increases, the coordination/subordination contrast is no longer significant.

The interaction between Bsize and CoordSubord is significant for speech rate but in a different way. For speech rate, CoordSubord is a significant predictor at all levels of Bsize. When Bsize = 0, subordinated segments are produced significantly faster. But when Bsize is 1 or larger, coordinated segments are produced faster. Furthermore, the speech rate of subordinated segments falls mildly as Bsize increases. By contrast, coordinated segments show a steady increase in speech rate as Bsize increases. This suggests that much of the effect of speech rate occurs in the coordinated segments. So, why would coordinated segments be spoken faster as Bsize increases, but subordinated segments would not? One way coordinated and subordinated segments differ is in terms of novelty. A new coordinated segment is creating a new space in the discourse, while a new subordinated segment is providing more information about something already under discussion. Perhaps relative novelty leads to a different interaction with boundary size. It is also possible this effect is related to the speech being monologic, read speech. In this experiment, speakers did not have a listener present with whom to interact and for whom to adjust their speech. It is possible that speakers behaved more with respect to their own needs than if listeners were present. They also were tasked with reading a text aloud verbatim without speech errors. In this task, the speech planning process involved reading instead of planning in one's own head. Perhaps the reliance on text for linguistic material affected speaking rate. It is unclear why novelty, monologue or reading aloud would lead to faster speech after larger boundaries, but these factors may be involved in an explanation. A separate explanation would say that listeners expect a default speech rate for default interpretations. Since large boundaries are less common than smaller ones, it is a relatively marked context. Perhaps this study's speakers were using marked prosody, i.e. faster speech, as a way of conveying this marked discourse context.

This study also examined measures which capture temporal information about how fast a speaker gets to pitch and intensity peaks. These measures reveal different patterns for pitch and intensity peaks: coordinated segments have earlier pitch peaks and later intensity peaks relative to subordinated segments. Intensity peaks also were later in a segment as Bsize increased. These results demonstrate there is potentially meaningful prosodic variation along this temporal dimension. Therefore, a fuller account of discourse prosody will need to take into account the location of prosodic peaks in addition to the values of those peaks.

As the above discussion shows, there are a number of significant correlations between discourse structure and speakers' prosody in the context of this study. Especially notable is how similar the correlations are between Bsize and the prosodic measures of pause duration, max pitch and max intensity. This raises the question of whether all three measures independently correlate with discourse structure, or whether there is some underlying prosodic category that gets fed forward to the phonetic realization of pause duration, pitch and intensity. If we posit a direct relationship, then we miss the potential generalization across the prosodic measures. Instead, we could posit an underlying category that mediates between discourse and the acoustic measures. Instead of speakers directly connecting discourse to the acoustics, they would have a representation of something like discourse prosodic emphasis; cf. Smith's (2004) term ''greater prominence'' to refer to similar patterns across prosodic measures (p. 250). In this case, one way discourse structure would interface with prosody is by generating more or less prosodic emphasis, which would itself then get spelled out in terms of phonetic measures like pause duration, pitch and intensity.

But if there is an underlying category behind the overt manifestations of pause duration, pitch and intensity, it raises questions about why there is still so much variability from measure to measure. It also raises the question of what motivation there could be for providing redundant cues to the structure of discourse. One possible explanation is that this variability provides necessary flexibility for discourse prosody to convey information about discourse structure. There are many factors that can affect the realization of pause duration, pitch and intensity, and can have an effect in different ways for the different prosodic measures. While overall the prosodic correlates of discourse structure may appear to be redundant, individual productions could exploit only some subset of the three prosodic measures. For example, in cases where pause duration is determined by other factors, speakers can still draw on pitch or intensity. What may from a macro-perspective seem redundant, in more individual instances could be important flexibility. Hirschberg and Grosz (1992) make a similar conclusion when they write that ''different configurations of intonational features may be employed to convey the same discourse information in different contexts. For while our aggregate statistics show certain trends, not every token exhibits all these differences'' (p. 446). Furthermore, redundancy of cues can also reinforce meanings that could otherwise be difficult to convey. In fact, redundant cues in production appear to facilitate the perception of discourse prosody (Mayer et al., 2006; Silverman, 1987).

It is also worth mentioning that this study tested for correlation, not causation. Subsequent research could try to determine whether discourse structure *causes* prosodic correlates. For example, if speakers are presented with a single discourse that could be interpreted in two ways corresponding to two different discourse structures, would speakers produce the different structures with different prosody? And could listeners successfully communicate to listeners which interpretation they intended? Holding the lexical and syntactic information constant while varying the discourse structure would help isolate the discourse structure as the cause of the prosodic correlates.

A potential criticism is that using acoustic measures only indirectly corresponds to theorized underlying categories of an abstract prosodic-phonological structure. This criticism may claim the utility of acoustic measures is hampered by variation for non-prosodic reasons like segmental perturbations or performance errors. The problem of variation due to performance errors was likely reduced by excluding disfluent speech segments from the analysis and including data from many speakers with many observations per speaker. Any errors that remain could add some noise to the data, but are unlikely to be the reason for the patterns identified. Segmental effects on f0 also exist, where F0 is higher immediately after voiceless consonants than after voiced consonants (Hanson, 2009; Lofqvist et al., 1989; Ohde, 1984). These effects are unlikely to account for the results above because each discourse segment from which the acoustic measures were extracted would likely contain numerous voiced and voiceless consonants, and there is no reason to expect voiced or voiceless segments to be distributed in ways that correlate with the predictors of interest.

And while there may be difficulties in using acoustic measures, their use for the analysis of prosody has a long history, including the discourse prosody studies of den Ouden et al. (2009) and Hirschberg and Grosz (1992), Breen et al.'s (2010) study of acoustic correlates of information structure, and Lehiste's (1982) analysis of the phonetic characteristics of speech as part of her interest in ''higher-level phonological structures'' (p. 117). Moreover, the intuitive judgments of prosodic categories need be grounded at least somewhat in acoustic information given that it is by listening to speech that we make our judgments. Some scholars (Price et al., 1991) have included both intuitive categories and acoustic measurements in the same study. And Podesva (2006) argues for the importance of phonetic in addition to phonological measures, claiming that ''disregarding fine-grained phonetic detail is an act of erasure (Irvine, 2001; Irvine and Gal, 2000), rendering invisible some differences in form that in fact exist in the realm of speech production, acoustics, and perception. While [he] would certainly not suggest that adopting a categorical view is misguided, adopting a solely categorical view

may be'' (p. 5). The use of prosodic phonological categories is also not without its problems, including that a system like ToBI has low inter-annotator agreement (Syrdal and McGory, 2000) and that existing phonological categories may not actually represent the phonological and phonetic facts (Dilley, 2010). This suggests that relying too much on transcriptions like ToBI may at times obscure rather than reveal the facts.

The point of this discussion is to emphasize that the study of prosody can benefit from approaches that use acoustic measures as well as those that use intuitive judgments. This study's findings are about phonetic correlates of discourse structure, and do not explain how those results may be mediated by phonological prosodic categories (e.g. the pitch accents and boundary tones of a transcription system like ToBI). The role of such a phonology is beyond the scope of this paper, which made no assumptions about a theory of phonology. Nevertheless, it does suggest potentially valuable avenues for future research exploring how the phonetic correlates discussed here interact with phonological categories. For example, exploring the role of pitch accents and pitch reset could help explain what led to the different pitch maxima results between Coord and Subord discourse segments.

4.1. Paraphrase analysis

In discourse prosody research, a common concern has been how to get a good representation of the discourse's structure independent of any prosody. When using spoken data as the basis of the structural analysis, there is a risk of circularity where prosodic information motivates the structure that is then correlated with prosodic measures. Scholars have tried to solve this problem by focusing on discourses with relatively uncontroversial structures, like BBC news broadcasts (Wichmann, 2000), or using naïve participants to mark discourse boundaries (Swerts, 1997). Others have used a specific discourse theory, like the Grosz & Sidner model (Grosz and Hirschberg, 1992) or Rhetorical Structure Theory (den Ouden et al., 2009), taking the theory to provide a good approximation of how participants were representing the structure of discourse. Similarly, this study used a specific discourse theory (SDRT) and took the annotations to be a good approximation of how participants were representing the discourse's structure. Unlike previous studies, this one had participants paraphrase the discourse before reading it aloud, which could provide one way to check whether participants' sense of key points corresponded well to key points in the SDRT representation. If the main topics of the paraphrases line up well with the main topics of the SDRT analysis, then we have evidence to support the claim that SDRT is capturing something about how participants are representing the discourse.

To explore whether such a correspondence existed, I first listened to each participant's paraphrase of the article and noted the topics mentioned. Then, I examined the SDRT representation to see what topics are in those discourse segments after the largest boundaries (level 3). I am using the term *topic* here to capture something like discourse topic, i.e. what the content is about. For a full list of topics and results of this analysis, see Appendix D.

Results show that nearly all participants mentioned topics 1 and 2, with fewer mentioning subsequent topics. The first topic, the overall topic for the article, was not captured by the boundary size measure. This is not surprising given that the overall topic of the article was mentioned at the very beginning, before enough has been said for there to be a large boundary. The second topic, addressing specific concerns with a crime bill, was also mentioned by nearly all participants. Subsequent topics received a minority of mentions, apparently with decreasing mentions as the topics were later in the article. This suggests paraphrasing may highlight topics introduced earlier in a discourse, with subsequent topics deemed less integral. Furthermore, each paraphrase was short, lasting between 45 and 90 s. In this time, participants only covered 2–4 topics, clearly choosing to emphasize the first two. These paraphrases suggest participants understood the discourse, and that the boundary size measure grounded in the SDRT representation captures some relevant aspects of how participants understood the discourse.

5. Conclusion

The significance of this study's findings are constrained in two ways. First, the generalizability of these findings is limited by the use of read speech. Because read speech has been shown to differ from spontaneous speech (Blaauw, 1994; Laan, 1997), we cannot assume that this study's findings would necessarily show up in non-read speech. But even in this constrained context of read speech, it does show that speakers can produce speech in such a way as to carry discourse structural information. The skill of the reader has also been found to be an important dimension for variation in read speech. As noted by Esser (1988) and Wichmann (2000), amateur and professional readers differ in how they read aloud, with professional readers tending to more consistently use prosodic features. This study used amateur readers and was still able to identify prosodic correlates of discourse structure.

Second, by using only SDRT to represent the discourse structure with which prosodic measures were correlated, comparisons cannot be made between the representations of SDRT and other theories. Were one to annotate the structure of a single discourse using multiple theories, then prosodic correlates could reveal which theory had the strongest correlations and permit comparisons between theories (e.g. den Ouden (2004) using RST and the Grosz &

Sidner model). This information could help identify which theories have the strongest prosodic correlates and potentially adjudicate between them.

Having demonstrated some ways prosodic measures correlate with discourse structure, this study does motivate follow-up work that could test whether listeners exploit that information in their perception. I see two kinds of issues discourse prosody perception studies could address: disambiguation and facilitation. If identical lexical material has multiple possible discourse structures, could prosody bias interpretation towards one or another? If stimuli are created where there is a mismatch between the prosody and the discourse structure, would listeners show processing difficulties? A study in Auran (2007) suggests mismatching prosody can induce processing difficulties in the interpretation of French discourse. Also, if one discourse is produced with distinct discourse prosody and another without, would listeners rate the speech with the discourse prosody as easier to understand or as more effective? Would comprehension or retention increase? It is possible one aspect of what makes good speakers easier to understand is their use of discourse prosody.

Finally, the findings discussed in this paper may lend themselves to practical applications, e.g. speech synthesis and speech training. Because speech synthesis systems currently tend to suffer from unnatural-sounding prosody, perhaps the correlates identified here could help inform ways to improve them. And if discourse prosody is found to facilitate comprehension and assessments of speaker effectiveness, teaching people to use discourse prosody could help them become more effective speakers.

Acknowledgments

I would like to thank Ezra Keshet and Robin Queen for their comments, contributions and support over the course this project. I am also grateful for the insights provided by the three anonymous reviewers, Nicholas Asher, Delphine Dahan, Carlos Gussenhoven, Gisela Redeker, Julie Boland, Jason Kahn, Lauren Squires.

Appendix A. Full text of newspaper article used in production study with paragraphing removed, as presented to participants

Politics & policy: blacks' increasing vocal opposition to violence is matched by strong opposition to crime bill -- by Joe Davidson staff reporter of The Wall Street Journal. The Rev. Jesse Jackson, the often fiery Rainbow Coalition president, was subdued, reflective, nearly rhymeless. At a recent hearing of the Congressional Black Caucus brain trust on crime, he spoke solemnly, his voice breaking, of how some young black men feel ''more secure in jails than on our streets.'' With tears in his eyes, he spoke of death in his own neighborhood here and the precarious position of black youth. ''Nearly half of all murder victims are black,'' he said. ''More blacks kill each other each year than were killed in the entire history of lynching.'' Yet, the Rev. Jackson assailed one of the prime legislative vehicles for dealing with that explosion of violence – the Senate-passed crime legislation that President Clinton backed in his State of the Union address. The measure, he declared, is an ''ill-conceived bill'' and a ''Draconian... expensive non-remedy.'' The bill has widespread bipartisan support in the Senate. Lawmakers contend it represents the toughest and most comprehensive government attack yet on violent crime, an issue at the top of the public's list of concerns in opinion polls. But at a time when African-Americans increasingly are speaking out against black criminals and the ''gangsta rap'' that seems to glorify violence, the Black Caucus and others say the Senate bill is too concerned with punishment, and not enough concerned with the alleviation of the conditions that cause crime. The strong opposition to the measure presents a problem for President Clinton, whose support for the legislation places him at odds with a core group of Democrats who elected him. Citing Mr. Clinton's embrace of one provision of the Senate bill – mandatory life sentences for criminals convicted of three violent felonies – the dean of the Black Caucus, Democratic Rep. John Conyers of Michigan, decries the ''lock-'em up and throw away the key'' approach that ''only fools the public into believing that we're doing something about crime.'' The White House will try to assuage at least some opponents' concerns as Congress undertakes to reconcile the Senate bill with a much different House measure. Justice Department officials, who were criticized for not visibly exerting influence over the Senate bill last year, will play a more overt role in removing or modifying the more extreme provisions this year. Deputy Attorney General Philip Heymann plans to testify at House crime legislation hearings, and Mr. Clinton himself held out the carrot of help to endangered youth in his speech to Congress. ''We have got to stop pointing our fingers at these kids who have no future,'' he said, ''and reach our hands out to them.'' The question, though, is whether enough changes can be made to the bill to soften opposition to it. In addition to the Black Caucus, a range of others – including the American Bar Association, the American Civil Liberties Union, the National Conference of State Legislators and many federal judges and prosecutors – oppose stringent sentencing provisions in the bill. Other less controversial provisions in the 22.3 billion dollar legislation include authorization for 100,000 additional police officers, drug treatment and other crime-prevention programs. Some black leaders, such as the leadership of the Nation of Islam, have long spoken out against crime and for the kind of values that make it unacceptable. But the mainstream civil-rights leadership generally avoided the rhetoric of ''law and order,'' regarding it as a code for keeping blacks back. Law and order didn't mean justice, Mr. Jackson used to say, but ''just us.'' In

the past, many were hesitant to speak about crime in public because ''the larger community would talk about 'lock them up and throw the key away' and hide behind black leaders in doing it,'' explains Rep. Craig Washington, the Houston Democrat who led the caucus hearing. Now there is escalating discourse within the black community about what it can and must do to stop crime. Just after the new year, Mr. Jackson held the first of several conferences focusing on just that. ''The premier civil-rights issue of this day is youth violence in general and black-on-black violence in particular,'' he has said. His conference also noted the structural conditions that encourage crime – the sorry state of the black economy, high unemployment, poor education and a legacy of racism. ''The black leaders recognize that if they don't step out front and engage in the discussion, that basically our young people are turning themselves into slaves,'' says Rep. Washington. Within the black community, there is ''more public concern and debate about the appropriate level of response to increasing crime and violence.'' Many of the black leaders involved in the growing debate retain strong objections to the Senate bill, with its large number of mandatory minimum sentences, death penalties and federalization of local crimes. One of the Senate measures strongly opposed by most members of the Black Caucus has as its author one of its own, Illinois Democratic Sen. Carol Moseley-Braun. Her amendment would restrict prosecutorial discretion – a point opposed by Attorney General Janet Reno – by directing U.S. attorneys to prosecute as adults 13-year-olds charged with committing violent crimes with firearms. The provision would federalize many crimes currently prosecuted by the states. Yet, notes federal Judge Maryanne Trump Barry of Newark, N.J., who is chairwoman of the criminal law committee of the Judicial Conference of the U.S., there is no federal juvenile justice system to handle such cases – no federal juvenile prisons, for instance, and no federal youth probation officers. The National Conference of State Legislatures is so opposed to the federalization of state crimes – another provision in the bill, pushed by GOP Sen. Alfonse D'Amato of New York, would federalize all violent handgun crimes – that it recently wrote President Clinton to say ''the Senate bill is inimical to principles of federalism, and we must oppose it.'' And a measure that would require states to adopt certain federal sentencing guidelines, such as mandatory minimum sentences, to get federal prison building funds is ''coercive policy,'' complains Jon Felde, NCSL's general counsel. There are numerous mandatory minimum provisions in the legislation that Mr. Washington fears could be used in an unfair fashion against blacks who may be charged more harshly than whites for similar acts. And federal judges have ''consistently, vehemently, and virtually unanimously opposed'' mandatory minimum sentences, Judge Barry wrote to Senate Judiciary Committee Chairman Joseph Biden, Democrat of Delaware, in November. Other measures that caucus members say could be used in a discriminatory way are those that would make it a federal crime to conspire to participate in a criminal street gang and that provides the death penalty for drug kingpins even if no death can be shown to have resulted directly from their illegal activity. The Justice Department has warned Congress that it thinks the drug kingpin provision is unconstitutional; the anti-gang measure will also be hit on constitutional grounds in the House. But Sen. Biden insists that the final legislation will include enough significant prevention and punishment provisions that liberals and conservatives alike will be able to endorse it. After all, he says, ''everybody is kind of singing from the same hymnal on the broad strokes.''

Appendix B. Full text of newspaper article used in production study as segmented according to SDRT in the DISCOR corpus

0. Politics & policy:
1. blacks' increasing vocal opposition to violence is matched by strong opposition to crime bill—
2. by Joe Davidson
3. staff reporter of *The Wall Street Journal*
4. The Rev. Jesse Jackson, the often fiery Rainbow Coalition president, was subdued, reflective, nearly rhymeless.
5. At a recent hearing of the Congressional Black Caucus brain trust on crime,
6 he spoke solemnly,
7. his voice breaking,
8. of how some young black men feel ''more secure in jails than on our streets.
9. ''With tears in his eyes, he spoke of death in his own neighborhood here and the precarious position of black youth.
10. ''Nearly half of all murder victims are black,''
11. he said.
12. ''More blacks kill each other each year than were killed in the entire history of lynching.''
13 Yet, the Rev. Jackson assailed one of the prime legislative vehicles for dealing with that explosion of violence—
14. the Senate-passed crime legislation that President Clinton backed in his State of the Union address.
15. The measure, he declared, is an ''ill-conceived bill'' and a ''Draconian. . . expensive non-remedy.''
16. The bill has widespread bipartisan support in the Senate.
17. Lawmakers contend it represents the toughest and most comprehensive government attack yet on violent crime,
18 an issue at the top of the public's list of concerns in opinion polls.

19. But at a time when African-Americans increasingly are speaking out against black criminals and the ‘‘gangsta rap’’ that seems to glorify violence,
20. the Black Caucus and others say
21. the Senate bill is too concerned with punishment, and not enough concerned with the alleviation of the conditions that cause crime.
22. The strong opposition to the measure presents a problem for President Clinton,
23 whose support for the legislation places him at odds with a core group of Democrats who elected him.
24. Citing Mr. Clinton’s embrace of one provision of the Senate bill—
25. mandatory life sentences for criminals convicted of three violent felonies—
26 the dean of the Black Caucus, Democratic Rep. John Conyers of Michigan, decries the ‘‘lock-’em up and throw away the key’’ approach that ‘‘only fools the public into believing that we’re doing something about crime.’’
27 The White House will try to assuage at least some opponents’ concerns
28. as Congress undertakes to reconcile the Senate bill with a much different House measure.
29. Justice Department officials, who were criticized for not visibly exerting influence over the Senate bill last year, will play a more overt role in removing or modifying the more extreme provisions this year.
30. Deputy Attorney General Philip Heymann plans to testify at House crime legislation hearings,
31. and Mr. Clinton himself held out the carrot of help to endangered youth in his speech to Congress.
32. ‘‘We have got to stop pointing our fingers at these kids who have no future,’’
33. he said,
34. ‘‘and reach our hands out to them.’’
35. The question, though, is whether enough changes can be made to the bill
36. to soften opposition to it.
37. In addition to the Black Caucus, a range of others – including the American Bar Association, the American Civil Liberties Union, the National Conference of State Legislators and many federal judges and prosecutors – oppose stringent sentencing provisions in the bill.
38. Other less controversial provisions in the $ 22.3 billion legislation include authorization for 100,000 additional police officers, drug treatment and other crime-prevention programs.
39. Some black leaders, such as the leadership of the Nation of Islam, have long spoken out against crime and for the kind of values that make it unacceptable.
40. But the mainstream civil-rights leadership generally avoided the rhetoric of ‘‘law and order,’’
41. regarding it as a code for keeping blacks back.
42. Law and order didn’t mean justice,
43. Mr. Jackson used to say,
44. but ‘‘just us.’’
45. In the past, many were hesitant to speak about crime in public
46. because ‘‘the larger community would talk about ‘lock them up and throw the key away’ and hide behind black leaders in doing it,’’
47. explains Rep. Craig Washington,
48. the Houston Democrat who led the caucus hearing.
49 Now there is escalating discourse within the black community about what it can and must do to stop crime.
50. Just after the new year, Mr. Jackson held the first of several conferences focusing on just that.
51. ‘‘The premier civil-rights issue of this day is youth violence in general and black-on-black violence in particular,’’
52. he has said.
53. His conference also noted
54. the structural conditions that encourage crime—
55. the sorry state of the black economy, high unemployment, poor education and a legacy of racism.
56. ‘‘The black leaders recognize
57. that if they do n’t step out front and engage in the discussion,
58. that basically our young people are turning themselves into slaves,’’
59. says Rep. Washington.
60. Within the black community, there is ‘‘more public concern and debate about the appropriate level of response to increasing crime and violence.’’
61. Many of the black leaders involved in the growing debate retain strong objections to the Senate bill, with its large number of mandatory minimum sentences, death penalties and federalization of local crimes.
62 One of the Senate measures strongly opposed by most members of the Black Caucus has as its author one of its own,
63. Illinois Democratic Sen. Carol Moseley–Braun.

64. Her amendment would restrict prosecutorial discretion
65. — a point opposed by Attorney General Janet Reno
66. — by directing U.S. attorneys to prosecute as adults 13-year-olds charged with committing violent crimes with firearms.
67. The provision would federalize many crimes currently prosecuted by the states.
68. Yet, notes federal Judge Maryanne Trump Barry of Newark, N.J.,
69. who is chairwoman of the criminal law committee of the Judicial Conference of the U.S.,
70. there is no federal juvenile justice system to handle such cases – no federal juvenile prisons, for instance, and no federal youth probation officers.
71. The National Conference of State Legislatures is so opposed to the federalization of state crimes
72. — another provision in the bill, pushed by GOP Sen. Alfonse D'Amato of New York, would federalize all violent handgun crimes—
73. that it recently wrote President Clinton to say
74. ''the Senate bill is inimical to principles of federalism, and we must oppose it.''
75. And a measure that would require states to adopt certain federal sentencing guidelines, such as mandatory minimum sentences, to get federal prison building funds is ''coercive policy,''
76. complains Jon Felde, NCSL's general counsel.
77. There are numerous mandatory minimum provisions in the legislation that Mr. Washington fears could be used in an unfair fashion against blacks who may be charged more harshly than whites for similar acts.
78. And federal judges have ''consistently, vehemently, and virtually unanimously opposed'' mandatory minimum sentences,
79. Judge Barry wrote to Senate Judiciary Committee Chairman Joseph Biden, Democrat of Delaware, in November.
80. Other measures that caucus members say could be used in a discriminatory way are those that would make it a federal crime to conspire to participate in a criminal street gang
81. and that provides the death penalty for drug kingpins
82. even if no death can be shown to have resulted directly from their illegal activity.
83. The Justice Department has warned Congress
84. that it thinks the drug kingpin provision is unconstitutional;
85. the anti-gang measure will also be hit on constitutional grounds in the House.
86. But Sen. Biden insists
87. that the final legislation will include enough significant prevention and punishment provisions
88. that liberals and conservatives alike will be able to endorse it.
89. After all, he says,
90. ''everybody is kind of singing from the same hymnal on the broad strokes.''

Appendix C. Praat pitch settings used in automatic measurements

	Praat default setting	F0max setting	F0min setting
Voicing threshold	0.45	0.6	0.75
Octave cost	0.01	0.01	0.07
Voicing/voiceless cost	0.14	0.14	0.21

Appendix D. Results of paraphrase analysis

Topic #	Boundary size (level 3)	Discourse segment	Topic content	Speakers who mention topic (n = 10)
1			A Senate crime bill, including Jesse Jackson's concerns about it and Senate support for it	9
2	X	19	Bill is too focused on punishment and not enough on prevention; it is a ''lock'em up and throw away the key'' approach	9
3	X	35	Can changes be made to bill to soften opposition to it	3
4	X	49	Black community discussing what it can do to stop crime	2
5	X	56	Quote: Importance for black leaders to address issue of crime	
6	X	80	Other potentially discriminatory measures in bill	1

References

Arnold, J.E., 2008. Reference production: production-internal and addressee-oriented processes. Language and Cognitive Processes 23 (4), 495–527.

Asher, N., Lascarides, A., 2003. Logics of Conversation. Cambridge University Press, Cambridge, UK, xxii + 526pp.

Asher, N., Vieu, L., 2005. Subordinating and coordinating discourse relations. Lingua 115 (4), 591–610.

Auran, C., 2007. Discourse cohesion and its prosodic marking in French: interactions between intonation unit onsets and anaphoric pronouns in speech perception. Paper presented at the ICPhS XVI.

Auran, C., Hirst, D., 2004. Anaphora, Connectives and Resetting: Prosodic and Pragmatic Parameters Interactions in the Marking of Discourse Structure. Paper presented at the Speech Prosody, Nara, Japan.

Blaauw, E., 1994. The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. Speech Communication 14 (4), 359–375, http://dx.doi.org/10.1016/0167-6393(94)90028-0.

Boersma, P., Weenink, D., 2009. Praat: Doing Phonetics by Computer. Retrieved from http://www.praat.org.

Breen, M., Fedorenko, E., Wagner, M., Gibson, E., 2010. Acoustic correlates of information structure. Language and Cognitive Processes 25 (7), 1044–1098.

Couper-Kuhlen, E., 2001. Interactional prosody: high onsets in reason-for-the-call turns. Language and Society 30 (1), 29–53.

Danlos, L., 2010. Strong generative capacity of RST, SDRT and discourse dependency DAGSs. In: Benz, A., Kühnlein, P. (Eds.), Constraints in Discourse. John Benjamins Publishing Co, Amsterdam.

Davidson, J., 1994 January. Blacks' increasing vocal opposition to violence is matched by strong opposition to crime bill. The Wall Street Journal.

den Ouden, H., 2004. Prosodic Realizations of Text Structure. University of Tilburg, Tilburg.

den Ouden, H., Noordman, L., Terken, J., 2009. Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports. Speech Communication 51 (2), 116–129.

Dilley, L., 2010. Pitch range variation in English tonal contrasts: continuous or categorical? Phonetica 67 (1–2), 63–81.

Esser, J., 1988. Comparing Reading and Speaking Intonation. Rodopi, Amsterdam.

Grosz, B., Hirschberg, J., 1992. Some Intonational Characteristics of Discourse Structure. Paper presented at the Proceedings of the 2nd International Conference on Spoken Language Processing, Banff, October.

Grosz, B., Sidner, C., 1986. Attention, intentions, and the structure of discourse. Computational Linguistics 12 (3), 175–204.

Hanson, H.M., 2009. Effects of obstruent consonants on fundamental frequency at vowel onset in English. The Journal of the Acoustical Society of America 125 (1), 425–441, http://dx.doi.org/10.1121/1.3021306, (Research Support, NIH., Extramural).

Herman, R., 2000. Phonetic markers of global discourse structures in English. Journal of Phonetics 28 (4), 466–493.

Hirschberg, J., Grosz, B., 1992. Intonational Features of Local and Global Discourse Structure. Paper presented at the Proceedings of the Speech and Natural Language Workshop.

Hobbs, Jerry R., 1985. On the coherence and structure of discourse (Report No. CSLI-85-37). Stanford University, Center for the Study of Language and Information, Stanford, CA.

Irvine, J., 2001. ''Style'' as distinctiveness: the culture and ideology of linguistic differentiation. In: Eckert, P., Rickford, J. (Eds.), Style and Sociolinguistic Variation. Cambridge University Press, Cambridge, pp. 21–43.

Irvine, J., Gal, S., 2000. Language ideology and linguistic differentiation. In: Kroskrity, P.V. (Ed.), Regimes of Language: Ideologies, Polities, and Identities. School of American Research Press, Santa Fe, pp. 35–84.

Laan, G.P.M., 1997. The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. Speech Communication 22 (1), 43–65.

Lehiste, I., 1975. The phonetic structure of paragraphs. In: Cohen, A., Nooteboom, S.G. (Eds.), Structure and Process in Speech Perception: Proceedings of the Symposium on Dynamic Aspects of Speech Perception. Springer, Berlin/Heidelberg.New York, pp. 195–203.

Lehiste, I., 1982. Some phonetic characteristics of discourse. Studia Linguistica 36 (2), 117–130.

Ljolje, A., 2002. Speech recognition using fundamental frequency and voicing in acoustic modeling. Paper Presented at the Proceedings of ICSLP, Denver, USA.

Lofqvist, A., Baer, T., McGarr, N.S., Story, R.S., 1989. The cricothyroid muscle in voicing control. The Journal of the Acoustical Society of America 85 (3), 1314–1321, (Research Support, U.S. Gov't P.H.S.).

Mann, W.C., Thompson, S.A., 1988. Rhetorical structure theory: toward a functional theory of text organization. Text 8 (3), 243–281.

Mayer, J., Jasinskaja, E., Kölsch, U., 2006. Pitch Range and Pause Duration as Markers of Discourse Hierarchy: Perception Experiments. Paper presented at the Ninth International Conference on Spoken Language Processing, Potsdam, Germany.

Müller, F.E., 1996. Affiliating and Disaffiliating with Continuers: Prosodic Aspects of Recipiency. In: Couper-Kuhlen, E., Selting, M. (Eds.), Prosody in conversation: Interactional studies. Cambridge University Press, Cambridge, pp. 131-176.

Ohde, R.N., 1984. Fundamental frequency as an acoustic correlate of stop consonant voicing. The Journal of the Acoustical Society of America 75 (1), 224–230, (Research support, U.S. Gov't P.H.S.).

Podesva, R., 2006. Phonetic Detail in Sociolinguistic Variation: Its Linguistic Significance and Role in the Construction of Social Meaning. PhD, Stanford, Palo Alto.

Polanyi, L., 1988. A formal model of the structure of discourse. Journal of Pragmatics 12 (5–6), 601–638.

Price, P.J., Ostendorf, M., Shattuck-Hufnagel, S., Fong, C., 1991. The use of prosody in syntactic disambiguation. Journal of the Acoustical Society of America 90 (6), 2956–2970.

Quené, H., van den Bergh, H., 2004. On multi-level modeling of data from repeated measures designs: a tutorial. Speech Communication 43 (1–2), 103–121, http://dx.doi.org/10.1016/j.specom.2004.02.004.

Quené, H., van den Bergh, H., 2008. Examples of mixed-effects modeling with crossed random effects and with binomial data. Journal of Memory and Language 59 (4), 413–425, http://dx.doi.org/10.1016/j.jml.2008.02.002.

Reese, B., Denis, P., Asher, N., Baldridge, J., Hunter, J., 2007. Reference Manual for the Analysis and Annotation of Rhetorical Structure (v 1.0). Technical Report University of Texas, Austin.

Silverman, K.E.A., 1987. The structure and processing of fundamental frequency contours. Unpublished Doctoral Dissertation, University of Cambridge, Cambridge, UK.

Smith, C.L., 2004. Topic transitions and durational prosody in reading aloud: production and modeling. Speech Communication 42 (3–4), 247–270, http://dx.doi.org/10.1016/j.specom.2003.09.004.

Swerts, M., 1997. Prosodic features at discourse boundaries of different strength. The Journal of the Acoustical Society of America 101 (1), 514–521.

Syrdal, A., McGory, J., 2000. Inter-transcriber reliability of ToBI prosodic labeling. Paper presented at the Sixth International Conference on Spoken Language Processing.

Van Kuppevelt, J., 1995. Main structure and side structure in discourse. Linguistics 33 (4), 809–833, (338).

Wichmann, A., 2000. Intonation in Text and Discourse. Pearson Education Limited, Essex.

Yule, G., 1980. Speakers' topics and major paratones. Lingua 52 (1–2), 33–47.