TOWARDS MODELLING PAUSING PATTERNS IN ADULT NARRATIVE SPEECH

by

JEFFREY E. KALLAY

A DISSERTATION

Presented to the Department of Linguistics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2020

DISSERTATION APPROVAL PAGE

Student: Jeffrey E. Kallay

Title: Towards Modelling Pausing Patterns in Adult Narrative Speech

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

Dr. Melissa A. Redford	Chairperson
Dr. Melissa Baese-Berk	Core Member
Dr. Tyler Kendall	Core Member
Dr. Ulrich Mayr	Institutional Representative

and

Kate Mondloch   Interim Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded September 2020

DISSERTATION ABSTRACT

Jeffrey E. Kallay

Doctor of Philosophy

Linguistics

June 2020

Title: Modelling Pausing Patterns in Adult Narrative Speech

The study that is the focus of this dissertation had 2 primary goals: 1) quantify systematic physiological, linguistic and cognitive effects on pausing in narrative speech; 2) formalize a preliminary model of pausing behavior which integrates these effects with those related to individual differences. As a natural consequence of speech-language production, pausing behavior has been of interest to linguistics researchers since at least the mid-20th century. A large number of previous studies have demonstrated systematic effects from factors related to linguistic structure and cognitive processing on speech pausing patterns. Despite that work, a comprehensive understanding of how those factors interact in shaping pausing behavior remains elusive. This is largely due to the highly complex nature of pausing behavior, which also includes more idiosyncratic factors which are difficult to quantify. The current study was aimed at addressing this problem by adopting a more holistic framework.

Three separate within-subject experiments were conducted to alternatively assess effects from respiratory recovery, speech planning, language structure and discourse planning processes on narrative speech pausing behavior. Each of these factors was found to systematically influence both the frequencies and durations of the pauses that were produced. The results also suggested effects on the variability of pause durations, but these effects were primarily limited to the low-level respiratory and speech planning factors. These results together formed the basis of a preliminary model of narrative speech pausing which was tested in the final part of the study. In addition to the experimentally investigated factors, a separate component related to working memory capacity was included in the model to account for the role of individual differences. On the whole, the model was only able to account for a relatively small amount of the variation in the pausing patterns. The language structure component was found to be the most significant contributor. The implications of these results are discussed with reference to the limitations of the preliminary model proposed in the current study. These limitations provide several suggestions on the best directions for future work in attempting to refine the model toward more accurately characterizing narrative speech pausing behavior.

CURRICULUM VITAE

NAME OF AUTHOR: Jeffrey E. Kallay

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
The Ohio State University, Columbus, Ohio

DEGREES AWARDED:

Doctor of Philosophy, Linguistics, 2020, University of Oregon
Bachelor of Arts, Linguistics, 2012, The Ohio State University

AREAS OF SPECIAL INTEREST:

Speech Production
Phonetics

PROFESSIONAL EXPERIENCE:

Graduate Employee, University of Oregon, 2013-2020

PUBLICATIONS:

Kallay, J.E., & Redford, M.A. (2020). Clause-initial AND usage in a cross-sectional and longitudinal corpus of school-age children's narratives. *Journal of Child Language*. 22 pages. Published online April 23, 2020. doi: 10.1017/s0305000920000197

Kallay, J.E., Mayr, U., & Redford, M.A. (2019). Characterizing the coordination of speech production and breathing. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceeding of the 19th International Congress of Phonetic Sciences*, (pp. 1412-1416). Melbourne, Australia.

Kallay, J.E., & Redford, M.A. (2018). Coarticulatory effects on "the" production in child and adult speech. In *Proceedings of the 9th International Conference on Speech Prosody* (pp. 1004-1007). Poznan, Poland.

Kallay, J.E., & Redford, M.A. (2016). A longitudinal study of children's intonation in narrative speech. In *Proceeding of the 17th Annual Conference of* *the International Speech Communication Association*, *INTERSPEECH 2016* (pp. 1079-1083). San Francisco, California.

Kallay, J., & Holliday, J. (2012). Using spectral measures to differentiate Mandarin and Korean sibilant fricatives. In *Proceeding of the 13th Annual Conference of the International Speech Communication Association*, *INTERSPEECH 2012* (pp. 118-121). Portland, Oregon.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my PhD advisor Dr. Melissa Redford for all the support and guidance that she has provided throughout my time at the University of Oregon. The completion of this dissertation would not have been possible without her generous assistance. I would also like to thank the members of my dissertation committee for providing invaluable feedback from the outset of this study through to the final preparation of this manuscript. I am also deeply grateful to Dr. Mary Beckman for being such a great mentor and nurturing my interest in Linguistics during my undergraduate studies at the Ohio State University. Thanks are also due to all of the faculty, staff and students in the Linguistics department at the University of Oregon for providing me with such a welcoming and supportive environment in which to study and work. I cannot imagine completing this dissertation anywhere else. Finally, I would like to extend special thanks to my family for always supporting me in anything that I choose to do.

TABLE OF CONTENTS

Chapter Page

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

I. INTRODUCTION

Speech-language production is a dynamic process which unfolds over time and involves a multitude of influencing factors. As a fundamental consequence of speech-language production, pausing behavior is likewise a complex and multi-faceted phenomenon. Given the level of complexity involved in naturalistic speech pausing patterns, it is perhaps not surprising that despite a wealth of previous research we have yet to arrive at a satisfying and ecologically valid model of the pausing process as a whole. To do so would require simultaneously accounting for influencing factors from a variety of domains. Consider for example the following excerpt from a transcript of an adult speaker telling a spontaneous story (slashes indicate pauses).

> *one day a little boy went to the store to go run errands for his mom / and / that's when he saw a / pair of shoes that he decided that he might want / he proceeded to go home / and try on the shoes / and / he really liked them / and then he wanted to listen to a story so his mom / told him to nestle up on the bed next to her as she read him one of his favorite stories / and then / out of nowhere a spider / came down into her field of view / which frightened her …*

Observing the locations of the pauses relative to the language that was produced in this short excerpt, several conclusions can be drawn. First, it is apparent that the syntactic structure of the language plays an important role in influencing the locations of pauses in the discourse. Of the 13 pauses included in the transcript, a total of 7 of them occurred at clear clausal boundaries. Looking more closely, it is also evident that the relative strength of a given syntactic boundary is relevant to pausing locations. In the 4th pause-delimited utterance in the excerpt, we find that the main clause boundary at the word "shoes" is followed by 2 consecutive relative clauses. The referential identity of the word “that” in each of the relative clauses is dependent on the word “shoes” in the preceding clause, and so the clauses were produced as a single utterance without interruption from a pause. What is also apparent in the above example, however, is that the patterns of pause locations are prone to deviations from the systematic characterization just provided. For example, both the 9th and 12th pauses in the excerpt were produced clause-internally. However, it is important to note that in both cases the pause nonetheless occurred between the subject NP and the following VP. Thus, even when speakers produce within-clause pauses, there is still an apparent tendency to obey the constraints from the syntactic structure of the clause.

Beyond the clear influences from language structure in the transcript, there also appear to be influences from more cognitive-based factors related to speech-language planning. While 9 of the 13 pauses occurred either at clause boundaries or at phrase boundaries within clauses, the 3rd pause occurred within the syntactic NP “a pair of shoes”. This pause likely reflects some type of disturbance in the speech-language planning process. Similarly, the 3 remaining pauses in the transcript occurred immediately following what we might call discourse markers (e.g. *and*, *and then*), which appear to function as placeholders for the speaker and also likely reflect issues with the speech-language planning process. In other words, these pause-delimited discourse markers are serving a purpose much like that of more traditionally defined filled pauses (e.g. *uhh*, *umm*; see Maclay & Osgood, 1959). An alternative possibility is that these sorts of placeholder pauses simply reflect something about personal speaking style rather than cognitive planning effects (see Kendall, 2013; Kolly, Leeman, Boula de Mareuil, & Dellwo, 2015).

As the preceding discussion makes clear, pausing behavior is unquestionably influenced at a minimum by factors related to both the structure of the language and the planning of that language. But as Zellner (1994) has pointed out, “speech motor activity is largely an individual activity, [so] the occurrence of pauses depends to a considerable extent on the specific speaker (p.45).” Indeed, a comparison of this example to one produced by a different speaker performing the same story-telling task is sure to reveal clear and meaningful individual differences in pausing behavior. This additional idiosyncratic property of pausing behavior poses a serious problem if the goal is to formulate a valid model of that behavior. While it is clear that at least some types of linguistic and cognitive factors must be included in a formal representation of pausing, such a behavioral model will remain incomplete absent additional information on how individual differences interact with the more systematic influences on pausing. As such, it is argued in the current study that any serious attempt to model pausing behavior in spontaneous speech must be approached from a more holistic framework than has typically been the case in previous studies. Most importantly, it is crucial for such a model that pausing behavior be studied from a within-subject approach so that influences from individual speaker differences can be incorporated into the overall characterization of the pausing patterns that are observed.

1.1. Current Study

The overarching goal of the current study was to acquire a coherent set of behavioral data which could provide the foundation for a formal model of pausing patterns in narrative speech. To this end, 3 separate experiments were conducted, and each was aimed at investigating the role of separate physiological, linguistic, and cognitive factors on pausing behavior. Experiment 1 was specifically aimed at investigating the relative weighting of low-level anticipatory vs perseveratory effects on pause frequency, duration, and duration variability. Anticipatory effects were assumed to reflect speech-language planning and perseveratory effects were assumed to reflect respiratory recovery. Experiment 2 was specifically aimed at investigating the strength of mid-level language structure effects on pause frequency, duration, and duration variability. These language structure effects were assumed to be indexed by varying levels of local syntactic cohesion between adjacent clauses. Experiment 3 was specifically aimed at investigating the strength of higher-level discourse planning effects on pause frequency, duration, and duration variability. It was assumed that discourse planning effects could be indexed by the relative degree of coherence in narrative speech samples. The final aim was to synthesize the results of the 3 experiments in formulating and testing an exploratory preliminary model of pausing behavior during narrative speech production. Importantly, the goal was to arrive at a model that not only can account for the roles of the various factors which have already been discussed but can also do so in a manner that accounts for individual differences between speakers. For this reason, the entire study employed a within-subjects design in which each speaker participated in all of the speaking tasks.

Following a review of the relevant literature on pausing behavior, the remaining chapters will present and discuss the results obtained in each of the 3 experiments conducted. A preliminary model of pausing behavior will then be proposed based on the experimental results, and results from the testing of this model will be presented and discussed. This will be followed by a discussion of the limitations of the current study as they relate to the pausing model, and some concluding thoughts on suggestions for future work will be presented.

II. BACKGROUND

Researchers have had a longstanding interest in speech pausing that dates to at least the 1950s. In the earliest work, studies were focused primarily on the relationship between speech hesitation and anxiety level in clinical psychology settings (*e.g.*, Kasl & Mahl, 1965; Mahl, 1956; Panek & Martin, 1959; Pope & Siegman, 1962; Siegman & Pope, 1965). However, by the 1960s the study of pausing had already begun to get pulled more firmly into the realm of psycholinguistics. This was due in large part to the work of Goldman-Eisler and her colleagues, whose interest was in the influences of cognitive processes on the temporal and rhythmic structure of speech production (Goldman-Eisler, 1961a, 1961b, 1961c, 1967, 1968; Henderson, Goldman-Eisler, & Skarbek, 1965, 1966; see also Maclay & Osgood, 1959). Since that time, the pace of research in this arena has not subsided, with studies on pausing in both read and spontaneous speech continuing to be conducted to the present day. Despite the wealth of literature that has accumulated over the past several decades, a reliable model of speech pausing behavior remains largely elusive. This is due at least in part to the broadly idiosyncratic nature of the phenomenon. Pausing patterns are well-known to vary randomly across speakers as a function of numerous factors, including both speaker-extrinsic and speaker-intrinsic ones.

2.1. Idiosyncrasy of Pausing

One of the most prevalent sources of variability in pausing patterns is differences in speech situations. For example, is has been previously found that speakers tend to pause significantly more frequently and for longer durations in casual dialogic speech than do speakers in a more formal news reading setting (Gustafson-Capkova & Megyesi, 2001). These differences were argued to “partly be explained by the planning involved in spontaneous speech (*ibid.*, p. 933)"; a process which is considerably less necessary in read speech. Presumably, the difference is more specifically in terms of language planning as opposed to speech-motor planning, suggesting a cognitive basis of at least some of the idiosyncratic variability in pausing patterns. Goldman-Eisler (1961a, 1961b) found similar effects of speech formality on pause durations when comparing casual discussions to psychiatric interviews. She found that speakers produced generally longer pauses in the interviews than in the discussions, and it was argued that this reflected differences in the "intellectual activity" (i.e., degree of verbal planning) of the speech productions. In a more recent study by Strangert (2005) it has even been argued that between-style differences in pausing can reflect differences in the *type* of cognitive processing beyond just verbal planning. She concluded that while within-constituent pauses did "serve to counteract problems caused by the on-line speech processing" in interview speech, these same types of pauses occurred overwhelmingly before "semantically heavy words" in read news speech and served instead to stress the importance of what was being said (p. 3404). Even when looking within a given speaking style, however, pausing patterns have also been shown to vary significantly for the purpose of rhetorical effect. For example, it has been found that participants in a political debate systematically alter the distributions and durations of silences in their speech as a means of expressing assertiveness or otherwise creating a heightened sense of emotion (Braga & Marques, 2004). Similarly, an earlier study by Duez (1982) found that speakers spent a greater amount of time pausing relative to the time spent speaking in political speeches than in either formal or informal interviews. She concluded that the different pattern of pausing in the political speeches reflected stylistic decisions on the part of the speaker; the pauses served to “emphasize ideas and arguments (p. 21).”

The above examples of speaker-extrinsic effects already underscore the large degrees of freedom which are involved in producing the pausing patterns that emerge in natural speech-language production. The problem of developing a reliable characterization of speech pausing behavior becomes even more complicated when one begins to also consider more speaker-intrinsic factors such as mood, emotional state, or even age-related decline in cognitive or respiratory functioning. For example, it has been found that depressed patients tend to produce pauses with significantly shorter and less variable durations following successful treatment for their condition (Mundt, Vogel, Feltner, Lenderking, 2012; see also Mundt, Snyder, Cannizzaro, Chappie, & Geralts, 2007; Pope, Blass, Siegman & Raher, 1970; Szabadi, Bradshaw, & Besson, 1976). Reynolds & Paivio (1968) also found that speakers who are prone to audience-induced anxiety paused significantly more frequently during a public speaking task than did those without such anxiety. Yet another study observed that pause durations vary systematically between positive or negative emotional states of the speaker (Schroder, Cowie, Douglas-Cowie, Westerdijk, & Gielen, 2001). Beyond differences in mood and emotional state, effects of normal aging and age-related disease can also contribute to systematic differences in the temporal structure of speech production. For example, older adults tend to produce longer pauses than younger adults and this is likely reflective of natural age-related decline in cognitive processing speed (Cooper, 1990); a reasonable conclusion given the numerous studies demonstrating similar effects from declining processing speed outside the realm of language (e.g., Clay et al., 2009; Manard, Carabin, Jaspar, & Collette, 2014; Salami, Eriksson, Nilsson, & Nyberg, 2012). Likewise, Parkinson's patients have been found to produce shorter utterances than their healthy peers as a result of reductions in cognitive functioning (Huber & Darling, 2011) as well as respiratory functioning (Solomon & Hixon, 1993). Similar effects of cognitive impairment on pause frequency and/or duration have also been found in patients with a variety of other age-related disorders such as ALS (e.g., Yunusova et al., 2016), Alzheimer's disease (e.g., Gayraud, Lee, & Barkat-Defradas, 2011), and other forms of dementia (e.g., Yunusova et al., 2016). To complicate matters even further, there are also more unquantifiable effects related to individual differences in personal speaking style which must be considered on top of any of the more quantifiable ones. Taken together, this multitude of speaker-intrinsic factors that contribute to the temporal patterns of speech production presents a serious challenge in arriving at a general and reliable characterization of overall pausing behavior in natural speech situations.

Despite the seemingly random nature of pausing behavior, however, the patterns have also been found to vary much more systematically as a function of various physiological, linguistic, and cognitive factors. These factors are the primary experimental focus of the current study. More specifically, the experiments conducted were aimed at quantifying effects related to various aspects of linguistic structure as well as lower-level speech planning and respiratory processes.

2.2. Features of Pausing Patterns and their Affecting Factors

The patterns of pausing that surface during natural speech-language production can be broadly characterized in terms of the following 3 features: pause location (or likelihood), pause frequency and pause duration. Pause location refers to the different types of locations within the speech-language stream where pauses are more or less likely to occur. These locations are typically defined in terms of syntactic or discourse structure. Pause frequency simply refers to the rate of occurrence of pauses relative to the overall amount of speech-language produced. For example, frequency can be measured in terms of the number of words produced per pause or the duration of speaking time between pauses. Finally, pause duration simply refers to the amount of time that a given instance of pausing lasts before speech-language production resumes.

In the current study, a framework was adopted which assumes that the various factors which systematically affect each pausing feature can be broadly classified along a continuum from low-level to high-level processes. At the lowest level are factors related to speech-motor planning of an already structured phrase or sentence and respiratory recovery. Speech planning can be characterized as a low-level process in the sense that articulation emerges largely as the result of experience through the repetitive practice of speech production (see Smith, 2006; Smith & Zelaznick, 2004). In this sense, speech planning is a largely subconscious process which unfolds more or less automatically. Similarly, respiratory recovery is an inherent consequence of the tension between the constant airflow required for speech production and the instinctive physiological drive to inspire. To be more precise, “the probability of occurrence of a pause during continuous speech seems to depend on the amount of residual air in the lungs” (Vaissiere, 1983). Thus, while it is true that the breathing cycle is more highly modulated during speech production as compared to normal quiet breathing (Hoit & Lohmeier, 2000; Maclarnon & Hewitt, 2004), inhalation remains a physiological necessity and respiratory recovery is an essentially automatic process.

Pausing patterns are also affected by language planning processes related to linguistic cohesion. In very general terms, linguistic cohesion has been defined as "the degree to which the propositions… within a narrative are linguistically connected (Shapiro & Hudson, 1991, p. 960)." This factor operates at the local micro-structure level of discourse and is not concerned with larger units such as topics and themes. Linguistic cohesion is characterized as a mid-level process in this dissertation because it concerns the conceptual/semantic content of the discourse, but only in terms of the sequential relatedness of successive units (e.g., clauses, sentences). It requires not only that the speaker track the information flow of the discourse as is being produced, but also that they make use of the appropriate linguistic devices (including pauses) for signaling the local structure.

Finally, pausing patterns are also affected by higher-level discourse planning processes related to linguistic coherence. Coherence differs from cohesion in that it deals with the more global organization of themes or goals (see Kintsch & van Dijk, 1978; van Dijk, 1982). As defined by Kintsch & van Dijk (1978), "a discourse is coherent only if its respective sentences and propositions are connected, and if these propositions are organized globally at the macrostructure level (p. 365)." As such, the linguistic coherence of a discourse is characterized by the nature of relationships between units at various hierarchical levels of the discourse structure as opposed to the flatter sequential ones that are the purview of linguistic cohesion. This requires a higher level of cognitive processing on the part of the speaker as they must not only store and recall the various themes in working memory as the discourse progresses, but also to make use of an even wider array of linguistic devices to signal the relatedness of discourse units at various levels of the discourse structure.

2.3. Structural Effects on Pause Location

Of course, speech and language production itself is a complex and dynamic endeavor, and so there are many other factors which can act to constrain the systematicity of pause locations in spontaneous speech. For example, while speakers do generally favor strong syntactic boundaries as pause locations, they are at the same time sensitive to the position of those boundaries relative to the overall length of the utterance. That is to say, it has been demonstrated that speakers prefer to produce roughly symmetrical “performance units,” such that it is not unusual for a speaker to disprefer a strong boundary in favor of subsequent weaker one if doing so would more faithfully preserve the “performance structure” of the utterance (see Gee & Grosjean, 1983; Grosjean, Grosjean, & Lane, 1979). The notion of “performance structure” relates to the observation that speakers appear to structure their speech productions in a way that is extraneous to the syntactic structure. Additionally, spontaneous speech production is naturally characterized by alternating periods of relative fluency and disfluency, with the latter argued to correspond with “periods of planning and cognition (Henderson et al., 1965, p. 237).” Thus, in spontaneous speech production the patterns of pausing are expected to be at times more or less systematic due to cognitive limitations in the amount of language that can planned at one time. Pauses will be less strictly constrained by the relative strength of syntactic and discourse boundaries in the case of the latter.

2.4. Low-Level Planning and Recovery Effects on Pause Duration

There have been several previous studies aimed at characterizing the relationship between speech planning and pause durations. For example, Ferreira (1991) investigated the effects that length and syntactic complexity had on the amount of time it took speakers to begin repeating memorized sentences in response to a question. She found that it took speakers about 100 milliseconds longer on average to initiate production of longer sentences than shorter ones. Moreover, when the sentence was long the initiation times increased with increasing syntactic complexity of the sentence. She interpreted her findings as suggesting that "the sentence production system attempts to produce a structurally defined unit (p. 217)"; larger and more complex units take longer to plan. Zvonik and Cummins (2003) later conducted a similar study using connected read speech samples which allowed for a direct assessment of spontaneous planning effects on pause durations. They found that speakers took longer pauses before producing longer intonational phrases (IP), although the effect virtually disappeared once the length of the IP surpassed about 10 syllables. This finding led them to conclude that "pauses less than (approximately) 300 ms. may be special (p. 779)" since they never occurred before longer utterances. Taken together, these studies at least provide some preliminary evidence that speech planning systematically affects the durations of inter-utterance pauses.

In addition to speech planning, researchers have been interested in the role that speech breathing plays in determining pause durations. As in Ferreira's (1991) study, Whalen and Kinsella-Shaw (1997) were interested in the effect that the length of a sentence had on the amount of time it took to initiate its production. However, they were more specifically interested in the relationship between respiratory inspiration and that initiation time. Asking participants to fully expire to a set level and then begin reading a sentence upon receiving a “go” signal, they found that speakers tended to take longer breaths before saying longer sentences. This suggests that the speakers anticipated that the production of longer sentences would require a larger supply of air, and that they planned their breath intakes accordingly. In turn, longer breaths are logically expected to result in longer pauses. But as the authors themselves state, their “paradigm was not particularly natural, since it is seldom the case that speakers will spontaneously decide to exhale to a low volume and then utter a single sentence (p. 148).” This situation was remedied to a degree in a more recent study (Fuchs, Petrone, Krivokapic, & Hoole, 2013). Instead of asking speakers to utter sentences in isolation, the stimuli used in this study included a preceding carrier sentence of sufficient length to induce inter-sentence breaths. In this way, any breath planning had to take place more online as the speech stream was unfolding. The speakers in their study had to balance the competing processes of continuous and fluent speech production versus taking a large enough breath to complete the second sentence. Their results not only showed that pauses were generally longer before longer sentences, but more importantly that speakers took longer *and* deeper breaths before longer sentences. These findings suggest that the longer pauses resulted not so much from articulation planning per se, but rather they reflected some amount of speech breath planning.

The aforementioned studies provide compelling evidence that pause duration is a reliable acoustic correlate of speech-motor planning. However, it is equally plausible that the pause durations will vary just as systematically as a function of respiratory recovery.

Since the residual volume of air in the lungs naturally decreases across the span of continuous speech production, there is an expectation that the length of the preceding utterance would also be predictive of breath pause likelihood. Given the demonstrated relationship between breath duration and pause duration (e.g., Fuchs et al., 2013), breath pauses would in turn be expected to be longer than non-breath pauses. Unfortunately, the experimental designs used in previous speech planning studies prevented investigation of this recovery effect. A relationship between pause duration and preceding utterance duration has been suggested, however. Zvonik and Cummins (2003) also analyzed the effect of preceding IP length on pause durations and found the same effect as with the following length; pauses were longer after longer utterances. In fact, the recovery effect was found to be even stronger than the planning effect in their study, suggesting that the former provides an even stronger constraint on pausing behavior. However, since they did not conduct additional analyses on the presence or duration of breaths during pauses, a meaningful relationship between preceding utterance length and speech breathing can only be inferred from their results.

2.5. Mid-Level Linguistic Cohesion Effects on Pause Likelihood and Duration

In terms of linguistic cohesion, the relative likelihood of a pause at a given boundary between successive language units (e.g., clauses, sentences) can vary systematically in correspondence with the strength of that boundary. For instance, it has often been observed that speakers are sensitive to and largely obey the hierarchical syntactic structure of language such that there is a strong preference for placing pauses at relatively stronger syntactic boundaries than at weaker ones. In one study, Bailly & Gouvernayre (2012) analyzed recordings of French audiobook readings and found that the likelihood of pausing at the strong sentence-final boundaries was significantly higher than at any of the weaker sentence-internal boundaries. However, the speakers still produced a large number of sentence-internal pauses, and these occurred most often before prepositional phrases. It was also not uncommon for pauses to occur immediately before coordinating conjunctions and verb phrases, but they rarely occurred at any of the other weaker sentence-internal boundaries. In another study, Ferreira (1991) asked English-speaking participants to read SVO sentences which all contained the same number of phonological words but differed in the relative complexities of their subjects and objects. Among other things, she found that speakers were more likely to pause sentence-internally with increasing complexity of the object, but importantly those pauses occurred overwhelmingly between the subject and the verb rather than between the verb and the object. Thus, her speakers were also sensitive to the hierarchical syntactic structure of the sentence even when pausing at relatively weaker sentence-internal boundaries. In addition to the read speech that was used in the aforementioned studies, similar patterns have also been found in more spontaneous speech samples. For example, van Donzel and Koopmans-van Beinum (1996) found that Dutch speakers paused overwhelmingly at paragraph boundaries (80-100%) during a storytelling task, followed by clause boundaries (~56%) and then other within-clause word boundaries (~15%). Numerous other studies have largely supported the significant role that hierarchical syntactic structure and relative syntactic boundary strength plays in determining where pauses are likely to occur (e.g., Brotherton, 1979; Cooper & Paccia-Cooper, 1980; Grosjean et al., 1979; Hawkins, 1971; Snedeker & Trueswell, 2003; Strangert, 2004). Importantly, the aforementioned studies find remarkably similar patterns of pause locations relative to syntactic structure in both read speech (single sentence and connected) and spontaneous speech, suggesting that syntactic constraints on pausing behavior during speech and language production are robust and psychologically valid.

Given that pause durations have been shown to reliably signal the information structure of discourse (see Swerts & Geluykens, 1994; Swerts, 1997), we might also expect pause durations to vary systematically as a function of the relative cohesion between discourse units. In fact, in a study of read news speech it was found that sentence boundaries were generally accompanied by the longest pauses, followed by major clause boundaries and then subordinate clause boundaries (Fant, Kruckenberg & Ferreira, 2003). Bailly & Gouvernayre (2012) found a similar pattern in an analysis of professional audiobook reading. The speakers generally produced the longest pauses at paragraph boundaries, followed in turn by those at sentence-boundaries and then sentence-internally. While the pattern of decreasing pause duration with decreasing syntactic boundary strength that was found in these studies fits with expectations, it is not clear to what extent their findings are ecologically valid given the nature of the speech samples that were analyzed. During spontaneous speech production, the speaker is not provided with the same type of speech-external structural cues (e,g., punctuation) as are available in read connected speech. In fact, Strangert (2004) found that there were no systematic differences between clause-final and clause-internal pause durations in spontaneous interview speech. Contrary to Strangert’s finding, an earlier study on pausing in spontaneous speech found that the distributions of pause durations did indeed vary according to the strengths of the syntactic boundaries at which they occurred (Goldman-Eisler, 1972). The pauses were characterized in terms of being “fluent” (< 250 milliseconds), and it was found that the proportion of fluent transitions increased systematically from boundaries at sentences, coordinate clauses, subordinate clauses, and relative clauses respectively. Furthermore, approximately 65% of the pauses at sentence boundaries were at least 1 second in duration while the majority of pauses at clause boundaries were less than 750 milliseconds in duration. Hirschberg & Nakatani (1996) similarly found pause durations to be longer at sentence boundaries than at other non-sentence boundaries in spontaneous speech. Synthesizing the findings from these various studies, there is a suggestion that syntactic structure does constrain pause duration, but also that the syntactic constraint is perhaps not as robust with pause durations as found with pause likelihoods (see also Kendall, 2013). The role of syntactic structure in determining pause duration appears to be context and meaning dependent.

2.6. Higher-Level Linguistic Coherence and Cognitive Processing Effects on Pause Frequency and Duration

Pause locations have also been found to vary systematically as a function of the global discourse/information structure. Regarding discourse structure, van Dijk & Kintsch (1983) have broadly defined coherence as follows:

> “a coherent structure is built on the basis of the pattern of argument repetition among the semantic units in the text. Semantic units are added level by level… to the fragment of the already existing textbase still available in short term memory (p. 44).”

Crucial to this definition is the notion of a hierarchical network of discourse units, as well as the role of memory in maintaining (and appropriately signaling linguistically) the relations that hold between units at various levels. For as they note in an earlier paper, “the [memory] buffer must contain the information about macropropositions and presuppositions that is required to establish the global coherence of the discourse

(Kintsch & van Dijk, 1978, p. 370).” Thus, discourse coherence operates at the level of large and stratified conceptual structures whose navigation imposes considerable demands upon higher-level cognitive processes.

The hierarchical nature of discourse structure means that boundary strengths between successive discourse units are relative to their position in the structure; stronger boundaries exist at higher levels (e.g. topic boundaries). Utilizing perceptual judgments of discourse boundaries, van Donzel (1999) found that pauses accompanied 100% of the boundaries that were perceptually judged as *extra strong* in spontaneous speech samples, while only 89% of perceptually *strong* boundaries and 46% of perceptually *weak* boundaries were marked with a pause. Of course, it is possible that the perceivers based their judgments of boundary strengths at least partly on the pauses themselves, leaving uncertain the relationship between pausing and discourse boundaries during speech production. In other words, it is not clear from their results whether stronger boundaries are more likely to be accompanied by a pause or alternatively whether the occurrence of a pause is more likely to result in the perception of a strong boundary. To this point, an earlier study by Swerts & Geluykens (1994) explicitly noted the “danger of circularity” in defining discourse structure in terms of prosody. Approaching the problem more from an information structure framework, they instead investigated the realization of prosodic features at well-defined “topical unit” boundaries in instructional monologues (i.e., speech which facilitated topic boundary marking irrespective of prosody). They found that the speakers *always* paused at the transitions between “topical units”, and that they were more likely than not to pause immediately after a topic-introducing phrase or clause. It is also important to note that the speakers did occasionally pause at “shallow structural breaks, e.g. in between a preposition and a noun (p. 33)" and did not always pause at strong syntactic boundaries such as after major clause breaks. These findings led them to the conclusion that "it would be a mistake to interpret pausal structure purely in syntactic terms (ibid.)" and absent consideration of the higher-level information structure of the discourse. To this point, van Donzel & Koopmans-van Beinum (1996) also found that it was not uncommon for speakers to pause immediately following discourse markers or other connectives during storytelling, which further suggests that discourse structure operates independently from syntactic structure in determining where speakers are likely to pause. Similar findings on the role of discourse structure in determining pause locations have been reported in numerous other studies (e.g., Gustafson-Capkova & Megyesi, 2002; Megyesi & Gustafson-Capkova, 2001; Yang, 2004).

As with the previously discussed effects of local linguistic cohesion on pause durations, it is also reasonable to assume that the patterns of relationship between higher-level discourse boundaries and pause occurrence corresponds also with systematic differences in pause durations at the various boundaries. In the aforementioned study by Swerts (1997), a direct relationship was found between the durations of pauses and the relative strengths of discourse boundaries at which those pauses occurred. Whereas boundaries with pauses of less than 1 second in duration were almost never judged as representing a discourse boundary, the degree of agreement on the presence of a discourse boundary increased systematically with each increase of 1 second in pause durations at the boundaries. This finding suggests that, just as with more local syntactic-based boundaries, the global and hierarchical structure of the discourse plays a meaningful role in constraining the durations of pauses.

2.7. Overview of the Study

In the following chapters, the results of 3 primary experiments will be presented and discussed. These 3 experiments follow from the general framework as laid out in the preceding discussion: Experiment 1 was designed to test for the low-level speech planning and respiratory recovery effects on pausing; Experiment 2 was aimed at investigating the mid-level linguistic cohesion effects on pausing; and Experiment 3 the higher-level linguistic coherence effects on pausing. In each of the experiments, the dependent variables (DVs) of interest were pause frequency, pause duration, and pause duration variability. As discussed above, previous research has found pause frequency to systematically vary as a function of both the local linguistic structure and cognitive processing demands. Pause durations have also been shown to vary according to those same factors, as well as lower-level recovery and planning processes. Pause duration variability has received much less focus in previous work, but it was included in each of the experimental analyses to study the extent to which the effects of the experimental factors on durations reflect the relative systematicity of the functions of the pauses themselves. For example, the occurrence of breath pauses is systematic since they serve the physiologically necessary function of replenishing the air supply needed to speak. By contrast, non-breath pauses may reflect any number of different functions such as planning difficulties, personal stylistic choices, etc. The prediction was that greater functional systematicity would result in less variable durations.

In Experiments 1 and 2, frequency of pausing was measured in terms of whether or not a pause occurred between pairs of sentences. The effects of experimental condition and speaker were then assessed using multiple logistic regression models. For all other analyses, the effects of experimental condition and speaker were assessed using cross factorial ANOVA models. Given the increasing popularity of the more complex mixed effects model analyses in the field of experimental linguistics, I address the choice to use the relatively simpler ANOVA analyses in the current study before proceeding.

Crossed factorial ANOVA analyses were deemed appropriate primarily given the experimental design of the current study (see Burnham, 2015; NIST/SEMATECH, 2020). In all 3 of the experiments conducted, the levels of each experimental condition were fully crossed with those of the speaker variable, and the outcome variables of interested were all quantitative non-zero measures. Additionally, there was a concern that the structural simplicity of the analyses coupled with overly complex models would lead to high variance of the parameter estimations, yielding results that were not generalizable to other population samples (see Clark, 2013). Thus, it was preferable to adopt a more conservative approach and err on the side of underfitting as opposed to overfitting the data. Having said this, it should also be mentioned that initial analyses were also conducted using more complex mixed effects models and no meaningful differences were found when compared to the simpler analyses presented below.

Following the chapters with experimental results, a preliminary model of pausing behavior for adult narrative speech is proposed and preliminary modelling results are presented and discussed. The model is based on Moderated Mediation Analysis. This type of analysis assumes that there is some variable which has a direct effect on the outcome, as well as some other variable(s) which act(s) indirectly to mediate the direct effect and a separate variable which further moderates the relationship between those predictors. To foreshadow a bit the proposal, the basic idea is that clause boundaries provide the opportunity for pausing, but that taking advantage of this opportunity is mediated by planning process. In addition, since speech and language production is an individual activity, then the proposal is that the relationship between planning and boundaries is itself moderated by speaker-level differences. It must be emphasized that the modelling results presented herein are preliminary, and so simply lay the groundwork for a larger-scale project. That project will need to overcome several limitations of the current study, as discussed at the conclusion of this thesis.

III. EXPERIMENT 1: LOW-LEVEL ANTICIPATORY AND RECOVERY EFFECTS

Experiment 1 of the current study had the general aim of quantifying the effects of speech planning and physiological recovery on pause frequency, pause duration, and the variability of pause durations. The planning effects were investigated in relation to the length of the utterance that followed a pause while recovery effects were investigated relative to the length of the utterance that preceded a pause. As discussed in the preceding chapter, there is a large body of evidence in the previous literature that speech planning exerts considerable influence over both the frequency and duration of pauses in read and spontaneous speech. Speakers are more likely to pause and tend to pause for longer prior to the production of longer utterances. This effect also appears to operate independent of the linguistic content of the utterances. A smaller number of previous studies have also suggested effects of respiratory recovery such that the production of longer utterances leads to increased likelihood of pauses which are also longer in duration. What appears to be missing from the existing literature is a careful examination of the planning and recovery effects with respect to each other. The current experiment was intended to address this gap. The specific aims were threefold. The first aim was to conduct a more rigorous investigation of the recovery effect on pausing than has been reported in previous literature. The second aim was to assess whether the well-documented planning effect hold in the presence of a recovery effect. The third aim was to assess the degree to which the relative strengths of each of these effects might interact with the other.

Assuming that speech planning has a robust effect on pausing behavior, the hypotheses for the current experiment were as follows. First, it might be predicted that the frequency of pausing at sentence boundaries would vary as a function of the relative lengths of the upcoming sentences; longer sentences contain more speech units to be planned, leading to increased likelihood of pausing in service of planning. On the other hand, the stimuli were designed such that the there was a strong conceptual boundary between sentences, and so an alternative prediction would be that the length on the upcoming sentence would not have an effect on pause frequency. In terms of the durations of pauses, the prediction was that they would be longer before longer sentences as a reflection of the increased time required for the planning of a greater number of speech units (see Ferreira, 1991; Krivokapic, 2007). There are less suggestions from previous literature on the predictions for planning effects on the variability of pause durations, but it was assumed in the current study that pause duration variability is at least partly influenced by the functional systematicity of the individual pauses (see section 2.2.7.). Pauses before longer sentences are presumed to be more systematically linked with speech planning than pauses before shorter sentences, and so the prediction was that the variability of pause durations would decrease with increasing length of the upcoming sentence.

Assuming that respiratory recovery has an effect on speech pausing behavior, the prediction was that pause frequency would increase with increasing length of the preceding sentence as a function of greater need to replenish the air supply in the lungs (see Kallay, Mayr, & Redford, 2019). Respiratory intake is a process requiring time to complete, so breath pauses are expected to be longer in duration that non-breath pauses (see Trouvain, Fauth, & Mobius, 2016). Thus, it was also predicted that the durations of pauses would increase with increasing length of the preceding sentence. Finally, breath pauses are more functionally systematic that non-breath pauses, so it was predicted that the variability of pause durations would decrease with increasing length of the preceding sentence.

Each of these predictions were tested using an experimental procedure in which speakers were asked to produce pairs of sentences off a computer screen. Importantly, the sentences in each pair were not semantically related and they were constructed with identical syntactic structures. The sentences varied only in terms of their relative lengths. This design allowed for a direct assessment of the planning and recovery effects without the confounding effects of language content or structure that have been present in previous studies. The elicited speech was analyzed in terms of the relationships of both the preceding and following sentence lengths to the characteristics of pauses that occurred between the sentences. The degree of interaction between the sentences lengths was also investigated.

3.1. Methods

3.1.1 Participants

A total of 43 college-aged adults were recruited through the Human Subjects Pool at the University of Oregon for participation in the current study. Three of the subjects were excluded from all of the analyses due to failure to follow instructions during their study sessions. Of the remaining 40 participants who successfully completed the study, 11 were male and 29 were female. All participants were native speakers of American English, and all had a history of typical speech-language development as self-reported on a background questionnaire. Most of the participants self-identified as being White (N = 33), while the remaining 7 participants identified as being either Asian, African

American, White Hispanic, or Multiracial. Each participant completed all of the different speaking and assessment tasks during a single study session.

3.1.2. Study Sessions

All study sessions took place in a quiet study room at the Spoken Language Research Laboratories (SLRL) located at the University of Oregon. At the outset of the study session the experimenter described the types of tasks that the participant would be asked to complete and informed them their rights as a voluntary research participant. They were asked to provide written consent for their participation in the study and to fill out a brief background information survey. The background information included basic demographics (age, ethnicity, gender, etc.), as well as information related to current or previous medical conditions (e.g. asthma, cold, etc.) which may have potentially impacted their performance in the speaking tasks. They were also asked to report whether they had ever received speech-language therapy.

At the outset of each study session, participants were asked to perform a standard spirometry test of forced vital capacity (FVC) followed by 1 of 2 working memory capacity (WMC) assessment tasks. These WMC tasks discussed at some length in Chapter 7. The spirometry test was repeated 3 more times at evenly spaced intervals throughout the session, and the entire session concluded with the 2nd of 2 WMC assessment tasks. The results of the spirometry test of FVC were not analyzed in the current study. Following the first WMC assessment task they were provided with instructions for the first speech production task which forms the basis of the analyses in Experiment 1 of the current study. The experimenter left the room while the participant completed the first speaking task on his/her own. At the conclusion of the first speaking task they were instructed on the computer screen to let the experimenter know that they had finished. At that time the experimenter returned and provided instructions for the speech production task conducted in Experiment 2 of the current study, followed similarly for Experiment 3. A fourth speaking task was also conducted in which participants were asked to provide spontaneous directions to several locations around the university campus, but no analyses were performed with that data for the purpose of the current study. Following Experiment 1 of the study, the experimenter remained in the room with the participant for the remainder of the session.

For the entirety of the study session participants sat at a table situated on one side of the room while the experimenter sat across from them on the other side of the room. A Dell laptop computer was placed on the table directly in front of the participant and was used for the presentation of stimulus items during all 3 of the primary speech production tasks. The participants' speech productions were recorded with a Shure SM81-LC cardioid condenser microphone connected to a Marantz PMD660 digital audio recorder. The microphone was boom-mounted and placed at an angle of approximately 45 degrees toward the participants mouth so as not to obscure their view of the computer screen. While the participants were free to move around in their seats naturally, the microphone was placed close enough so that it was typically within 12 inches of their mouths in order to ensure high quality recordings for manual segmentation based on the acoustic signal.

3.1.3. Materials

The stimuli for Experiment 1 consisted of pairs of meaningful but semantically unrelated sentences. The sentences in each pair were separated by strong syntactic and conceptual boundaries at which pauses are expected to occur (see van Donzel, 1999; van

Donzel & Koopmans-van Beinum, 1996; Oliveira, 2002; Swerts & Geluykens, 1994). The sentence lengths varied between short (= 12 syllables) and long (= 24 syllables). While previous studies have typically found no effects of syntactic complexity on pausing behavior (see, e.g., Fuchs et al., 2013; Krivokapic, 2007), the syntactic structures of the sentences were also restricted to ditransitive constructions. More specifically, each sentence was of the following form: [SUBJECT]-[VERB$_{\mathrm{PAST}}$]-[INDIRECT OBJECT]-[DIRECT OBJECT]. The verbs used in the sentences were as follows: *bought*, *brought*, *built*, *gave*, *made*, *read*, *sent*, *showed*. A total of 16 sentences were constructed (8 short and 8 long), with each verb used for both 1 short and 1 long sentence. The words used in the sentences were limited to those found in the top 1000 in frequency in the Corpus of Contemporary American English (COCA: Davies, 2008-present). This was meant to control for observed word frequency effects; pauses tend to be more frequent and longer before lower frequency words (see Beattie & Butterworth, 1979; de Jong, 2016). Except for the verbs, no other content words were repeated across sentences.

Taking all possible combinations of short (S) and long (L) sentences yields the 4 distinct pair types that were used: SS, SL, LS, LL. The pair orders for each participant were randomized with a partial randomized blocking technique, which is described next. First, all possible pairs of sentences for each pair type were constructed. Within each block, 2 pairs of each pair type were then randomly sampled without replacement, providing 8 total pairs per block. The order of the 8 pairs within a block were then randomized before being presented. Each participant was presented with 5 consecutive blocks of 8 sentences, for a total of 40 sentence pairs. The total number of possible pairs (N = 296) was greater than the 40 that were presented, meaning that each participant was given a unique subset of the pairs. The following are examples of each pair type and the full set of sentences can be found in Appendix A:

SS: *The science teacher gave her students directions.*

Her husband made all of us some materials.

SL: *The woman showed her family the new picture.*

The respected institution read their likely employees the economic analysis.

LS: *The important medical doctor brought the public hospital some much needed experience.*

The officer read somebody a good story.

LL: *The serious individual showed the popular manager several new solutions*.

The American expert gave the foreign official a significant opportunity.

3.1.4. Procedures

The sentence pairs were presented one at a time on a computer screen with a program that was written using the Expyriment library (Krause & Lindemann, 2014) in Python. The 2 sentences were shown on separate lines on the screen, with a line of space between them. The pairs were presented in white text on a black background with a font size of 48. They were also centered on the screen both vertically and horizontally.

There were 3 separate phases of stimuli presentation per pair: *read*, *exhale*, *say*. Prior to the initial presentation of each pair, a visual timer would countdown from 3 to 1 on the screen. Following the countdown, the pair would be displayed along with the text “READ:” at the top of the screen. The participant was instructed to read the sentences to themselves at this time. This was intended to familiarize them with what they would be producing with the goal of creating more of a speaking task and less of a reading task.

The pairs stayed on the screen for 8 seconds during the *read* phase before being replaced with the *exhale* phase. At this time the screen would read “Exhale completely. Then click to read sentences.” The exhalation phase was meant to require that the participant fully inspire prior to each production, and the goal was to control for unwanted effects of variable initial volumes of air on the pause durations. Upon exhaling and clicking the button, the pair of sentences would reappear on the screen exactly as before except that the text “SAY:” was displayed at the top of the screen instead of “READ:”. Once the sentences were produced participants were instructed on the screen to click the button again to move on to the next pair. In addition to explaining the procedures to each participant prior to beginning, the following text was displayed on the screen at the start of the task:

In this experiment you will read and say sentence pairs.

After a countdown from 3, a sentence pair will be shown on the screen.

The pair will remain for 8 seconds for you to read it to yourself.

You will then be asked to exhale completely and then click to continue.

The pair will return, and you will be asked to read it aloud.

If you are ready to begin, click the mouse.

For this task, the experimenter left the room as soon as the participant indicated that they understood the task directions, and they were left to complete the task by themselves. Upon completion of the task they were instructed on the screen to notify the experimenter that they had finished.

3.1.5. Segmentation & Labelling

The onsets and offsets of all sentence pairs produced in Experiment 1 of the study were marked by trained researchers based on both the visible and audible acoustic signals using Praat software (Boersma & Weenink, 2017). All sentences were constructed to be sonorant-initial and -final, so no specific criteria for marking stop closure boundaries were required. For utterance-initial or utterance-final vowels, nasals, and liquids, onsets and offsets were marked at the point where the second formant became visible or disappeared, respectively. In the case of utterance-final fricatives, it was common for there to be a clear portion of spectral energy which was followed by some residual energy in a narrower band of frequencies. The boundaries of these were marked at the point where the preponderance of the spectral energy appeared to end and there was a visible decrease in amplitude.

The onset and offset boundaries of pauses were also marked according to the same criteria as above. Pauses were defined as any period of silence between the offset of the first sentence and the onset of the second sentence that was at least 50 milliseconds in duration. In early studies on pausing it was common to adopt a much higher threshold in classifying pauses, as it was argued that silences of less than 250 milliseconds represented “articulatory pauses” rather than “hesitation pauses proper” (Goldman-Eisler, 1961, pp. 232-233; see also Goldman-Eisler, 1968). In more recent studies the minimum duration required for a silence to be classified a pause has been more variable, but it is now common to use thresholds of less than the previously favored 250 milliseconds. More recent studies have alternatively adopted cutoffs of 200 milliseconds (e.g., Grosjean & Collins, 1979; Krivokapic, 2007), 150 milliseconds (e.g., van Donzel & Koopmans-van Beinum, 1996; Koopmans-van Beinum & van Donzel, 1996), and 100 milliseconds (e.g., Butcher, 1981; Trouvain & Grice, 1999). The rationale for using such a low threshold for pause classification in the current study follows from the aim of the experiment: to quantify the effects of surrounding sentence lengths on the relative degree of juncture between the sentences. The goal was not to categorize silences as pauses or not pauses. It should also be noted that using a shorter cutoff is not without precedent (e.g., Kendall (2013) used 60 milliseconds), and it has been found that pauses of less than 100 milliseconds in duration are psychologically valid to listeners in certain contexts (Butcher, 1981). Other studies have even classified any silent portion greater than 0 seconds in duration as a pause (e.g., Ferreira 1991, 1993)

3.1.6. Measures & Analyses

There were 3 dependent variables of interest in Experiment 1 of the study: pause frequency, pause duration, and pause duration variability. Pause frequency was measured simply in terms of the relative proportions of pairs which did or did not contain an inter-sentential pause. Pause durations for each inter-sentence pause were extracted directly from the segmented audio files. Pause duration variability was measured as the Coefficient of Variation of pause durations (CV_{dur}), which was calculated as the standard deviation of the durations divided by the mean duration within-condition for each speaker. The primary independent variables of interest in each of the analyses were the lengths of the preceding and following sentences as a well as their interaction (i.e. pair type). The main effect of speaker was also included in each of the models to assess the extent to which the variability in pausing behavior could be explained by individual differences independent of the experimental manipulation.

The effects on pause frequency were assessed by first fitting a logistic regression model with the aforementioned independent variables as predictors of the presence/absence of inter-sentence pauses. The statistical significance of each of the predictors in the logistic regression model were then tested using binomial ANOVA analysis. The effects on pause duration and CV_{dur} were assessed using two-way crossed factorial ANOVA models with effects structures that mirrored the logistic regression model for pause frequency. McFadden's pseudo-R^2 is reported as a measure of explained variability in the logistic regression model predicting pause frequency. A pseudo-R^2 value between .2 and .4 is generally interpreted as indicating a strong model fit (see McFadden, 1974, 1978). Partial-ω^2 is reported as a standardized effect size measure for all statistically significant effects on pause duration and CV_{dur}.

3.2. Results

3.2.1. Pause Frequency

There were significant main effects of both the preceding and following sentence lengths on the frequency of inter-sentence pausing [Preceding: $\chi^2(1,1556) = 13.242$, $p < .001$; Following: $\chi^2(1,1555) = 6.550$, $p = .01$], but there was not a significant interaction between the lengths. The main effect of speaker was also a significant predictor of pause frequency [$\chi^2(38,1517) = 175.275$, $p < .001$]. Figure 1 shows the proportions of pairs which contained a pause by both the preceding (Recovery) and following (Planning) sentence lengths. Speakers paused more frequently following longer sentences than shorter ones (Short: $\hat{p} \approx .94$; Long: $\hat{p} \approx .98$), and also preceding longer sentences than shorter ones (Short: $\hat{p} \approx .95$; Long: $\hat{p} \approx .97$). While the effect of pair type was not statistically significant, there was a trend for increasing frequency from the SS to LL sentence pairs (SS: $\hat{p} \approx .93$; SL: $\hat{p} \approx .96$; LS: $\hat{p} \approx .97$; LL: $\hat{p} \approx .99$). Overall, there was a strong tendency to produce pauses between the sentences regardless of the experimental condition ($\hat{p} \approx .96$), which is expected given that the pairs were designed to elicit inter-sentence pausing. The full model containing all of the fixed effects and interactions explained a large amount of the overall variance in pause frequency (*pseudo*-$R^2 = .38$), but it appears that the effect of speaker alone accounted for the majority of this explained variance.

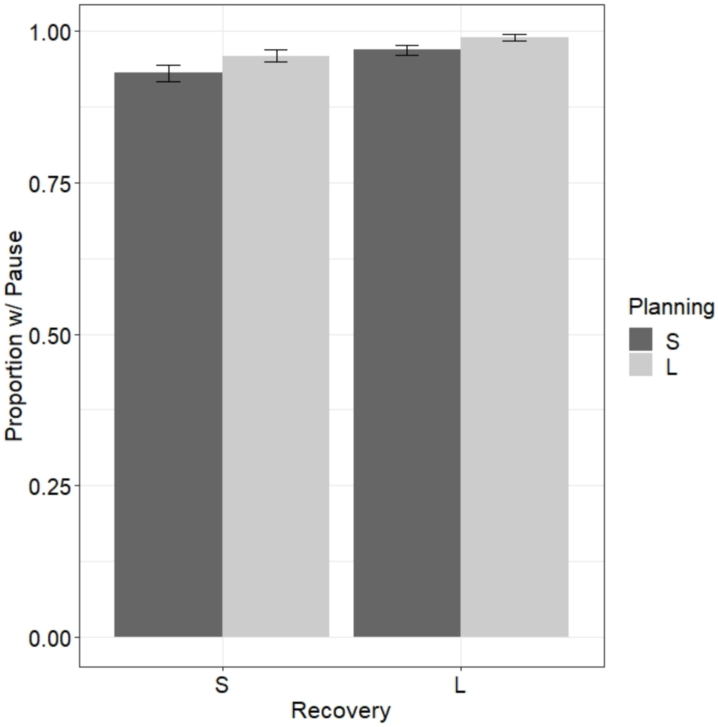


Figure 1. Proportion of pairs with a pause by preceding (Recovery) and subsequent (Planning) sentence lengths.

3.2.2. Pause Duration

There were significant main effects of both the preceding and following sentence lengths on the durations of inter-sentence pauses [Preceding: $F(1,1457) = 81.957$, $p < .001$, *partial*-$\omega^2 = .05$; Following: $F(1,1457) = 13.337$, $p < .01$, *partial*-$\omega^2 < .01$], but the interaction between lengths was not significant. The main effect of speaker was also significant [$F(38,1457) = 19.076$, $p < .001$, $partial\text{-}\omega^2 = .31$]. Figure 2 shows the mean pause durations by both preceding (Recovery) and following (Planning) sentence lengths; Figure 3 shows the pause durations by sentence pair types. It is visible in Figure 2 that the pauses were shorter following short sentences and longer following long sentences (Short: $M = 465$ msec., $SD = 269$; Long: $M = 556$ msec., $SD = 223$). Likewise, pauses were generally shorter preceding shorter sentences and longer preceding longer sentences (Short: $M = 494$ msec., $SD = 267$; Long: $M = 529$ msec., $SD = .233$). Despite the lack of statistical effect of the interaction between pair types, Figure 3 shows that there was a systematic trend for increasing durations from SS to LL sentences pairs (SS: $M = 442$ msec.; SL: $M = 488$ msec.; LS: $M = 543$ msec.; LL: $M = 569$ msec.). Moreover, the recovery effect appears to be stronger than the planning effect in that there were generally shorter pauses in the SS and SL conditions than in the LS and LL conditions. There overall fit of the full model was significant [$F(41,1457) = 19.79$, $p < .001$, $R^2 = .34$], but again the differences in effect sizes between the predictors suggests that this is driven primarily by the main effect of speaker.

In light of the study prediction that pause durations would increase with increasing length of the preceding sentence as a function of the need to breathe, the strong recovery effect on pause durations begs the question of whether speakers were actually breathing more frequently following longer sentences than following shorter ones. An additional simple logistic regression model was constructed to predict the presence/absence of breath intakes, where intakes were hand-coded based on visible and audible acoustic evidence. This analysis confirmed that the preceding sentence length had a significant effect on whether the pause contained a breath [$z(1557) = 11.713$, $p < .001$]. Figure 4 shows that speakers were significantly more likely to take a breath following a long sentence than following a short sentence. The participants took breaths during approximately 82% of the pauses when the preceding sentence was long compared to only 54% when the preceding sentence was short. There was also a strong correlation between pause duration and breath duration in pairs which contained a breath ($r = .53$), which further suggests that the observed recovery effect on pause durations was strongly driven by the physiological need for respiration.

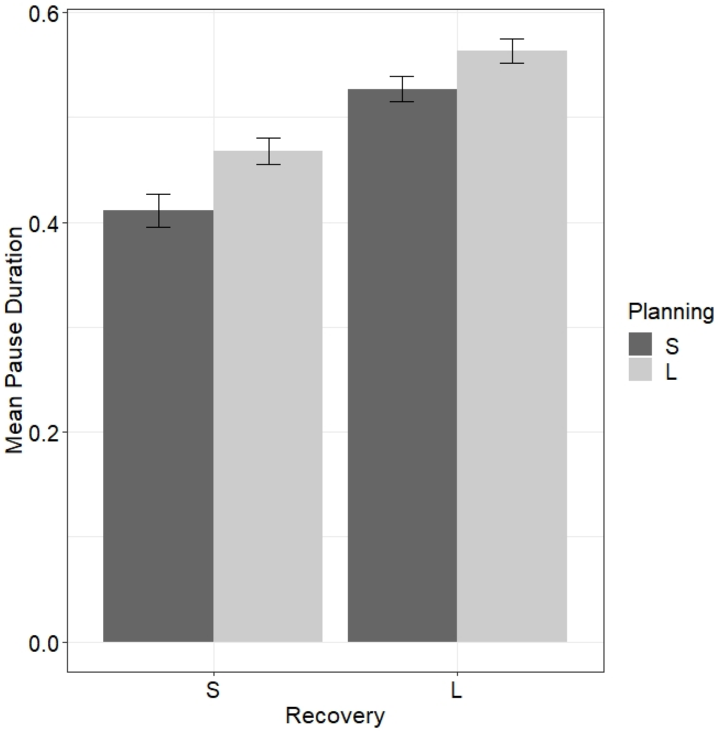


Figure 2. Mean pause duration by preceding (Recovery) and subsequent (Planning) sentence lengths. Bars represent the Standard Error of the Mean.

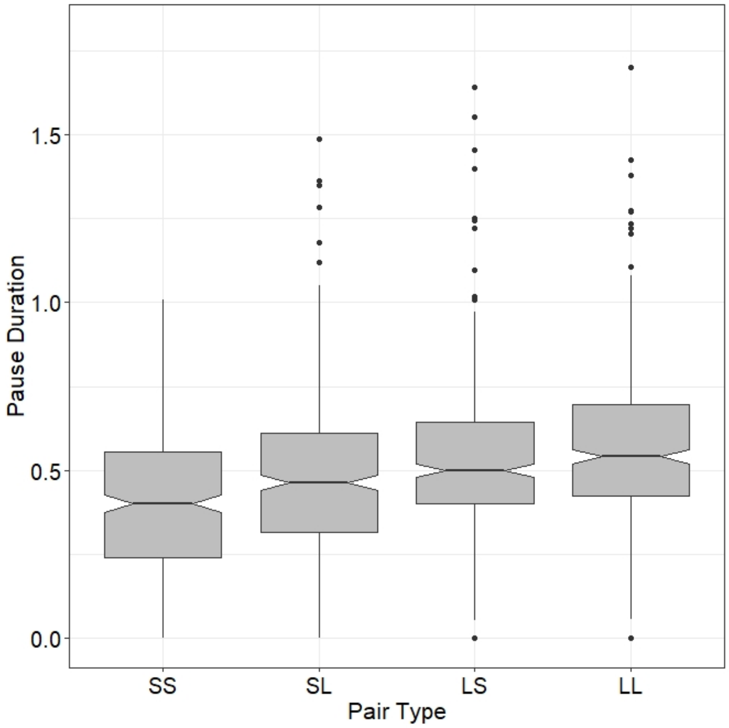


Figure 3. Pause durations by sentence pair types.

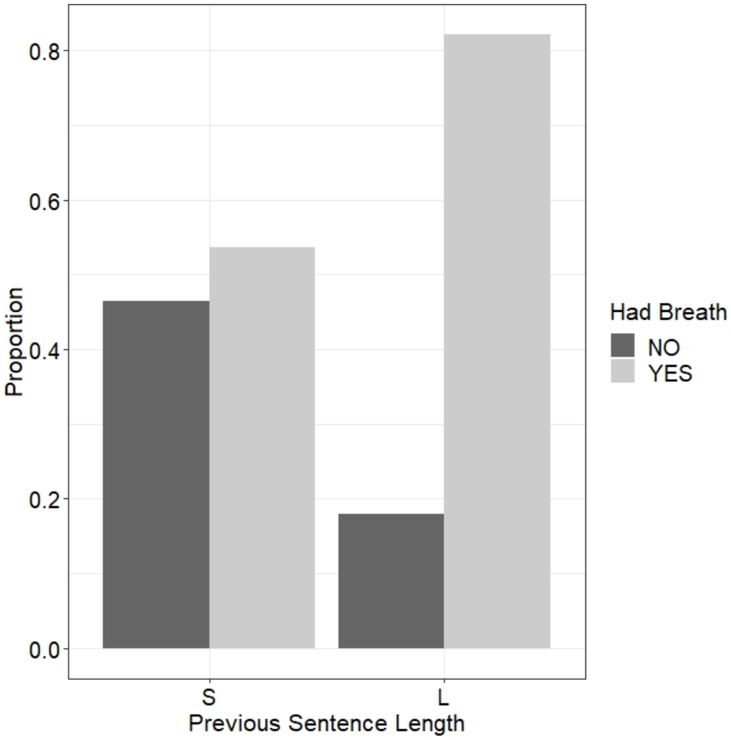


Figure 4. Proportion of pairs without or with an inter-sentence breath intake by preceding sentence length.

3.2.3. Pause Duration Variability

There was a significant main effect of the preceding sentence length on the CV_{dur} measure [$F(1,114) = 24.253$, $p < .001$, $partial\text{-}\omega^2 = .13$], but not an effect of the following sentence length or of the interaction between the lengths. The main effect of speaker was also significant [$F(38,114) = 3.002$, $p < .001$, $partial\text{-}\omega^2 = .33$]. Figure 5 shows the mean CV_{dur} measure by both preceding (Recovery) and following (Planning) sentence lengths.

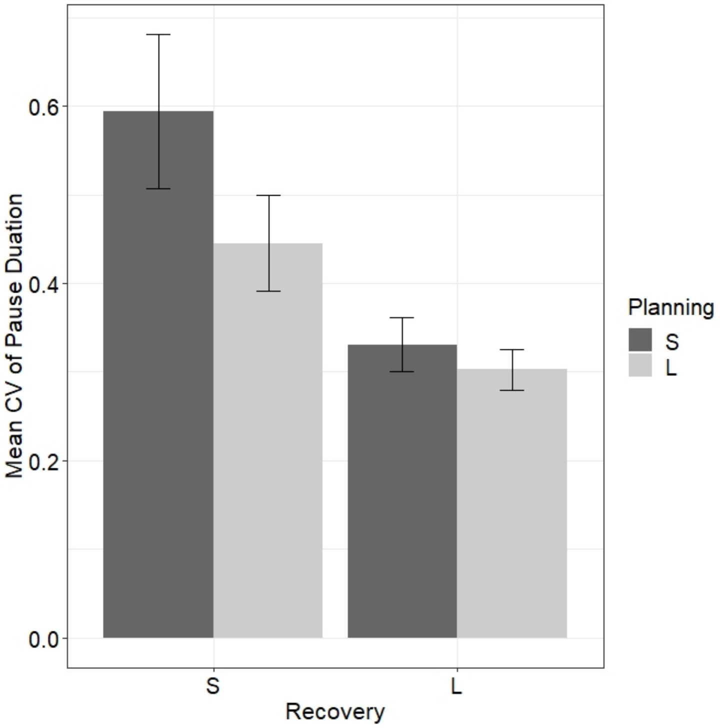


Figure 5. CV_{dur} by preceding (Recovery) and following (Planning) sentence lengths. Bars represent the Standard Error of the Mean.

It is shown that the variability of pause durations was greater when the preceding sentence was short than when it was long (Short: $M = 0.439$, $SD = 0.314$; Long: $M = 0.285$ msec., $SD = 0.127$). Despite the lack of statistical significance of following sentence lengths there was also a visible trend of greater variability of pause durations before shorter sentences than before long ones (Short: *M* = 0.389, *SD* = 0.288; Long: *M* = 0.337 msec., *SD* = 0.206). However, the difference in the CV_{dur} measure was much less pronounced between short and long following sentences than between short and long preceding sentences. The full model containing all of the main effects had a significant fit [$F(41,114) = 3.503$, $p < .001$, $R^2 = .40$], with the main effect of speaker explaining more of the variance than those of the sentence lengths.

3.3. Discussion

The results of Experiment 1 revealed strong effects of both respiratory recovery and speech planning on each dimension of the pausing patterns. There was a systematic pattern of increased frequency of pausing when both the preceding and following sentences were long compared to when they were short. The increase in pausing frequency before longer sentences suggests that the speakers required additional time with which to plan for production of a greater number of syllables. On the other hand, the increase in pausing frequency following longer sentences also suggests that with an increasing number of syllables produced the physiological drive to replenish the air supply becomes stronger. There was also evidence that the planning and recovery effects were additive in that the lowest frequency of pausing occurred in the SS sentence pairs, followed successively by the SL, LS, and LL pairs. This pattern could perhaps be interpreted as suggesting that the recovery effect was generally stronger than the planning effect. Of course, it is also important to bear in mind that while there was a systematic pattern of effects on pausing frequency, the rate of pause occurrence was extremely high across all 4 pair types. Even between 2 short sentences the speakers paused approximately 93% of the time. The sentences in each pair were semantically unrelated, so it is reasonable to assume that the strong junctures between them correspond with the notion of topic boundary in more natural discourse. The finding in the current study fits with previous findings that speakers are overwhelmingly more likely to pause at “major changes in the information flow (Swerts & Geluykens, 1994, p. 29)” relative to other locations within the discourse (see also van Donzel, 1999; Oliveria, 2002).

There were also systematic effects of both the preceding and following sentence lengths on the durations of pauses. The speakers took longer pauses both after and before producing longer sentences than shorter ones, suggesting that both respiratory recovery and speech planning processes are significant determinants of relative pause durations. These effects also appeared to be additive such that the pause durations in the SL and LS sentence pairs were intermediate to those in the SS and LL sentence pairs.

The observed planning effect on pause duration in the current study is in line with the findings from previous studies. Fuchs et al. (2013) found speakers to pause longer before longer sentences (= 24 syllables) than before shorter sentences (= 12 syllables) regardless of syntactic complexity (see also Ferreira, 1991). Interestingly, the difference in mean durations between the short and long sentence conditions was remarkably similar in both their study (~38 milliseconds) and the current one (~35 milliseconds). In an earlier study, Whalen and Kinsella-Shaw (1997) also found that speakers took longer breaths before reading longer sentences than they did before reading shorter ones with a similar lack of effect from syntactic structure. Their finding suggests that the effect of upcoming sentence length on pause durations may reflect motor planning for speech breathing as opposed to phonological-phonetic planning. To this point, Fuchs et al. (2013) also found that in addition to producing longer pauses speakers took longer and deeper breaths before longer sentences. In the current study, there was also some evidence of motor planning related to speech breathing in that speakers took a breath during roughly 66% of the pauses that occurred before short sentences as opposed to during 75% of the pauses before long sentences. While this difference is not nearly as dramatic as the difference in proportions of breath pauses *following* short and long sentences, it nonetheless suggests that these speakers were perhaps anticipating the need for a greater amount of air to produce the longer sentences.

The current study differed from previous ones in that it was also aimed at quantifying the effects of physiological recovery in addition to the demonstrated planning effects. While speech production researchers have occasionally acknowledged that “respiratory pauses” are “physiological necessities” (Vaissiere, 1983, p. 54; see also Zellner, 1994), the precise effect of respiratory recovery on pause durations remains largely unquantified. Having said that, the observed presence of both recovery and planning effects in the present study is not entirely without precedent. Zvonik & Cummins (2003) asked participants to read aloud passages from a novel and found that their pause durations varied systematically as a function of the number of syllables in both the preceding and following IPs. Moreover, their study found similarly that these effects were additive in that durations varied systematically as a function of the combined lengths of surrounding IPs. Importantly, the results from the present study revealed the same level of systematic recovery and planning effects even in the absence of potentially confounding linguistic factors. Whereas the sentences in a novel passage display semantically cohesive relations and contain highly variable syntactic structures, the semantic and syntactic features of the sentence pairs used in the present study were tightly controlled. The findings here suggest that while global pausing patterns are shaped to some extent by language factors, pause durations are clearly constrained by lower-level speech planning processes.

There was also a systematic effect of preceding sentence length on the CV of pause durations. When the preceding sentence was short the pause durations were considerably more variable than when the preceding sentence was long. The mean $\mathrm{CV}_{\mathrm{dur}}$ value was nearly twice as large for SS pairs (*M* = .489) than for either of the pair types in which the first sentence was long (LS: *M* = .284; LL: *M* = .286). This pattern is similar to that found by Zvonik and Cummins (2003) where pause durations increased with increasing length of the preceding utterance, but only up to approximately 10 syllables. The mean durations of pauses following longer utterances were much more uniform. A likely explanation for the differences in duration variability between the preceding sentence lengths lies in the observed respiratory recovery effect on pause durations. Recall that the participants took a breath intake following 82% of the long sentences while they only did so following 54% of the short sentences. This suggests that the pauses served a more systematic function when they followed longer sentences. Greater systematicity corresponds with less variability by definition. This interpretation is supported by the fairly strong correlation between pause duration and breath duration in the breath pauses (*r* = .53).

While there were not significant effects of the following sentence length or of the interaction between preceding and following lengths (pair type) on duration variability, there were nonetheless noticeable patterns in both regards. When the preceding sentence length was 12 syllables, the mean $\mathrm{CV}_{\mathrm{dur}}$ was higher before short sentences (*M* = .489, *SD* = .361) than it is was before long sentences (*M* = .388, *SD* = .253). Furthermore, when the preceding sentence length was 24 syllables the mean $\mathrm{CV}_{\mathrm{dur}}$ values were nearly identical between following short and long sentences, suggesting the presence of an interaction between lengths. So, while not statistically significant, these patterns appear to be systematic and indicate that the cumulative length of both sentences in a sentence pair play some role in determining the overall variability of pause durations.

3.4. Conclusions

Taken as a whole, the results presented here largely support both the planning and recovery hypotheses as presented at the beginning of this chapter. First in terms of planning, the predicted effects of the following utterance lengths on both pause frequency and pause duration were confirmed. Longer sentences led to preceding pauses which were both more frequent and longer. Only the predicted effect of planning on pause duration variability was not upheld. That prediction was based on an assumption that pauses before longer sentences would be more functionally systematic, but the overwhelming tendency to pause regardless of length suggests that the pauses were equally systematic in terms of signaling the boundary between unrelated sentences. All 3 of the predicted recovery effects were also found. The hypothesized positive relationship between preceding sentence length and pause frequency was confirmed, as was the similar prediction for pause duration. In contrast to the planning effect, the predicted recovery effect on pause duration variability was also evidenced. This supports the assumption that breath pauses are more functionally systematic than non-breath pauses. In terms of the relationship between the 2 types of effects, there was clear evidence of an interaction between planning and recovery, but there was also a strong suggestion that the latter exerts greater influence on pausing patterns than does the former.

The results from this experiment indicate that a complete model of pausing behavior must account for influences from low-level speech planning and respiratory recovery processes. Given the design of the stimuli used in the experiment, it also clear that these processes should be assumed to operate at least somewhat independently of additional processes related to syntactic or discourse structure. Experiment 2 of the current study is presented in the following chapter, and it was aimed instead at illuminating the role that syntactic structure might play in a larger model of pausing.

IV. EXPERIMENT 2: MID-LEVEL LINGUISTIC COHESION EFFECTS

The second experiment in the current study had the general aim of quantifying the effects of mid-level cohesion factors on pause frequency, pause duration, and the variability of pause durations. These factors were investigated by varying the degree of semantic relatedness in successive clauses. Although the existing literature on pausing suggests that pausing patterns are robustly constrained by the local syntactic structure of language productions, the relative strengths of the syntactic junctures affect both the likelihood of a pause occurring at that boundary as well as the relative durations of those pauses. In the current study, the strength of a major syntactic juncture (i.e. the clause) was further modulated by the degree of cohesion between the successive clausal units. For instance, successive independent clauses in narrative speech are considered cohesive in the sense that they are both embedded within the scope of some shared higher-level discourse unit (e.g. theme, topic). The degree of cohesion between them is considered to be relatively low, however, in that there is no direct overlap in their linguistic content; each can be understood independently of the other. Contrastively, the boundary between a main clause-subordinate clause pair is characterized by relatively high cohesion because a full understanding of the communicative message conveyed by the latter is only achievable with reference to the linguistic content of the former. In other words, a subordinate clause uttered in isolation leaves the listener wondering as to its meaning. Most previous studies on syntactic structure effects have focused on mostly broad distinctions such as paragraph vs sentence or sentence vs clause. What is lacking is an understanding of how pausing patterns vary between successive units of the same type (e.g. clauses) but different subtypes (e.g. coordinated, relative, etc.). Experiment 2 addresses this question.

Several predictions follow from the assumption that cohesion modulates juncture strength. The first prediction is that pausing frequency will decrease with increasing cohesion between consecutive clauses. Pairs of more cohesive clauses are expected to form a dependent relationship wherein the meaning of one is entirely dependent of the context of the other. For example, overt subject references are often dropped in dependent clauses and so they are not minimally complete information units on their own. Thus, as both clauses together create a unified unit of meaning the prediction is that a pause is less likely to occur between them. Likewise, in the case that a pause does occur between highly cohesive clauses the prediction is that the duration of that pause will be shorter than ones between less cohesive clauses. This prediction is based on the suggestion that pause duration is a salient cue to the information structure of discourse (see Swerts, 1997). In terms of pause duration variability, the expectation of shorter pauses between more cohesive clauses leads to a prediction of less variability of those durations. This prediction is based largely on the assumption that shorter pauses overall are more constrained in the range of their possible durations leading to greater uniformity.

Each of these hypotheses was tested using an experimental procedure that was very similar to the one used in Experiment 1. Participants were once again tasked with reading pairs of sentences from a computer screen, but these sentence pairs consisted of meaningfully related clauses. The degree to which the clauses in each pair were related was manipulated to create 3 distinct experimental cohesion conditions. Importantly, the lengths of clauses in terms of number of words was held constant in contrast to the sentence pairs used in the first experiment. This allowed for assessment of the effect of clausal cohesion on the pausing patterns in a way that was independent of confounding speech planning or respiratory recovery effects. The clause pair productions that were collected from this experiment were analyzed in terms of the effects that the experimentally manipulated coherence conditions had on the patterns of inter-clausal pausing.

4.1. Methods

The participants in the current experiment were the same as those in Experiment 1 (see Section 3.1.1.). The details of the study sessions were also identical (see Section 3.1.2.). All other methodological differences are given as follows.

4.1.1. Materials

The stimulus items used in Experiment 2 of the study consisted of pairs of semantically related clause pairs. The pairs that were constructed for this production task varied according to the relative strengths of the syntactic boundaries between them. For this purpose, boundary strength was defined with reference to *referential coherence* and *embeddedness*. As defined by Kintsch & van Dijk (1978, p. 367), "referential coherence corresponds to argument overlap among propositions" such that 2 successive clauses are considered referentially coherent if they share a common argument (see also Garnham, Oakhill, & Johnson-Laird, 1982; Kintsch, 1995). The *embeddedness* of the clause pairs was characterized with respect to Swerts' (1997) representation of discourse structure wherein subtopical information units become embedded within higher-level topical units. It is in this way that hierarchical discourse structure emerges, and the degree of *embeddedness* varies relative to position in that structure (see also Gee & Grosjean, 1983; Thompson & Mann, 1987).

The clause pairs used in the study were classified according to 3 boundary strength conditions: Separate, Coordinated and Relative. In the Separate boundary condition, the pairs consisted of related but separate clauses which were connected by a discourse marker such as *and*, *then*, or *and then* (see Fraser, 1999; Schiffrin, 2001). In each of these pairs, the subjects of the two clauses were always different and so there was no *referential coherence*. Nonetheless, the second clause was *embedded* as a subtopic of the first clause. In the Coordinated boundary condition, the pairs consisted of related and coordinated main clauses. These pairs were similar to those in the Separate boundary condition in that they were connected with *and*, but they differed in that the clauses shared a subject which was expressed pronominally in the second clause. Thus, the second clause in the Coordinated pairs were also embedded as a subtopic of the first clause, but there was a higher degree of referential coherence than in the Separate condition pairs. In the Relative boundary condition, the pairs consisted of a main clause followed by a relative clause. More specifically, the object of the first clause was relativized in the subsequent clause. The structure of the Relative clause pairs can be represented as:

$[X_1 \text{ V Y } [\textit{that}_{Y}\ X_2 \text{ V}\_{Y\ \ldots}]]$.

As a type of subordinate clause, relative clauses are entirely *dependent* on the main clause for referent resolution and are thus strongly embedded within the main clause. Furthermore, the degree of referential coherence between the 2 clauses in these pairs was very high given that the embedded object was expressed with a zero form rather than pronominally. As such, the identity of the referent is presupposed and only recoverable given the discourse context (see Nariyama, 2004).

Since Experiment 2 was aimed at investigating language-related rather than speech-related effects on pausing, all pairs were comprised of 24 syllables distributed evenly between the 2 clauses. In all cases, the connecting linguistic device (e.g. *and*, *then*, *that*) was included in the length of the second clause. Similar to the stimuli used in Experiment 1, the words used in the clause pairs for Experiment 2 were also limited to those found in the top 1000 in frequency in COCA to control for potential word frequency effects (Davies, 2008-present). Three separate clause pairs were constructed for each boundary strength condition for a total of 9 unique pairs. Each participant was presented with 5 blocks of the 9 pairs for a total of 45 pairs produced in the task. The blocks of clause pairs were presented to participants in 1 of 4 predetermined randomized orders. Two of the orders were obtained by creating lists of the 5 blocks and then randomizing the orders of the 9 pairs within each block. A second set of 2 orders was then obtained by taking the reverses of the 2 randomized orders. The sets of pairs were then numbered and rotated through for each participant such that if the previous participant was given order 1 then the current participant was given order 2 and the subsequent participant order 3, and so on. In this way, each of the 4 orders was used for exactly 10 of the 40 participants. The following are examples of each pair type, and the full set of pairs can be found Appendix B:

Separate: *My brother and I visited the property*
and then the manager showed us many units.

Coordinated: *The executive understood the benefits*
and he identified the most simple response.

Relative: *I have yet to read any of the science books*
that my professor gave me for personal use.

4.1.2. Procedures

The procedures for Experiment 2 were nearly identical to those for Experiment 1. The clause pairs were presented on separate lines on a computer screen with an empty line of between them. The text was white in 48 point font and shown against a black background. It was centered on the screen both vertically and horizontally.

There were 3 phases of stimuli presentation: *read*, *exhale*, *say*. Prior to the initial presentation of each pair, a visual timer would countdown from 3 to 1 on the screen. Following that, the pair would appear with the text “READ:” at the top of the screen. The participants were instructed to read the clause pair to themselves at this time in order to familiarize themselves with what they would be asked to say. After 8 seconds the text would then disappear and be replaced by new text prompting the participant to exhale completely before clicking the button to continue. Once the button was clicked a “ding” sound was played, and this was a prompt for the experimenter to read a predetermined question to the participant. The pair would then reappear on the screen 3 seconds after the sound with the text “SAY:” at the top of the screen instead of “READ:”. At this time the participant would read the clause pair out loud in response to the question. The goal of this elicitation method was to make the speaking task somewhat more natural. In particular, the goal was to elicit more natural prosody.

4.1.3. Segmentation & Transcription

The onsets and offsets of each clause pair that was produced were marked by trained researchers based on the visible and audible acoustic signals using Praat software (Boersma & Weenink, 2017). All sentences were constructed to be sonorant-initial and -final, so no specific criteria for marking stop closure boundaries were required. For utterance-initial or -final vowels, nasals, and liquids, onsets and offsets were marked at the point where the second formant became visible or disappeared respectively. Utterance-final fricatives often displayed a clear portion of spectral energy which was followed by some residual energy in a narrower band of frequencies. These fricative boundaries were marked at the point where the bulk of the spectral energy appeared to end and there was a visible decrease in amplitude. The onset and offset boundaries of pauses were also marked according to the same criteria as above. Pauses were defined as any period of silence between the offset of the first clause and the onset of the second clause that had a duration of at least 50 milliseconds (see Section 3.1.5.).

4.1.4. Measures & Analyses

There were 3 dependent variables of interest in Experiment 2 of the study: pause frequency, pause duration, and pause duration variability. Pause frequency was measured in terms of the relative proportions of pairs which did or did not contain an inter-clausal pause. The inter-clausal durations of each pair were extracted directly from the segmented audio files. Pause duration variability was measured as the Coefficient of Variation of pause durations (CV_{dur}). The CV_{dur} measure was calculated as the standard deviation of the durations divided by the mean duration within-condition for each speaker separately. The primary independent variable of interest in each of the analyses was the clausal Cohesion conditions of the pairs. As in Experiment 1, the main effect of speaker was included in each of the models to assess to overall variance that it explained independent from the experimental condition effect.

The effects on pause frequency were assessed by first fitting a logistic regression model with the 3 IVs as predictors of the presence or absence of an inter-clausal pause. The statistical significance of each of the predictors in the logistic regression model were then tested using binomial ANOVA analysis. The effects on pause duration and CV_{dur} were assessed using two-way crossed factorial ANOVA models with effects structures that mirrored the logistic regression model for pause frequency. Separate analyses were run to assess differences between each pairwise combination of the 3 coherence conditions on pause frequency, duration, and variability. The additional analyses were identical to those just described, but with the main effect of clause coherence condition as the lone IV. The models predicting pause duration and CV_{dur} included only those pairs for which there was an inter-clausal pause. McFadden's pseudo-R^2 is reported as a measure of explained variability in the full model predicting pause frequency (see McFadden, 1974, 1978). Partial-ω^2 is reported as a standardized effect size measure for all statistically significant effects in the models predicting pause duration and CV_{dur}.

4.2. Results

4.2.1. Pause Frequency

There was a significant main effect of the experimentally defined Cohesion condition on the frequency of inter-clausal pausing [$\chi^2(2,1751) = 206.29$, $p < .001$]. There was also a significant main effect of speaker [$\chi^2(38,1713) = 323.80$, $p < .001$]. Figure 6 shows the differences in pausing frequency between the 3 Cohesion conditions, and it is visibly apparent that there were significant differences for the pairwise combinations of Relative versus Coordinated pairs [$\chi^2(1,1167) = 82.491$, $p < .001$], Relative versus Separate pairs [$\chi^2(1,1167) = 205.962$, $p < .001$], and Coordinated versus Separate pairs [$\chi^2(1,1168) = 31.591$, $p < .001$]. The speakers paused least frequently between the Relative clause pairs ($\hat{p} \approx .04$) followed consecutively by the Coordinated clause pairs ($\hat{p} \approx .22$) and the Separate clause pairs ($\hat{p} \approx .36$). It is also worth noting that pausing was relatively infrequent overall regardless of Cohesion condition ($\hat{p} \approx .21$). The full model including the main effects of Cohesion condition and Speaker along with their interaction accounted for a large amount of the overall variance in pausing frequency (*pseudo*-$R^2 = .35$).

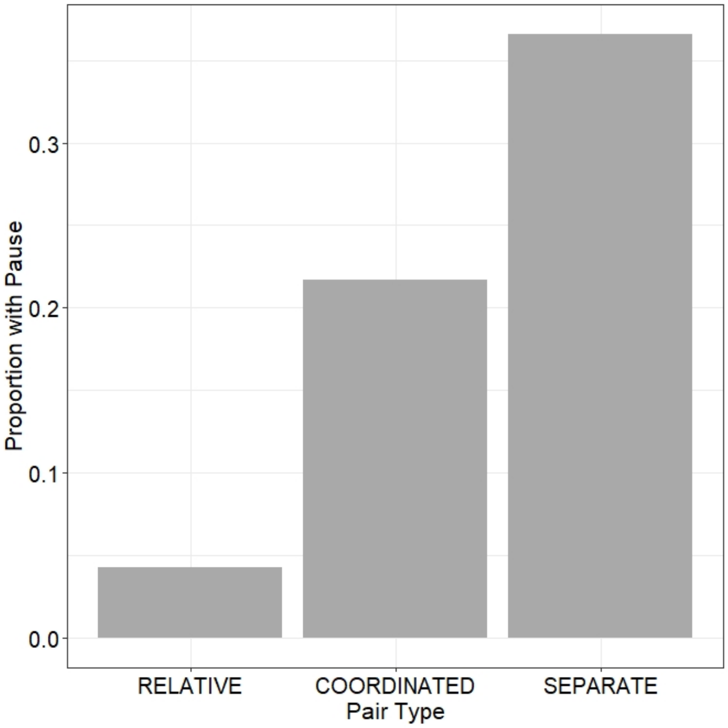


Figure 6. Proportion of clause pairs with a pause by clause pair type.

4.2.2. Pause Duration

There was a significant main effect of the experimentally-defined Cohesion condition on the durations of inter-clausal pauses [$F(2,328) = 4.667$, $p = .01$, *partial*-$\omega^2 = .02$]. There was also a significant main effect of speaker on the pause durations [$F(36,328) = 2.722$, $p < .001$, *partial*-$\omega^2 = .14$]. Figure 7 shows the pause durations by Cohesion condition, and it can be seen that there were significant differences between the Relative and Coordinated conditions [$F(1,113) = 8.725$, $p < .01$, *partial*-$\omega^2 = .06$] as well as between the Relative and Separate conditions [$F(1,190) = 6.404$, $p = .01$, *partial*-$\omega^2 = .03$]. Pause durations were generally longer in the Relative condition while the mean pause durations were very similar between the Coordinated and Separate conditions (Relative: $M = 222$ msec., $SD = 147$; Coordinated: $M = 148$ msec., $SD = 122$; Separate: $M = 151$ msec., $SD = 117$). Despite the clear pattern of differences in pause durations shown in Figure 7, caution should be exercised in drawing any sort of meaningful conclusion from these results. Very few of the Relative clause pairs were produced with an inter-clausal pause overall ($N = 26$, $\hat{p} \approx .04$), and of those with a pause exactly ½ of them were produced by only 3 of the speakers. No other speaker paused in the Relative condition more than twice out of the total 15 relative clause pairs. Thus, there is not only an insufficient amount of data in the Relative clause condition, but there is also the possibility that the productions of the 3 speakers who did pause in this condition were simply abnormal. There overall fit of the full model was significant [$F(38,328) = 2.85$, $p < .001$, $R^2 = .16$], and the effect size measures of each predictor suggests that the main effect of speaker accounted for a greater amount of the variance in pause durations.

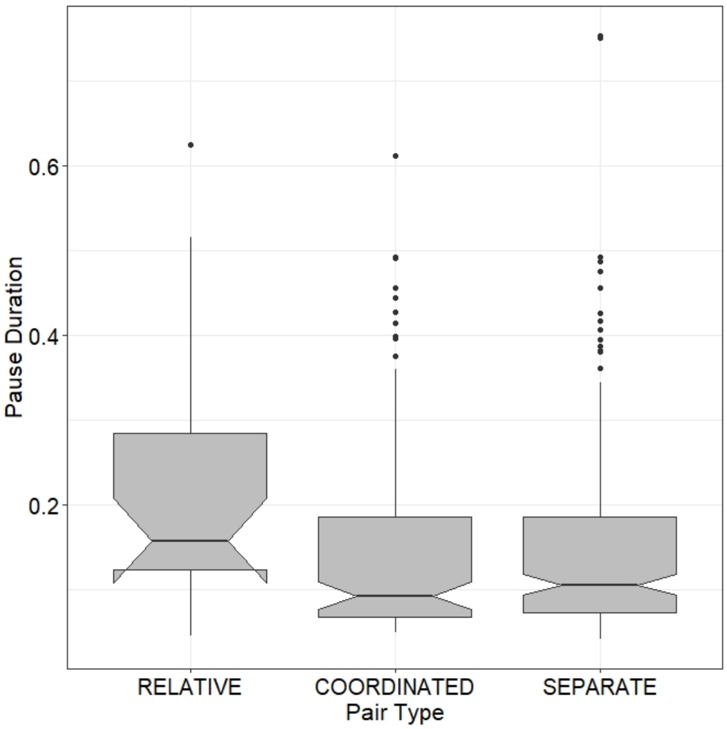


Figure 7. Inter-clausal pause duration by pair type.

4.2.3. Pause Duration Variability

There was not a significant effect of Cohesion condition on the CV_{dur} measure, nor was there a significant main effect of speaker. There were also no significant differences between any of the pairwise combinations of Coherence conditions. Figure 8 shows the CV_{dur} values by condition and it is apparent that despite the lack of statistically significant effect of Cohesion, there was nonetheless a systematic increase in variability with decreasing cohesiveness (Relative: *M* = 0.394, *SD* = 0.199; Coordinated: *M* = 0.469, *SD* = 0.323; Separate: *M* = 0.581, *SD* = 0.280). As with the results on the pause durations, however, the number of data points is considerably variable within individual Coherence conditions and there are very few observations in the Relative clause condition (*N* = 6). So it is possible that the lack of significant effects reflects the unequal distributions across conditions. There was not a significant overall fit of the full model predicting pause duration variability**.**

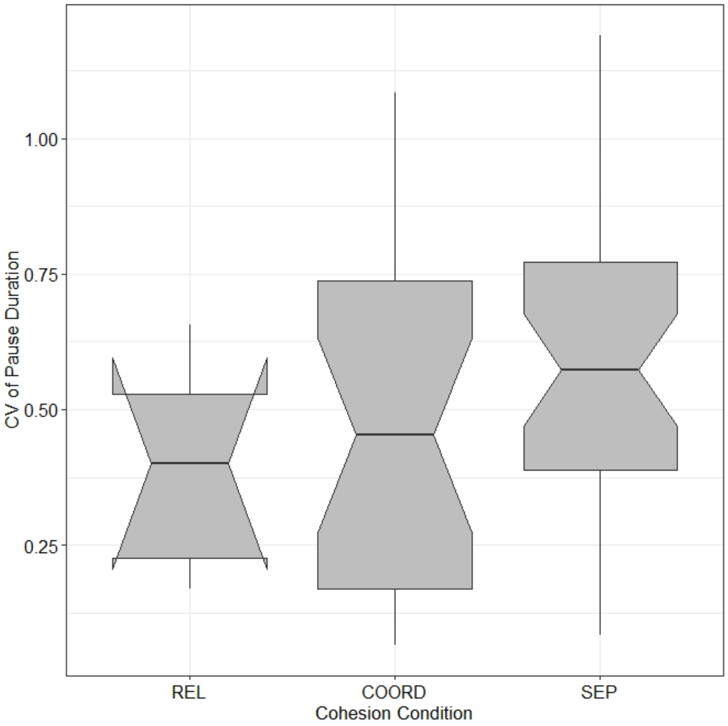


Figure 8. CV of pause durations by Cohesion condition.

4.3. Discussion

The results from Experiment 2 revealed strong and systematic effects of clausal cohesion on the frequency of intra-clausal pausing. The occurrence of pauses before relative clauses was found to be exceedingly rare, with speakers pausing in less than 5% of the relative clause pairs. Moreover, only 13 of the 40 speakers who participated in the study produced any pauses before relative clauses, meaning that such pauses are uncommon both across- and within-speaker. This result is in line with expectations based on the findings of previous studies. Referring back to the adopted definition of cohesion (Section 4.1.1.), relative clauses display a very high degree of *referential coherence* with preceding main clauses in terms of shared arguments. Not only are the relative clause subject and main clause object co-referential, but the former has been relativized and its referent is only retrievable in the context of the latter. By virtue of this relationship, the relative clause is also deeply *embedded* within the main clause; the former is an overt subtopic of the latter. As such, the inter-clausal boundary strength is very weak, and numerous studies have demonstrated a systematic relationship between boundary strength and pause location. In one particularly early study on pausing in spontaneous speech, Goldman-Eisler (1972) found that nearly 80% of sentence boundaries were accompanied by a pause of least 500 milliseconds while about two-thirds of clause boundary transitions were either “fluent” (i.e. no pause) or marked by a pause of less than 500 milliseconds More directly relevant to the present discussion, she found that nearly two-thirds of relative clause transitions were fluent compared to only one-third of coordinate clause transitions. Subsequent studies have found similar relationships between syntactic structure and pausing (Duez, 1982).

In terms of the relative frequency of intra-clausal pausing in the other 2 cohesion conditions, the results of the present study continued to follow the same systematic pattern described by Goldman-Eisler (1972). Whereas the speakers almost never paused at relative clause transitions, they paused at nearly one-fourth of coordinated clause transitions and one-third of related but separate clause transitions (i.e. coordinated sentence boundaries). One important difference between the current study and the earlier Goldman-Eisler study lies in the relative frequencies of pausing at each boundary type. While there was also a clear relationship between boundary strength and pause occurrence in Goldman-Eisler’s study, her speakers paused much more frequently overall than did those in the current study. This fact is likely due to often observed differences between spontaneous and read speech as regards pausing behavior. For example, pausing patterns have been shown to be less systematic in spontaneous speech than in read speech with pauses occurring at more regular intervals and having less variable durations in the case of the latter (Henderson et al., 1965). This is no doubt a reflection of important differences in the cognitive demands associated with each speaking style; the language structure must be entirely generated by the speaker during spontaneous speech. Nevertheless, the robust relationship between boundary strength and pausing that has been replicated in the current study speaks to the importance of local syntactic structure in constraining the temporal rhythmic structure of speech production. To underscore this point, recall that in Experiment 1 the speakers paused between roughly 93% of the SS sentence pairs, which matched the clause pairs produced here in terms of length. Thus, it appears that with just the addition of linguistic cohesion the relative frequency of pausing was found to decrease at least threefold. However, we should be cautious in interpreting this difference given that the speakers had an interlocutor in Experiment 2 but not in Experiment 1. This means that the elicitations from Experiment 2 may represent more naturalistic speech productions.

The results from the current experiment also revealed a significant main effect of coherence condition on the relative pause durations. However, unlike with pause frequency there was not a stepwise difference in pause durations from the least to the most cohesive pair types. The longest durations were found in the relative clause pairs, but there was virtually no difference in overall durations between the coordinated and separate clause pair conditions. More importantly, the finding that speakers paused for longer prior to a relative clause than prior to either of the other clause types runs counter to expectations. Numerous studies have demonstrated instead that pause durations systematically decrease with decreasing boundary strengths. For example, boundary strengths obtained from naïve boundary judgements using text transcripts of spontaneous speech have been shown to be positively correlated with the durations of pauses at those junctures (Swerts, 1997). And Ferreira (1991) found not only that intra-sentence pauses were more likely to occur immediately preceding the verb (i.e. between-phrase) than immediately following it (i.e. within-phrase) in read sentences, but also that the pre-verb pauses were significantly longer than the post-verb ones. Interpretation of the opposite pattern observed in the present study is complicated by the general infrequency of pausing in the relative clause condition. Moreover, since the pauses were not expected to be frequent in that condition given the high degree of intra-clausal cohesion, it is entirely likely that the pauses which *did* occur did so as a function of disfluency. This explanation fits with previous findings that pauses resulting from disfluency tend to be longer in duration than fluent speech pauses (see Henderson et al., 1965; Kendall, 2013). In sum, the current results make it difficult to draw strong conclusions about the relationship between linguistic cohesion and pause duration. In order do so would likely require collection of a much larger data set.

The results on the variability of pause durations was similarly inconclusive. There was evidence of a systematic increase in CV_{dur} with decreasing clausal cohesion, albeit not a significant one. But the analysis of variability suffered from the same lack of data in the relative clause condition as did the duration analysis. Once again, if the relationship between linguistic cohesion strength and pause duration variability is in fact robust and meaningful, a conclusive demonstration of that relationship will likely only emerge with a much larger set of speech data than was collected in the current study.

4.4. Conclusions

While systematic trends were found for each of the 3 pausing variables in the current experiment, the relationship between linguistic cohesion and pause frequency was the only one that appears to be reliably strong. The hypothesis was that pauses would become more frequent between less cohesive clauses, and that pattern was very clear in the current results. The speakers almost never paused between the relative clause pairs but pausing became increasingly more common in the coordinated and separate pairs, respectively. As far as the pause durations and their variability, the results of the current study unfortunately provide more questions than they do answers. The predicted decrease in both durations and variability with increasing cohesion were not able to be confirmed in these analyses.

On a more positive note, the infrequent occurrences of pausing between these clause pairs, especially in comparison to the unrelated sentences in Experiment 1, provides strong evidence that local language structure in the form of clausal cohesion provides a considerable constraint against pausing. Thus, the results of this study indicate that a complete model of pausing behavior must also account for influences from local linguistic structure, if even just in terms of cohesive vs incohesive. Experiment 3 of the current study is presented in the following chapter, and it was aimed at providing information about how the factor of global discourse planning might fit in a larger model pausing behavior.

V. EXPERIMENT 3: HIGH-LEVEL LINGUISTIC COHERENCE EFFECTS

The third and final experiment of the current study had the general aim of quantifying the effects of higher-level discourse planning on pause frequency, pause duration, and the variability of pause durations. These effects were investigated in narrative speech that was elicited using comic book style sequences of pictures which varied according to the degree of relatedness between pictures in the sequence. This design was meant to experimentally manipulate discourse coherence. The analyses then tested whether this manipulation contributed to meaningful differences in pausing behavior. Based on the assumption that discourse planning has a significant effect on pausing, the predictions were as follows. Regarding pause frequency, the prediction was that pausing would be more frequent overall when the pictures used to elicit the narratives were less clearly related to each other than when they were more clearly related to each other. This prediction assumes that more clearly related pictures represent topics or events which are characterized by relatively weak conceptual junctures. Previous studies have demonstrated a positive relationship between discourse boundary strength and pause likelihood (e.g., van Donzel, 1999; van Donzel & Koopmans-van Beinum, 1996; Gustafson-Capkova & Megyesi, 2002). Conversely, as pictures become less closely related, the strength of conceptual boundaries between them will increase, leading to more frequent pausing.

It was furthermore assumed that construction of a story based on more clearly related events imposes less cognitive demands than one based on unrelated events. In the case of the former, the speaker must do a lot more work to fill in the blanks between the pictures in order to make an incoherent sequence of events into a coherent one. It was expected that this work would result in more frequent pausing when pictures were unrelated.

Regarding pause durations, the prediction was similarly that narratives elicited using unrelated pictures would be characterized by longer pauses overall than those elicited using related pictures. This prediction also follows from the assumption that narratives based on incoherent events require a greater degree of planning, and planning takes time (see Ferreira, 1991; Henderson et al., 1965; Krivokapic, 2012).

Finally, regarding pause duration variability, it was expected that pausing patterns in narratives elicited using highly related pictures would be less variable than those elicited using less related pictures. This prediction is again related to the idea that cognitive demands are lower when the speaker does not have to impose a discourse structure on the narrative because it is contained within the related pictures. Absent cognitive demands for discourse planning, pausing will follow from lower-level factors that are more systematic.

5.1. Methods

The participants in the current experiment were the same as those in Experiment 1 (see Section 3.1.1.). The details of the study sessions were also identical (see Section 3.1.2.). All other methodological differences are given as follows.

5.1.1. Materials

Comic-strip style series of wordless cartoon pictures were used to elicit narrative speech samples from the participants. Each picture set consisted of 5 pictures designed to create 3 separate coherence conditions: High, Moderate, and Incoherent. In the High coherence condition, the pictures consisted of a sequence of local actions which could be described as comprising as single *Event Schema* (see Mandler, 1984; Schank & Abelson, 1977). According to *Story Schema Theory*, narrative stories are hierarchically organized with larger units such as *Settings* and *Episodes* at the higher levels and smaller *Event Schema* units at the lower levels (see also Mandler & Johnson, 1977; Rand, 1984; Rumelhart, 1975; Thorndyke, 1977). Individual event components (i.e., actions) occur within an *Event Schema* and can proceed either sequentially, temporally, or causally at the most local level of the narrative. Given the tightly bounded nature of the pictures in the High coherence condition, the expectation was that these stories would be the easiest to conceptualize and recount. Figure 9 show an example of a High coherence picture set that was used in the study and it illustrates the locally connected nature of the component event within the *Event Schema* of building a birdhouse.

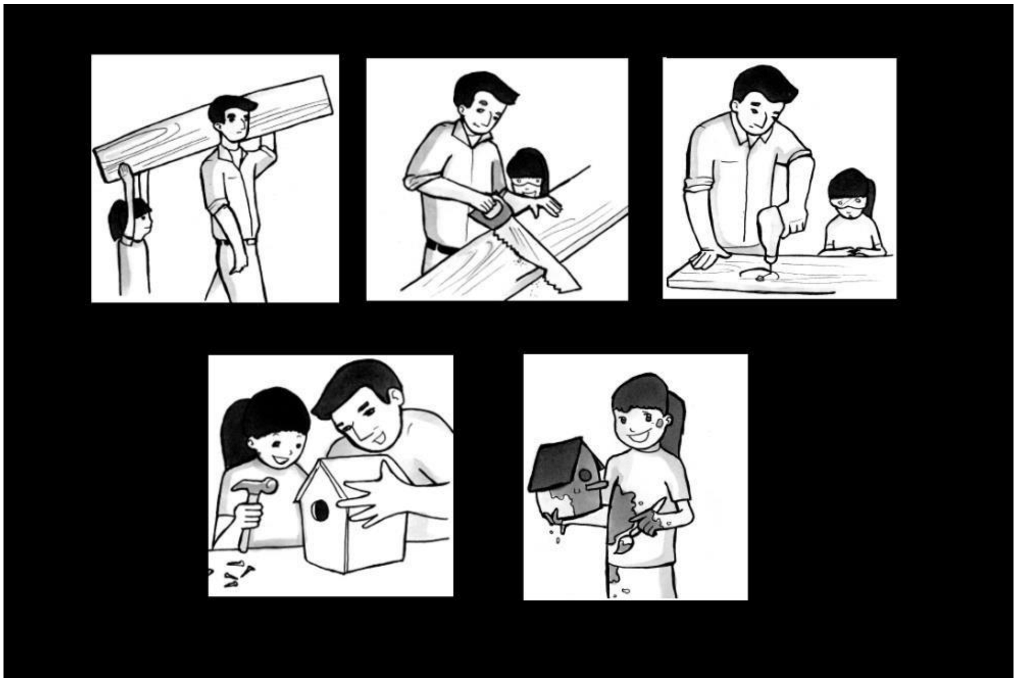


Figure 9. High coherence picture set example from Experiment 3 of the study.

In the Moderate coherence condition, the pictures consisted of more temporally distant actions (i.e. scenes) which can be characterized as comprising a sequential series of related *Event Schemas* within a larger *Episode* (Mandler & Johnson, 1977). The expectation was that these stories would be more difficult to conceptualize and recount than those in the High coherence condition since additional information is required beyond what is depicted in the pictures in order for the story to be complete and logically organized. Whereas the High coherence stories unfold in a way that is directly conditioned by the scenes in the pictures themselves, it was expected that the stories in the Moderate condition would impose greater cognitive demands on the speaker as they must draw more from working memory in conceptualizing the story. Figure 10 shows an example of a Moderate coherence picture set that was used in the study and illustrates the more temporally distant nature of the depicted events.

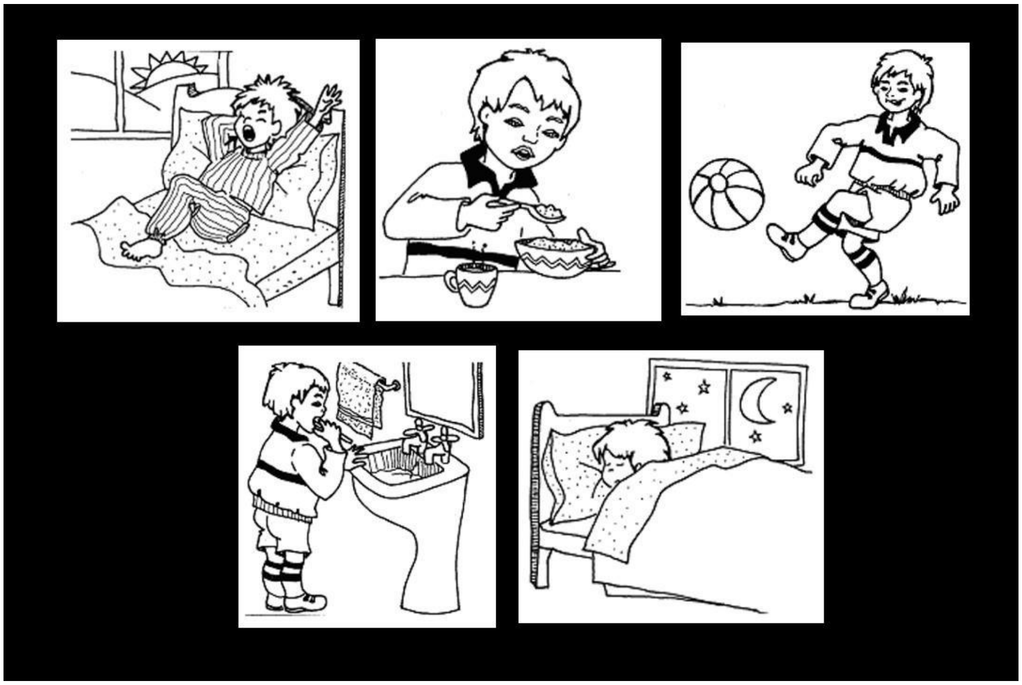


Figure 10. Moderate coherence picture set from Experiment 3 of the study.

In the Incoherent coherence condition, the pictures consisted of a series of meaningfully unrelated events which were performed by separate actors; they were comprised of completely independent *Episodes*. The expectation was that these stories would be the most difficult to conceptualize given that the speakers were tasked with synthesizing disjointed actions in formulating an otherwise coherent and well-organized story. Accordingly, it was expected that pauses would become more frequent and longer as a function of the strong conceptual boundaries at each transition between pictures. Figure 11 shows an example of an Incoherent picture set that was used in the study.

Figure 11. Incoherent coherence picture set from Experiment 3 of the study.

Two separate picture series were constructed for each coherence condition providing a total of 6 different series. Each participant was presented with 2 blocks of picture series, with each block containing one series from each coherence condition. The picture series were randomized within-block and then presented to participants in 1 of 4 pre-determined orders. The picture series orders were rotated through each participant such that if the previous participant was given the first order, then the current participant was given the second order, and so on.

5.1.2. Procedures

The picture series were presented to participants one at a time on a computer screen. The pictures within a series were arranged in order from left to right with 3 pictures on the top row and 2 pictures on the bottom row and they were presented against a black background (see Figures 9-11). At the outset of the task, participants were instructed that they would be shown several sets of pictures and be asked to tell a short story to the experimenter for each of them. The specific instructions given were as follows:

> *For each set of pictures that you see, I want you to spend about 15-30 seconds conceptualizing a short story that is based on those pictures. The story does not have to be of a certain length, but what we don't want you to do is just give simple descriptions of what is depicted in each picture. Try to make the events in the pictures go together as part of a larger coherent story. It might help to pretend that you are telling the story to a child instead of to me. For each set of pictures, once you have come up with your story you can begin telling it whenever you are ready. You can then click the button to move on to the next story when you finish with the current story. There are 6 stories total.*

5.1.3. Segmentation & Transcription

The audio recordings of each story were segmented into pause-delimited utterances using Praat software (Boersma & Weenink, 2017) and the utterances were transcribed. For segmenting purposes, a pause was defined as any period of silence of at least 100 milliseconds in duration, in line with several previous studies (e.g. Butcher, 1981; Trouvain & Grice, 1999). An additional 50 milliseconds was added to this criterion in the case of utterance-initial or unreleased utterance-final stops to account for stop closure time.

A stricter durational definition than the one used for Experiments 1 or 2 of the study was chosen primarily because of differences in the objectives of pause identification. The preceding analyses on read sentence pairs were more simply intended to characterize the distributional properties of silent intervals in speech production rather than to categorize those silences as pauses (cf. Krivokapic, 2007; Smith, 2004). By contrast, it was crucial to classify silent intervals as pauses in Experiment 3 of the study since the utterances were speaker defined. A major methodological hurdle in any study on pausing remains that there is a lack of "a well-founded theoretical rationale for what length should constitute a minimum pause (Klatt, 1980, p. 115)." The decision to adopt the criteria of 100 milliseconds in the current study was based largely on previous findings suggesting that silent intervals in the proximity of this duration are at least "psychologically functional" in the sense that they are not merely "of articulatory origin" (Hieke, Kowal, & O'Connell, 1983; see also Kendall, 2013).

5.1.4. Measures & Analyses

As before, there were 3 dependent variables: pause frequency, pause duration, and pause duration variability. Pause frequency was measured in terms of the ratio of the number of words spoken to the number of pauses within a story. The calculation of pause frequency excluded filled pauses (e.g., *uhh, umm, ok*) and clear speech disfluencies (e.g., truncated words, speech-like sounds, etc.). Pause durations were extracted automatically from the segmented audio files. The variabilities of pause durations within stories were again measured as the Coefficient of Variation of the durations. The main independent variable of interest in each of the analyses was the coherence condition defined by the picture series. In order to also account for effects of individual differences, the main of effect of speaker was also included in all of the analyses.

Separate two-way crossed factorial ANOVA models were constructed to assess the effects on pause frequency, pause duration, and the variability of pause durations. Each model contained the main effects of condition and speaker as predictors. In addition to the full models for each analysis, separate models were also constructed to test for similar effects with each pairwise combination of coherence conditions. These additional models included only the main effect of coherence condition as the lone predictor. Partial-ω^2 is reported as a standardized effect size measure for all statistically significant effects in each of the analyses.

5.2. Results

5.2.1. Story Characteristics

Table 1 provides a summary of the amount and quality of language use in the stories. In terms of the amount of language, speakers tended to produced stories that were both longer in duration and which contained a greater number of words in the Incoherent condition than they did in either the High or Moderate conditions; the story durations and numbers of words were very similar between the 2 latter conditions. In term of the quality of the language produced, the opposite pattern was found. In addition to producing stories which contained longer utterances in the High and Moderate coherence conditions than they did in the Incoherent condition, the speakers also produced stories with a higher Lexical Density (LD) in those same conditions (.661 and .679 versus .631). LD is measured as the proportion of content words to the total number of words, and it is often used in second language acquisition research as a means of characterizing the "richness" of spoken or written language production (Laufer & Nation, 1995; Lu, 2012). Lexical Density is interpreted here as a measure of the relative amount of information that speakers have packed into their utterances (see also Johansson, 2008). Taken together, these story characteristics suggest that speakers had a more difficult time formulating the language required to conceptualize a coherent story in the Incoherent condition. While they generally spoke for longer and said more when the picture prompts were incoherent, the language that they produced tended to be less productive or expressive.

Table 1. Mean (SD) values of language characteristic measures across and within Coherence conditions.

	Overall	High	Moderate	Incoherent
Story Dur. (sec.)	22.77 (13.46)	20.29 (12.47)	20.02 (10.60)	27.86 (15.41)
Words per Story	59.11 (33.69)	53.02 (31.24)	53.50 (29.11)	70.81 (37.45)
MLU (words)	6.67 (2.59)	6.72 (2.93)	6.99 (2.45)	6.33 (2.34)
Lexical Density	.657 (.057)	.661 (.049)	.679 (.057)	.631 (.054)

5.2.2. Pause Frequency

There was a significant main effect of the experimentally-defined coherence conditions on the frequency of pausing within the stories [$F(2,194) = 4.849$, $p < .01$, *partial-*$\omega^2 = .03$]. There was also a significant main effect of speaker [$F(39,194) = 4.556$, $p < .001$, *partial-*$\omega^2 = .37$]. Figure 12 shows the differences in pause frequency between the 3 conditions, where it can be seen that there were significant differences between the High and Incoherent coherence conditions [$F(1,79) = 8.568$, $p < .001$, *partial-*$\omega^2 = .04$] and between the Moderate and Incoherent conditions [$F(1,77) = 15.010$, $p < .001$, *partial-*$\omega^2 = .07$]. Speakers produced more words per pause on average in the High and Moderate conditions than they did in the Incoherent condition (High: $M = 8.59$, $SD =$

4.83; Moderate: *M* = 8.65, *SD* = 4.33; Incoherent: *M* = 7.26, *SD* = 3.16). The frequencies of pauses in the High and Moderate conditions were very similar and their difference was not statistically significant. There was a significant overall fit of the full model [$F(41,194) = 4.551$, $p < .001$, $R^2 = .38$], and the effect sizes measures suggest that this driven mostly by the main effect of speaker.

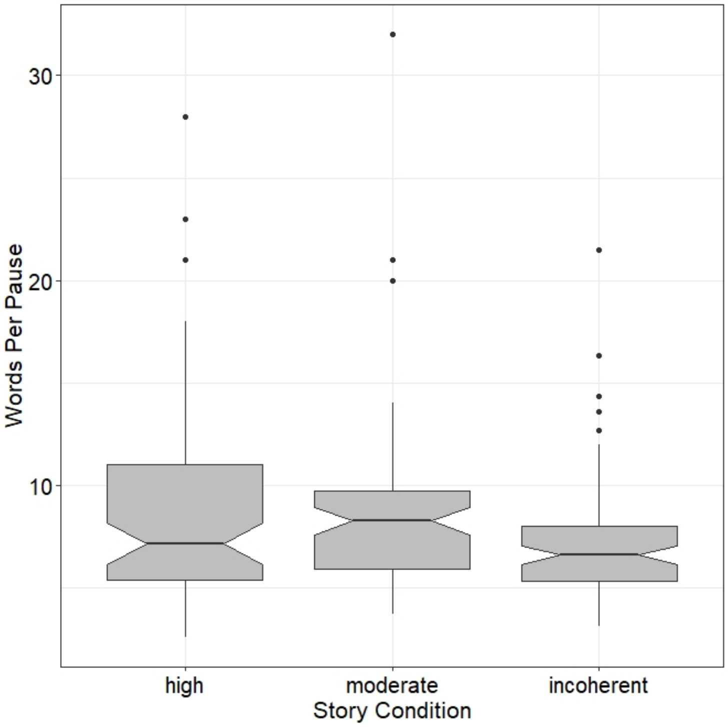


Figure 12. Pause frequency (in words-per-pause) by story coherence condition.

5.2.3. Pause Duration

There was a significant main effect of the experimentally-defined coherence conditions on the mean pause durations within the stories [$F(2,187) = 13.379$, $p < .001$, *partial*-$\omega^2 = .09$],as well as a significant main effect of speaker [$F(39, 187) = 4.032$, $p < .001$, *partial*-$\omega^2 = .34$]. Figure 13 shows the differences in mean pause durations between the 3 conditions. There were significant differences between the High and Incoherent conditions [$F(1,76) = 14.346$, $p < .001$, *partial*-$\omega^2 = .08$] as well as between the Moderate and Incoherent conditions [$F(1,77) = 24.113$, $p < .001$, $partial\text{-}\omega^2 = .12$]. Speakers produced shorter pauses on average in the High and Moderate conditions than they did in the Incoherent condition (High: $M = 0.521$, $SD = 0.195$; Moderate: $M = 0.502$, $SD = 0.167$; Incoherent: $M = 0.621$, $SD = 0.230$). The pause durations in the High and Moderate conditions were very similar and their difference was not statistically significant. The fit of the full model was also significant [$F(41,187) = 4.432$, $p < .001$, $R^2 = .38$], and the effect size measures again indicated that speaker was the stronger of the 2 main effects.

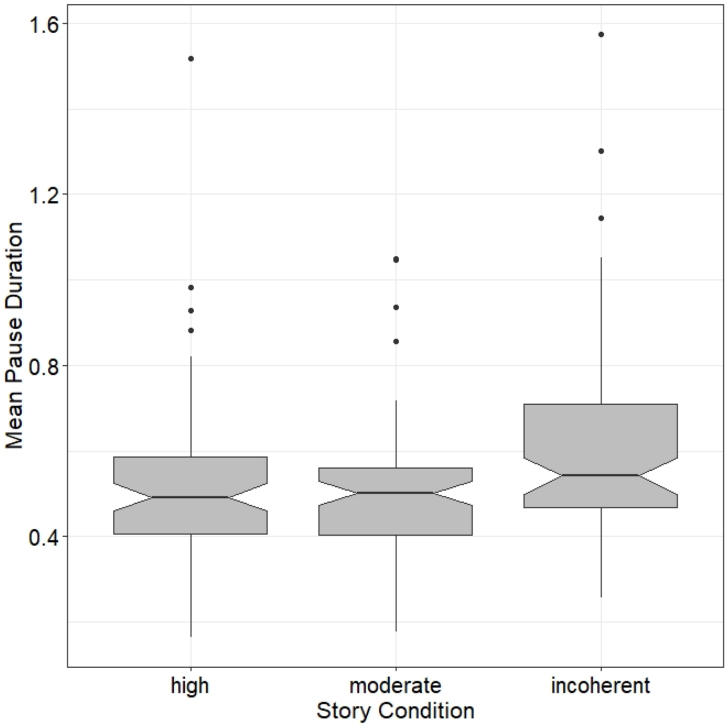


Figure 13. Mean pause durations by story coherence condition.

5.2.4. Pause Duration Variability

There was not a significant main effect of the experimentally defined coherence conditions on the variability of pause durations within the stories, but there was a significant effect of speaker [$F(39,187) = 2.239$, $p < .001$, $partial\text{-}\omega^2 = .17$]. Figure 14 shows the differences in duration variability between the 3 conditions. There are no apparent differences between the High and Moderate condition or between the Moderate and Incoherent conditions. However, there was a small but statistically significant difference in pause duration variability between the High and Incoherent conditions [$F(1,77) = 4.506$, $p = .03$, *partial*-$\omega^2 = .02$]. Speakers generally produced less variable pause durations in the High condition than in the Incoherent condition, while the variabilities of durations in the Moderate condition were intermediate (High: $M = 0.585$, $SD = 0.207$; Moderate: $M = 0.617$, $SD = 0.193$; Incoherent: $M = 0.645$, $SD = 0.217$). Finally, the full model predicting pause duration variability was significant, due almost entirely to the significant main effect of speaker [$F(41,187) = 2.226$, $p < .001$, $R^2 = .18$].

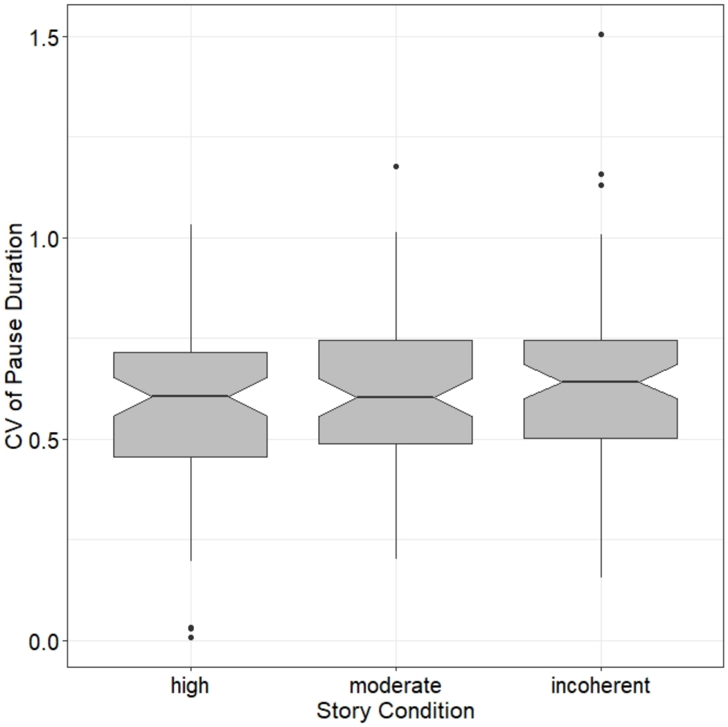


Figure 14. CV of pause durations by story coherence condition.

5.3. Discussion

The results from Experiment 3 revealed a systematic effect of story coherence on the frequency of pausing. Speakers were found to pause less frequently in the High and Moderate coherence conditions compared to the Incoherent condition. This pattern of increasing pause frequency with decreasing picture series coherence fits well with the findings of Duez (1982), who compared differences in pausing patterns between casual interviews (CI) and political interviews (PI). She found that speakers paused less frequently in CI speech than in PI speech. The difference was due at least in part to the suppression of between-phrase pauses in CI speech, where speakers were more likely to string multiple phrases together within an utterance. The different speaking styles in Duez's study can be related to the different coherence conditions of the current study in the following way. The message being conveyed in CI speech carries less weight than in PI speech, and so CI speech is likely to be less cognitively demanding in terms of message formulation, leading to fewer pauses. In contrast, language must be more carefully considered in PI speech, which in turn leads to more frequent pausing. The Incoherent story condition in the present study is more similar to PI speech in that the speaker must formulate the story in addition to simply producing it; by and large, the more coherent picture series in the High and Moderate coherence conditions formulate the story for the speaker.

The explanation of coherence effects on pause frequency as due to cognitive load is further supported by findings from Reynolds and Paivio (1968), who found that, when speakers were asked to devise and verbalize definitions for words, they paused at a significantly higher rate when the word being defined was abstract as opposed to concrete. Lay and Paivio (1969) also found that speakers paused significantly more frequently when asked to describe a cartoon as opposed to describing themselves, and even more frequently when asked to interpret and describe proverbs.

The results from this experiment also revealed similarly significant and systematic effects of story coherence on the relative durations of pauses. Once again there was little difference in mean pause durations between the High and Moderate coherence stories, but the mean pause durations were on average roughly 100 milliseconds longer in the Incoherent stories. This pattern is very similar to one found by Goldman-Eisler (1961a). Looking at the distributions of pause durations between various speaking style conditions, Goldman-Eisler found that speakers produced an increasing proportion of short pauses (< 500 milliseconds) and a decreasing proportion of long pauses (> 3 seconds) as they repeated a cartoon description task several times. Moreover, when she compared the distributions of pause durations in the cartoon description task to ones from more spontaneous speech samples (i.e. psychiatric interviews), she found that the proportion of short pauses decreased significantly in the latter and long pauses became even more common. This effect of task on duration may again be mediated by the relative cognitive demands associated with the formulation of the concepts to be conveyed as well as the planning of language required to express those concepts. These cognitive demands are even more amplified in truly unstructured and spontaneous speech situations such as interviews. As Goldman-Eisler explained, of the different speaking styles she investigated, the interviews represented “the most intellectual speech production, i.e. those requiring the highest level of verbal planning (p. 237).” Similarly, the results of the current experiment suggest that the finding of shorter pause duration in narrative elicited under the more coherent conditions reflects the relative ease of discourse planning when presented with sets of related pictures than when presented with sets of unrelated pictures.

Finally, although the results of the Experiment 3 did not reveal an overall effect of coherence condition on the variability of pause durations within stories, there was a significant difference in duration variability between the High and Incoherent conditions. Also, even though the main effect of condition was not significant, there was a systematic increase in duration variability from the High to Moderate to Incoherent conditions, respectively. Thus, there was at least a suggestion that pause duration variability tends to increase with decreasing narrative coherence. This pattern fits with findings from an early study by Henderson et al. (1965), in which they analyzed differences in the temporal structuring of speech productions between read and spontaneous speech. A key finding from that study was that the spontaneous productions had a tendency to fluctuate between "fluent" and "disfluent" segments as measured in terms of the ratios of pause durations to the durations of their surrounding utterances; the read speech productions displayed a much more regular structure with little fluctuation. Such fluctuations in fluency naturally lead to greater overall variability in the pause durations. Their study also found significant differences in the relative proportion of breath pauses between the two speaking styles, which is of interest here as it pertains to the role of cognitive processing in the results of the present study. In the read speech samples, approximately 77% of all silent pauses were accompanied by a breath, while only 34% of the pauses in the spontaneous speech samples contained breaths. This suggests that pauses occur in read speech largely as a function of the physiological need to breathe, whereas pausing in spontaneous speech varies more idiosyncratically as a reflection of higher-level processes related to language and discourse planning.

5.4. Conclusions

By and large, the results of Experiment 3 provide evidence in support of the predictions as laid out at the beginning of this chapter. There was an overall main effect of Coherence condition on the frequency of pausing, and the general trend was in the predicted direction. There was not a significant difference in pause frequency between the High and Moderate conditions (i.e. the “coherent” conditions), but there were significant differences between each of those conditions and the Incoherent condition; speakers paused more frequently in the latter condition. A similar pattern was found for pause durations, with significantly longer pauses overall in the Incoherent condition than in either of the High and Moderate conditions, in which durations were remarkably similar. Both findings support the hypothesis that narratives elicited in response to unrelated pictures require a greater amount of discourse planning than narratives elicited in response to related pictures, and that the increased cognitive processing load associated with discourse planning will lead to more and longer pauses. The hypothesized relationship between cognitive load and pause duration variability was not as evident in the results, but there was nonetheless a systematic trend for increasing variability with decreasing picture coherence. This pattern supports the assumption that pausing patterns in the more coherent conditions largely reflect the more systematic effects associated with lower-level planning and recovery factors.

As with the 2 preceding experiments, the results from this storytelling experiment suggest that a complete model of pausing behavior must also account for additional influences from high-level cognitive processing demands as they relate to effects from global discourse planning. In contrast with previous studies, these results furthermore demonstrate that the high-level cognitive effects operate not only between different speech situations but are also manifest within a single type of speech situation as a function of variable degrees of cognitive load. In the remaining chapters, the results of the 3 experiments presented thus far will by synthesized toward formalization of an exploratory model of pausing behavior in monologic narrative speech. The model was hypothesized to involve several interacting components as suggested by the current results, and the validity of model was preliminarily tested using the narrative speech samples analyzed in the current experiment.

VI. TOWARDS A MODEL OF NARRATIVE PAUSING BEHAVIOR

The results of the 3 experiments presented thus far provide clear suggestions for a model of pausing behavior. The discussion of how each factor investigated in the current study affects narrative pausing behavior is presented below in the order of the levels at which they are assumed to operate within a fuller model; a point which will be more clear by the conclusion of this discussion. The full details of the model are presented in the following chapter.

6.1. Clause Boundaries

Based on the experimental results of the current study, there is clear evidence that the output from the local language production system represents a strong primary component affecting pausing behavior in narrative speech. More specifically, there is empirical evidence that the local syntactic structure of a narrative has systematic effects on the pause frequency, pause duration, and the variability of pause durations. In Experiment 1, it was found speakers had a strong tendency to pause at the strong syntactic boundaries between the separate sentences (i.e. independent clauses). This was interpreted as suggesting that strong syntactic boundaries between successive language production units provide the most grammatical opportunities for the speaker to pause relative to the local structure of the language. The results of Experiment 2 served to provide additional evidence that the likelihood of pausing at clausal boundaries is further modulated by the relative degree of embeddedness between the clauses. As successive clauses become increasingly related it becomes increasingly rare for the speaker to pause between them. Taken together, these sets of results suggest that speakers are acutely aware of the generally hierarchical nature of local syntactic structure. This awareness is evidenced by the fact that the speakers in this study generally obeyed the constraints of the syntactic structure in their choices of where and when to pause.

A comparison of the pauses produced in the Experiments 1 and 2 in this study also provides strong evidence that the local syntactic structure contributes directly to the durations of pauses. The pauses which occurred at the strong sentence boundaries in Experiment 1 were on average more than 3 times as long than the pauses between the related clauses in Experiment 2. Even when controlling for effects from the lengths of clauses around the boundary by restricting the comparison to the SS pairs in Experiment 1, the mean pause duration between the 12-syllable unrelated sentences was 422 milliseconds compared with only 155 milliseconds between the 12-syllable related clauses. Thus, not only are speakers attuned to the relative grammaticality of a given inter-clausal pause, but moreover this awareness gets reflected in the durations of the pauses that they produce at boundaries of differing strengths.

The results from the current study provided less clear suggestions about the role that local syntactic structure might play in influencing the variability of pause durations in a model of narrative pausing behavior. Taken as a whole, the duration variability was found to be slightly higher for pauses between unrelated sentences than it was between related clauses (Experiment 1: M = .79; Experiment 2: M = .68), suggesting at least some degree of effect of boundary strength. The results from Experiment 2 also provided further evidence that pause durations become more variable with increasing boundary strength between related clauses, but that inference must be approached with caution given the relative lack of pausing which occurred between clauses in each condition overall. Nonetheless, it is reasonable to assume that the general trend of increasing pause duration variability with increasing syntactic boundary strength is reliable given that pause durations in general are shorter between more cohesively connected clauses. In other words, if pauses at weaker syntactic boundaries tend to be shorter than those at stronger boundaries, there is an expectation that the overall range of durations at the weaker boundaries will be smaller resulting in less variability overall. In this way, a narrative produced with more complex local structure is expected to contain less variable pause durations as a function of more frequent embedding of clauses (i.e. more weak clause boundaries).

6.2. Speech Planning Domain

In addition to the primary effect of clause boundaries on the opportunity to pause, the results of the current study also suggest that speech planning processes play a role in pausing behavior. Experiment 1 demonstrated that pausing patterns vary systematically as a function of the amount of language about to be produced. When speakers were tasked with producing longer sentences, they were more likely to pause and to pause for longer than before shorter sentences. This was interpreted as reflecting the relative amount of planning required in each condition. This interpretation was also supported by the utterance length patterns found in Experiment 3. The Incoherent stories were predicted to be more difficult to plan, and speakers produced significantly shorter utterances on average in those stories than in the more coherent ones. This might simply reflect the greater cognitive demands imposed by the higher-level discourse planning processes, but presumably it also indexes at least some degree of differences in the speech planning processes between the conditions.

For the purposes of the preliminary model, it was assumed that the effects from speech planning processes operate at a level at which they mediate the direct effect of an opportunity to pause. As discussed above, it is clear that clause boundaries play an essential role in determining both where pauses will occur and how long those pauses are likely to be. However, the language expressed in narrative speech is not coincidental, and some degree of planning is required by the speaker prior to its production. It has been shown in previous studies that disturbances in speech planning abilities result in meaningful differences in language production between disordered and typical speakers (e.g. Costello & Warrington, 1989). All of the speakers in the current study reported no history of language disorder, yet it is reasonable to assume that they differ to some degree in terms of their respective speech planning abilities. An obvious consequence of such a discrepancy would be differences in the overall complexity of the language produced. A higher capacity for speech planning is expected to allow for the planning of more complex language structures. For this reason, it is proposed that speech planning processes operate as mediators of the direct effect of clause boundaries on pausing behavior.

6.3. Discourse Planning

The current study also provides empirical evidence that higher-level discourse planning processes play a meaningful role in shaping narrative speech pausing patterns. This evidence comes specifically from the spontaneous narrative production task in Experiment 3. When speakers were asked to tell stories based on a set of coherently related pictures, they were found to pause less frequently on average than when the relationship between pictures was incoherent. In addition to pausing more frequently, it was also found that the durations of those pauses were significantly longer overall. There was less evidence of a coherence effect on the variability of pause durations, but the results nonetheless displayed a systematic trend for increasing variability with decreasing coherence.

In light of findings from previous studies, these results were not surprising. As noted by Kintsch & van Dijk (1978), a coherently produced story is characterized by not only a local "hierarchical sequence of propositions in which coreferential expressions occur (p. 365)," but also a more global hierarchical structure in which those coreferential expressions are further embedded within broader discourse topic or topics. A greater amount of embedding of events or ideas has systematic effects on relative boundary strength between utterances, which in turn results in meaningful variation in pausing patterns. It is with this hierarchical characterization of narrative discourse structure in mind that it is assumed that higher-level discourse planning processes also operate at a level at which they mediate the direct effects from the language production system. While speech planning abilities shape the complexity of the language structure at a local clause-by-clause level, discourse planning processes are assumed to do the same at the global discourse level. As the discourse structure is built up through the embedding of language units at successive levels, a speaker's relative ability to plan fluent narrative productions is assumed to have an effect on the structure of the language at any given level of the discourse structure.

6.4. Individual Differences

The discussion to this point has been aimed at making a case for inclusion of 3 components in a preliminary model of narrative pausing behavior. The justifications for inclusion of each of those components have been presented specifically with reference to the experimental findings of the current study. Before proceeding with formalizing and testing the model, however, it is argued here that an additional component related to individual cognitive processing abilities should also be included.

It is proposed that individual cognitive processing abilities can be appropriately represented in the model in terms of individual differences in working memory capacity (WMC). More specifically, WMC is assumed to play the role of a moderator on the relationship between clause boundary strength and speech and discourse planning. Put another way, the relationship is conditional on influences from language-independent working memory abilities. In the existing literature, it has long been customary to study the link between language production and WMC from the perspective of second language learning. For example, numerous studies have demonstrated a positive relationship between WMC and oral production proficiency among L2 language learners (e.g. Ahmadian, 2013; Linck, Osthus, Koeth, & Bunting, 2014). While perhaps less common, others have also investigated the link between WMC and speech-language production in L1 speakers. For example, Daneman (1991) found that performance on a Speaking Span assessment of WMC was positively correlated with general fluency in both spontaneous and read speech. Similarly, Daneman & Green (1986) found that there was a relationship between WMC and word-finding effects such that speakers with lower WMC had more difficulty selecting alternative words in a read speech task. These results are unsurprising since speech-language production is inherently a cognitive task which requires a balancing of limited processing, storage and recall resources (see, e.g., Alario, Costa, Ferreira, & Pickering, 2006; Lively, Pisoni, van Summers, & Bernacki, 1993; McClain & Goldrick, 2018).

At the local utterance level, a storyteller needs not only to formulate the primary message to be conveyed, but also decide how much and what type of elaboration is necessary to effectively convey that message to the listener. This means that the speaker must be attuned to their internal language production goals while simultaneously attending to the external needs of the person listening. In this way, it is predicted that an individual's WMC will moderate their speech planning capacity and so also by extension the opportunities that will emerge within their narrative speech for pausing. Looking more globally at the discourse level, a speaker must also store and access information that is relevant to the construction of a coherent narrative as a whole. Among other things, they must track the information status of various referents as the story unfolds, such as whether they are being introduced for the first time, are being re-introduced, or are currently in focus (see Colozzo & Whitely, 2014; Schneider & Hayward, 2010). It is also necessary to continuously manage the major themes or goals of the narrative, as well as the types of relationships which hold between them. With this in mind, it is also predicted that an individual's WMC will moderate the relationship between their discourse planning capacity and the opportunities that emerge for pausing with clause boundaries.

6.5. Conclusions

In summary, the experimental results of the current study suggest that an ecologically valid model of pausing behavior in narrative speech should at least include separate components which account for influences from clause boundary strength, speech planning processes related to message length, and higher-level discourse planning processes. Furthermore, an argument has been made that the relationships between these components are likely conditional on an additional cognitive processing factor in the form of WMC. Notably absent from this characterization of the model is a component accounting for effects from individual differences in respiratory functioning. Based on the results from Experiment 1, it appears undeniable that pausing patterns are systematically influenced by the physiological need for respiratory recovery. Thus, a truly complete model of pausing will ultimately need to account for this effect. Unfortunately, this was not possible in the current study due to the lack of a reliable means of characterizing individual respiratory functioning from the data at hand.

VII. EXPLORATORY MODEL OF NARRATIVE PAUSING BEHAVIOR

A preliminary model of pausing behavior is proposed and tested on the narrative speech elicited from the High and Moderate coherence conditions in Experiment 3. The model seeks to characterize pausing patterns as emergent from moderating and mediating factors investigated in this dissertation. The results are preliminary due to the small number of participants in the current study ($N = 40$). The type of model proposed here requires sample sizes of hundreds of participants. To this point, Fairchild & McKinnon (2009) note that "models that simultaneously examine mediation and moderation effects are at … [a] disadvantage as they involve several interaction terms as well as estimation of indirect effects (p. 96)." This is especially problematic for analyses of conditional effect of the moderator variable given that participants must be split into subgroups (see section 7.8).

Before moving on to the specific details of the proposed model, it will be useful to first provide a brief and general overview of Moderated Mediation Analysis within which the preliminary model of pausing behavior is proposed. A representation of the typical architecture of a typical Moderated Mediation model can be seen in Figure 15. This model relies on a set of assumptions which can be summarized as follows (for a more complete description, see Jose, 2013). The first assumption is that there is some IV (X) for which there is a significant main effect on the DV(s) of interest (Y). This IV is considered to have a Direct Effect on the DV. In other words, absent any other effects in the model there is a predicted direct relationship between the X and the Y. In addition to the Direct Effect, it is assumed that there are 1 or more mediating Indirect Effects (M) which serve to further explain the relationship between the IV and DV. For example, a hypothetical Moderation Mediation model might assume that there is a direct effect of age on the likelihood of heart disease. Furthermore, it might also assume that the relationship between age and heart disease is mediated by an additional effect of smoking. That is, the effect that age has on the likelihood of heart disease is in some way dependent on whether a subject is also a smoker. The final assumption in a Moderated Mediation model is that the Indirect Effect is also Conditional on some additional moderating effect (*W*). In the hypothetical example, one might assume that the Indirect Effect of smoking is Conditional on socioeconomic status (SES); the degree to which smoking has an effect will differ between participants with lower or higher SES.

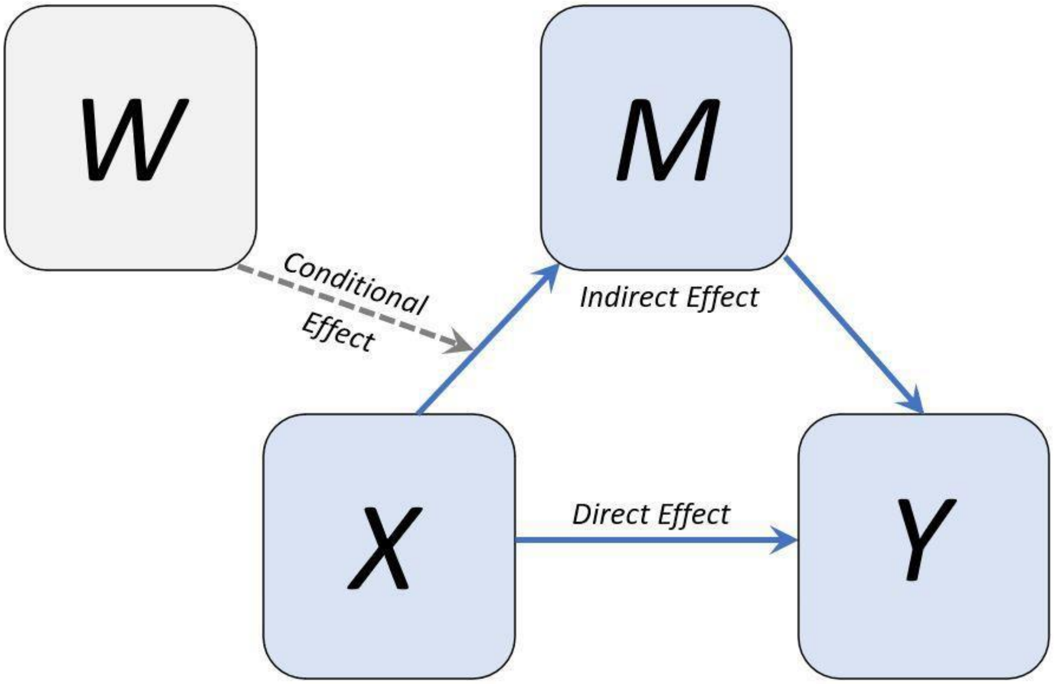


Figure 15. Architecture of a generic Moderated Mediation model.

The details of the preliminary model of narrative pausing behavior that is proposed here is shown in Figure 16. In this model, clause boundary strength is assumed to have a Direct Effect on the pausing outcomes in that such boundaries provide opportunities for pausing. The model also assumes that the Direct Effect of boundary strength is mediated by speech and discourse planning processes. For example, as show in Chapter 4, the speech planning domain is larger when adjacent clauses are highly cohesive; speakers rarely pause at relative clause boundaries. Finally, it is proposed that the Indirect Effects are moderated by a Conditional Effect of WMC. Thus, the relationship between boundaries and planning are predicted to vary as a function of individual WMC differences. Included in the representation shown in Figure 16 are specific measures which are assumed in the present analyses to reliably index each of the effects in the model. The justifications for each language-related measure as an index of its respective component are presented next.

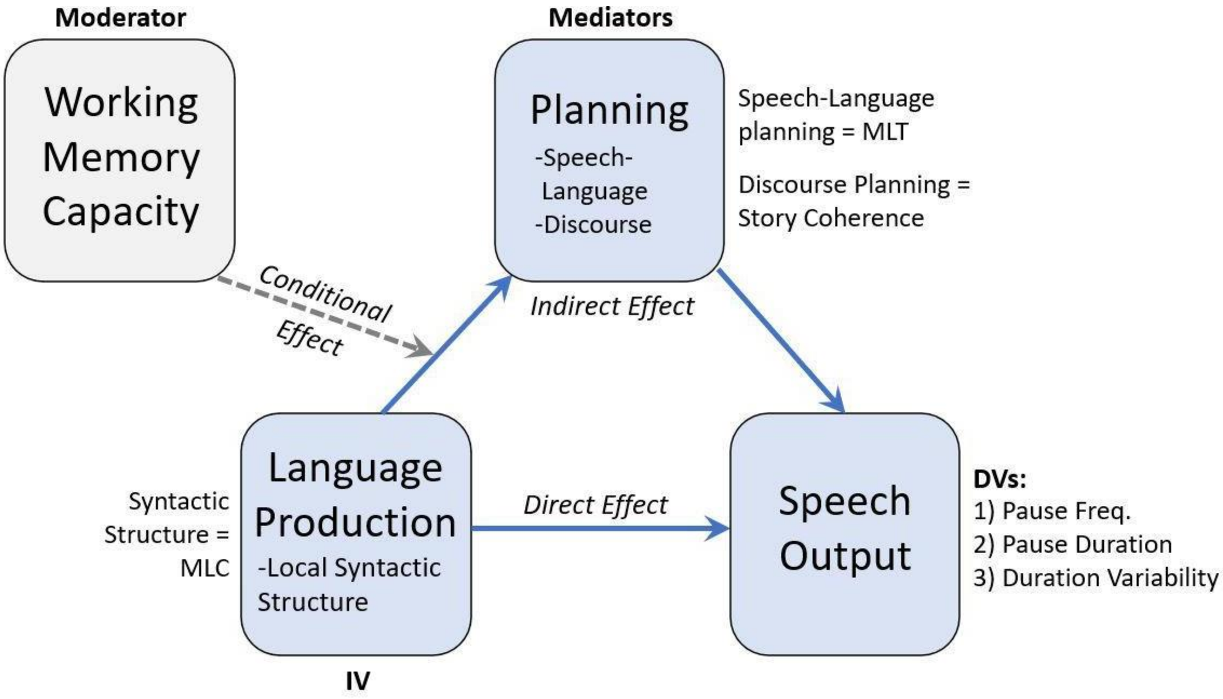


Figure 16. Formal representation of the preliminary model of narrative pausing behavior.

7.1. Indexing Syntactic Boundaries

In the analyses of the preliminary model, strong syntactic boundaries are indexed by a measure of the mean length of clauses (MLC) within a narrative. As already discussed, speakers have a very strong preference for pausing at sentence/clause boundaries and are much less likely to pause at other within-clause locations (see e.g. Bailly & Gouvernayre, 2012; van Donzel & Koopmans-van Beinum, 1996). Thus, clause boundary pauses can be characterized as “grammatical,” in contrast to the less natural “ungrammatical” pauses that occur within clauses. For this reason, the expectation might be that higher MLC will correspond with less frequent pausing since shorter clauses afford more frequent opportunities for grammatical pausing. On the other hand, it is reasonable to assume that MLC also corresponds with the relative complexity of syntactic structure, such that more frequent embedding of subordinate clauses will correspond with lower MLC values. This expectation follows partly from the fact that speakers often either drop shared arguments from embedded clauses or reduce them to pronominal forms (see, e.g., Givon, 1983), which naturally leads to fewer words being expressed. Additionally, there are both cognitive and physiological limitations to the amount of language that can be planned and produced as a single cohesive unit in spontaneous speech (see Chafe, 1987; Wlodarczak & Heldner, 2017). If multiple clauses are embedded within a single utterance, then the lengths of the individual clauses will necessarily become shorter to accommodate these limitations. Thus, there is the suggestion that both higher and lower MLC values might predict less frequent pausing. If a speaker tends to chain longer independent clauses (i.e. higher MLC), then they will have fewer opportunities for grammatical pausing relative to the number of words produced. If instead a speaker tends to chain several shorter embedded clauses together within their utterances (i.e. lower MLC), then they will have relatively fewer strong syntactic boundaries at which to pause. However, it is also reasonable to assume that while embedded clauses are expected to be shorter than independent clauses, they are not expected be so much shorter that a chain of embedded clauses will not result in a longer utterance than a single independent clause. This being true, a lower MLC value is predicted to correspond with less frequent opportunities for grammatical pausing, and a corresponding decrease in overall pause frequency.

Following from the above prediction for pause frequency, the predicted effects of MLC on both mean pause duration and pause duration variability are more straightforward. In Experiment 1, pause durations were found to be systematically longer when they followed longer 24-syllable sentences than when they followed shorter 12-syllables sentences. The suggestion from the analyses was that the increased durations were tied to the physiological need for breathing because longer speaking time leads to a more depleted air supply. Since respiratory intake necessarily requires a certain amount of time to complete, breath pauses on average will be longer than non-breath pauses. Thus, if lower MLC values correlate with less frequent opportunities for grammatical pausing, it is predicted that there will also be a negative relationship between MLC and mean pause duration. In addition to having longer durations, the pauses following long sentences were also found to have less variability of those durations than did pauses following short sentences. This finding was again interpreted as reflecting speech breathing effects; breath pauses serve a more systematic function than do non-breath pauses. As follows, it is predicted that lower MLC values will correspond with less variable pause duration in the model.

7.2. Indexing Speech Planning Domain

In the preliminary pausing model, it is assumed that the scope of speech planning can be indexed by a measure of the mean length of T-units (MLT) within a narrative. The notion of a T-unit has been around in second-language studies since the 1960s (see Hunt, 1965), and it has been widely used as a general measure of language complexity. In the most general terms, a T-unit is a unit of language which is “grammatically capable of being terminated with a capital letter and a period” (Hunt, 1965, p. 37). More specifically, a T-unit consists of a main clause plus any accompanying subordinate clauses. A more complex T-unit in which multiple clauses have been embedded within each other is expected to be longer than one comprised of single main clause. This presumption is supported by evidence that MLT is strongly correlated with the mean number of clauses per T-unit (Lu, 2011). Thus, a higher MLT value is interpreted in the current study as reflecting a greater tendency of the speaker to elaborate on the events in the narrative, rather than simply producing sequential chains of individual events. In other words, elaboration requires the planning of multiple inter-related ideas to convey more complex messages, as opposed the one-by-one planning of individual ideas.

The prediction was that MLT would be negatively correlated with pause frequency in the stories. This prediction follows from the above assumption that the production of longer T-units reflects the planning of more complex syntactic structure in the form of clausal embedding. As with the above discussion of MLC, this structural complexity is expected to result in fewer opportunities for pausing at stronger boundaries and correspondingly to result in less frequent pausing. Likewise, less frequent pausing is expected to result in generally longer pauses overall, so MLT is predicted to be positively correlated with pause duration. Finally, the expectation that opportunities for grammatical pausing will decrease with increasing MLT suggests that those pauses which do occur will be more functionally systematic. Thus, it is predicted that the variability of pause durations will decrease with increasing MLT.

7.3. Indexing Discourse Planning

In the preliminary model, it is assumed that the scope of discourse planning can be indexed by a measure of judged narrative Coherence. More specifically, it is expected that narratives which are well-structured hierarchically around the events and themes of the story will be judged as more Coherent than those in which events are simply conveyed sequentially without elaboration. As noted in previous studies, one consequence of hierarchically embedding language units within each other is greater variability of juncture strengths within the discourse (see Gee & Grosjean, 1983; Swerts, 1997). This being true, a coherently produced story will be characterized by fewer strong syntactic boundaries relative to the total number of boundaries. As strong syntactic boundaries represent the most grammatical opportunity for pausing in discourse, more coherent stories were predicted to have less frequent pausing than less coherent ones. Similar to the prediction for the MLT measure of local structure, more globally coherent narratives are also predicted to have generally longer pauses. As opportunities for grammatical pausing decrease, the likelihood of breathing during the pause will increase leading to longer pause durations. The systematic function of breath pauses following longer stretches of speech productions is then predicted to result in less overall variability in the durations of those pauses.

7.4. Indexing Individual Differences

Individual cognitive processing abilities were indexed using a composite measure from the 2 separate WMC assessment tests that were administered in the current study. This measure of WMC was intended to reflect an individual’s abilities to simultaneously juggle the processing storage and recall of various types of information. Such capacity is relevant to narrative speech production as the speaker must construct and produce meaningful language units at various hierarchical levels. The hypothesis was that an individual’s relative ability to plan language at both the local syntactic and global discourse levels would be constrained by their WMC. It was predicted that speakers with higher WMC would produce more complex language structures and so pause less frequently than those with lower WMC. This decrease in pause frequency was predicted to result in longer pauses overall for the higher WMC speakers given the experimentally demonstrated effects of respiratory recovery and speech planning. Speaking for longer increases the drive to inspire and the planning of longer utterances takes longer than the planning of shorter ones. The decrease in pause frequency for higher WMC speakers was also predicted to lead to less variability of pause durations.

7.5. Methods

7.5.1 Speech Samples

The speech samples used to evaluate the preliminary model consisted of a subset of the semi-spontaneous narrative productions that were collected in Experiment 3 of the current study. This subset was comprised on the 4 “coherent” stories produced by each participant (i.e. the High and Moderate stories). Four of these stories did not contain a pause and were therefore excluded, so the analyses reported below were conducted on a total of 156 stories. The choice to exclude the Incoherent condition stories from the analyses was motivated by 2 main considerations. First, the goal was to run a preliminary test of the model on speech samples. Second, there was a concern that pausing behavior would be more similar between speakers in the Incoherent condition leading to less meaningful differentiation between speakers overall with the inclusion of those stories.

7.5.2. Direct Effect: MLC

The primary IV in each of the moderated mediation analyses was a measure of mean clause length (MLC). This measure corresponds to the Direct Effect component in the model shown in Figure 16 (above). To calculate MLC, the transcripts of each narrative were first segmented into individual clauses. A clause was defined as "any unit containing a unified predication, whether in the form of a verb or adjective (Berman & Slobin, 1994, p. 26)." As an example, the following excerpt contains a total of 3 clauses (delimited with a slash):

Gene's really good / at finding fire materials / and starting a fire.

The 1st clause contains a predicate adjective while the 2nd and 3rd clauses both contain verbal predicates. Note that while the subject of each clause is the same ("Gene"), and so not explicitly expressed in clauses 2 and 3, they still represent separate clauses under the above definition. In the case of complex verbs forms such as "wanted to go," the clausal segmentation depended on the presence or absence of an intervening non-VP. In the given example, the VP "wanted to go" was treated as a single verbal predicate. If there were an intervening NP such as in "wanted him to go," then the verbs "wanted" and "to go" were treated as separate verbal predicates and coded as belonging to separate clauses.

Once all transcripts were segmented, the number of fluently produced words was extracted for each clause. Filled pauses (e.g. *uhh*, *umm*) and truncated words (e.g., *tr-*, *spo-*) were excluded from the word counts. For each speaker, a single measure of MLC was then calculated as the mean number of words per clause across all 4 of their narratives.

7.5.3. Mediating Effect I: MLT

A measure of mean length of T-unit (MLT) was calculated to index the scope of speech planning per speaker. This measure corresponded to 1 of the 2 Mediating Effects in the preliminary model (Figure 16). As originally defined by Hunt (1965), a T-unit consists of "one main clause with all the subordinate clauses attached to it (p. 20)." At a minimum a T-unit must contain a single independent clause, but it may contain more clauses. To calculate the MLT measure, the transcripts of each story were first segmented into individual T-units. For this purpose, adverbial clauses, relative clauses, and other types of dependent clauses were considered to be subordinate.

Adverbial clauses were defined as any clause which contained an adverb (e.g. *after*, *as soon as*, etc.) as well as its own unified predication.. It was not uncommon for some speakers to embed multiple adverbial clauses within a single T-unit, as in the following example:

*He falls asleep / **after** brushing his teeth / **as soon as** he gets in bed*

Relative clauses were defined as any unit with a clause-initial relativizer (e.g. *that*, *which*, etc.) and its own unified predication. The following example of a T-unit contains 1 main clause and 1 relative clause:

*This is a man / **who** is a baker*

Other types of dependent clauses were defined as those for which some argument was dependent on the main clause for resolution. The majority of these were clauses which were coordinated with the main clause and without explicit mention of a subject (i.e. zero-anaphora). In the follow example of a T-unit there is 1 main clause and 2 dependent clauses:

They bought some wood / and Ø cut it up / and Ø made a birdhouse

Any other clause which could either stand alone as an independent clause or which was coordinated with another clause but contained an overtly expressed subjects was considered a main clause. In the following example there are 2 clauses which also represent 2 separate T-units:

This is a baker / and he is baking cookies.

Similar to the calculation of MLC, the number of fluently produced words was extracted for each T-unit in the segmented transcripts. Once again, filled pauses and truncated words were excluded. For each speaker, MLT was then calculated as the mean number of words in the T-units from all 4 of their stories.

7.5.4. Mediating Effect II: Narrative Coherence

A measure of narrative Coherence was also calculated as a separate mediating factor and was intended to index discourse planning abilities across speakers. The Coherence measure was based on judgments by naïve raters independent of the experimentally defined Coherence conditions. It was expected that stories which were more well-structured hierarchically around the overall themes of the narrative would be judged as more coherent. The details of the Coherence judgment collection procedure are as follows.

7.5.4.1. Participants

A total of 80 workers were recruited through Amazon's Mechanical Turk crowdsourcing platform to provide ratings of Coherence for each of the stories. Participation in the task was limited to workers residing in either the United States or Canada who had previously completed a minimum of 5000 HITs ("Human Intelligence Tasks") with an acceptance rate of at least 95%. Each worker was paid $1 for completing the task. Workers were given up to 35 minutes with which to complete the task, and the average time to completion was approximately 14 ¼ minutes (*SD* = 9 minutes). One of the initial 80 workers failed to provide ratings for 3 stories so an additional worker was recruited to rate those same stories. The ratings of the 1st worker were replaced by those of the 2nd worker.

7.5.4.2. Procedure

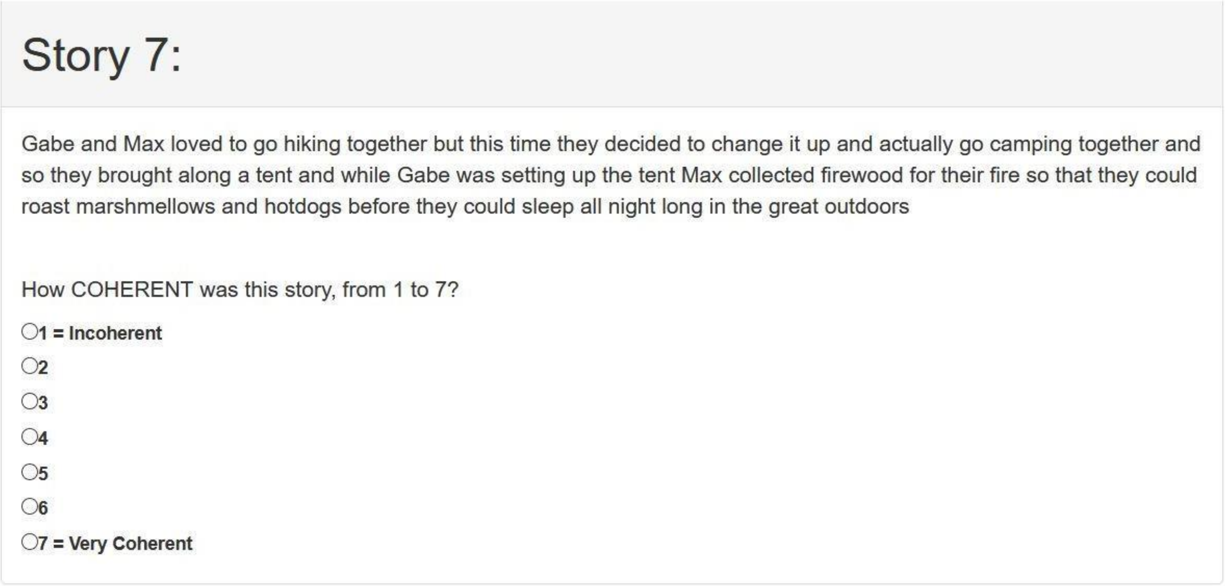
Prior to the collection of coherence judgments, the order of the 160 stories was randomized and then blocked into 8 different subsets consisting of 20 unique stories each. Each worker was presented with 1 subset of 20 stories and asked to rate them on a 7-point scale according to their level of Coherence. A rating of 1 corresponded to an "incoherent" story while a rating of 7 corresponded to a "very coherent" story. It is more common to use a 5-point likert scale in Linguistics and Psychology research, but the use of a 7-point scale has been found to yield slightly more reliable results (Lee & Paek, 2014; Preston & Coleman, 2000). It has also been argued that providing a larger variety of response options is likely to yield more fine-grained responses which are thus more differentiating (Joshi, Kale, Chandel, & Pal, 2015).

The task instructions provided participants with a general characterization of narrative coherence, as follows:

> *A coherent story is one that is well-organized around the theme or themes of the narrative, and in which the events are structured in a manner that is easy to follow. Coherent narratives also tend to be ones that are interesting or creative, and in which the narrator elaborates on the themes and events rather than simply stating them as a series of facts.*

The global organization of themes and the local structuring of events in the first part of this characterization correspond to Kintsch & van Dijk's (1978) notions of narrative macro- and micro-structure respectively (see also van Dijk & Kintsch, 1983). The relationship of creativity and elaboration to narrative coherence correspond to the characterization provided in the Test of Narrative Language assessment (Gillam & Pearson, 2004), which adopts a Story Grammar framework (see Stein & Glenn, 1979). Importantly, participants were also advised that the stories they would be rating were transcripts of natural speech productions, and so they would occasionally contain speech errors (e.g., truncations, repetitions, etc.). They were asked to ignore such errors and to base their ratings solely on the content of the stories. The goal was to ensure elicitation of ratings that were based principally on the conceptual discourse structure of the narratives and not on more general speech fluency issues.

Each subset of stories was rated by 10 separate workers. The stories were presented one at a time, and the participants could scroll from one to the next at their own pace. This meant that it was possible for participants to rerate previously rated stories once they were exposed to more examples, but they were not explicitly instructed to do so. Figure 17 shows an example as it appeared to participants.



Story 7:

Gabe and Max loved to go hiking together but this time they decided to change it up and actually go camping together and so they brought along a tent and while Gabe was setting up the tent Max collected firewood for their fire so that they could roast marshmellows and hotdogs before they could sleep all night long in the great outdoors

How COHERENT was this story, from 1 to 7?

○1 = Incoherent
○2
○3
○4
○5
○6
○7 = Very Coherent

Figure 17. Example story as seen by participants in the Coherence rating task.

7.5.4.3. Coherence Measure

Once all ratings were collected, each individual story was given a composite coherence score, calculated simply as the mean rating from the 10 different raters for that story. For each speaker, an overall coherence score was then calculated as the mean composite coherence score for that speaker's 4 stories. The latter score was used as the measure of the Coherence in the Moderated Mediation analyses reported below.

7.5.5. Moderating Effect: WMC

As means of characterizing Moderating effects of individual differences in cognitive abilities in the preliminary model, all participants in the study completed 2 separate WMC assessment tests. The 2 tests were administered as the first and last tasks of each study session respectively, and the order in which they were given alternated from one subject to the next. The WMC values that were used in the models were calculated as the average proportion of correct responses between the 2 tests. The details of each WMC assessment test are described as follows.

7.5.5.1. Counting Span Assessment

A Counting Span assessment test was administered to obtain a general measure of WMC. As a type of complex span task, this test was designed to require storage and recall of primary information while simultaneously engaging in a secondary processing task (see Engle, Tuholski, Laughlin, & Conway, 1999; Mathy, Chekaf, & Cowan, 2018; Unsworth & Engle, 2007). The Counting Span task in particular is visual in nature and involves very minimal linguistic load (see Danahy, Windsor, & Kohnert, 2007); it was chosen with the aim of obtaining a measure of WMC that is independent of verbal WMC. More specifically, the goal was to characterize each participant’s ability to effectively split available cognitive resources between different types of concurrent processing tasks. Such an ability is relevant to the production of narrative speech in that planning must occur concurrently at both the levels of local speech-language and global discourse structures.

Materials. During the administration of the test, participants were shown sequences of pictures and asked to count aloud the numbers of target objects that were contained in each picture. At the conclusion of a picture sequence, they were tasked with recalling all of the target object counts in the order that they appeared. Each picture contained between 2 and 7 targets objects (blue stars), along with twice as many non-target objects (red circles). For example, a picture containing 4 targets objects would also contain 8 non-target objects. There were an equal number of pictures for each number of targets.

The objects were arranged randomly without overlap within a square outline in each of the pictures. Previous studies on WMC in children have used a similar Counting Span task, but with a range of 3-8 target objects (see, e.g., Ransdell & Hecht, 2003; Towse, Hitch & Hutton, 1998). However, it has been previously found that accurate recall of stimulus items is constrained by the amount of time that a subject is required to retain those items in memory, and a larger number of objects to be counted necessarily takes a longer time to count (see Barrouillet & Camos, 2001; Towse & Hitch, 1995). The decision to instead use 2-7 target objects in the current study was aimed at minimizing these unrelated effects of temporal decay to the extent that it was possible.

Procedure. To ensure that participants understood the task directions, the assessment test began with 2 practice sets that each consisted of a 2-picture sequence. Following successful completion of the practice items, participants then completed 4 blocks of test items, with each block containing 3 sets of picture sequences. In the first testing block, each set consisted of 3 pictures, and the set size was increased by 1 picture in each successive block. Thus, the final testing block contained 3 sets of 6 pictures. The pictures within each set and block were presented to all participants in the same pre-determined randomized order. It was important that all participants be given the pictures in the same order because it has been shown that scores decrease systematically with increasing numbers of objects in the final picture of a set, and also that scores are strongly correlated with the overall number of objects in the entire picture set (Towse et al., 1998).

The assessment was scored by assigning a 1 to each correct picture count, *in the correct order* of appearance; incorrect responses were scored as 0. For example, if a participant was shown 3 pictures containing 4, 2, and 7 targets respectively, a response of

4, 7, and 2 would yield a total of only 1 point since the 2nd and 3rd picture counts were recalled out of order. The requirement of recalling in the correct order was intended to reduce the likelihood of correct responses resulting from simply guessing a number from the limited set of possible numbers (i.e. 2-7). The sum of the picture set scores was divided by the total number of pictures to yield a proportional measure of WMC which ranged from 0 (none correctly recalled) to 1 (all correctly recalled).

7.5.5.2. Competing Language Assessment

All participants were also administered a Competing Language assessment of WMC. This test was more specifically intended to obtain a characterization of verbal WMC. Verbal WMC is relevant to the current study in that it is clearly linked to the speech and language production system (see Acheson & McDonald, 2009). The materials and procedures used for this task were taken from Gaulin & Campbell (1994) and are described as follows.

Materials & procedures. In this task the experimenter read series of simple 3-word sentences to the participant, and they were first tasked with judging the veracity of each sentence by responding aloud with either True or False. For example, if the experimenter read the sentence “Bananas are blue” then the participant was expected to respond with “false.” At the conclusion of a set of sentences, the participant was then asked to recall aloud the last word from each sentence. The test began with 2 practice sets which each consisted of 2 sentences to ensure that the task directions were clear to the participant. Following successful completion of the practice items, the participant then completed 6 blocks of testing items which each contained 2 sets of sentences. In the first test block each set consisted of only a single sentence, and the number of sentences per set was increased by 1 in each successive block. Thus, in the sixth and final block each set contained 6 sentences. Within each block, exactly half of the sentences were true statements and the other half were false statements. Prior to the start of each block, the experimenter explicitly informed the participant of how many sentences would be in each set.

For each sentence, a participant's response of True or False was recorded and correct responses were given a score of 1. Additionally, each sentence was scored for recall of the final word, with correct responses given a score of 1 and incorrect responses given a score of 0. In this way, each sentence could receive a possible score of 0, 1 or 2 points. The maximum score for the test was 84 points (2 points times 42 sentences). As with the Counting Span test, the final score for the Competing Language test was calculated proportionally as the number of points that the participant scored divided by 84.

7.5.6. Dependent Variables: Pausing

The dependent variables used in each of the Moderated Mediation analyses mirror those investigated in each of the experimental parts of the current study: pause frequency, pause duration, and pause duration variability. Pause frequency was calculated as the MLU across all 4 stories for each speaker. As with the measure of MLC and MLT, the total number of fluently produced words was extracted for each pause-delimited utterance and this number was then divided by the total number of for the speaker. Note that MLU is a measure of words produced per pause, and so a higher MLU value corresponds with less frequent pausing. For the mean pause duration measure, the total duration of pauses was extracted across all 4 stories for each speaker and then divided by the total number of pauses. As before, the duration variability measure was calculated as the Coefficient of

Variation across all pause durations per speaker (CV_{dur}). The SD of all pause durations was first calculated and then that value was divided by the mean pause duration that was calculated above. A higher CV_{dur} value corresponds with more variable pause durations.

7.5.7. Moderated Mediation Model Analyses

The moderated mediation analyses were performed in 2 separate phases with R software in order assess effects between the various relationships within the full model. In the second stage, the other overall performance of the model was assessed with analyses performed using the mediation library (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014). Separate mediation analyses were conducted for each combination of mediators and DVs for a total of 8 models. In each model, all variables that were included in an interaction effect were first mean-centered in order to facilitate interpretation of the parameter estimates (see Aguinis & Gottfredson, 2010).

7.5.7.1. Phase I: Linear Regression Model Analyses

In the first phase of analysis for each mediation model, 2 separate linear regression models were fit. The first model was meant test for relationships of both the IV and the moderator variable with the mediator variable. For this, a multiple linear regression model was constructed with the main effects of the IV and moderator as predictors of the mediator (i.e. the Mediator Model). The interaction between the IV and moderator was also included as a predictor. The purpose of the second model was to test for effects of the IV, moderator and mediator on the DV (i.e. the Pausing Model). A new multiple linear regression model was constructed with the main effects of the IV, moderator and mediator as predictors of the pause related DVs. The interaction between the IV and moderator was again included in the pausing model. The full regression tables are provided for each of the multiple linear regression models that were constructed. Both the *F*-statistic and adjusted R^2 value are also reported for each of the full models.

7.5.7.2. Phase II: Conditional Model Analyses

The second phase of analysis was intended to assess the conditional effect of the moderator on the direct and mediated effects within the full model, as outlined in Figure 16 (above). In more specific terms, this analysis tested for the extent to which variable levels of WMC moderated the effects of clause boundary strength and planning on each of the pause related DVs of interest. The conditional analyses were performed with the mediate function from the mediation package in R software using both the Mediated Model and Pausing Model that were fit in Phase I. To assess the effects at different levels of the moderator, low and high WMC speaker groups were specified according to the speakers with the lowest and highest 15 scores each respectively. For the Causal Mediation analyses, each of these groups was compared separately to the other 25 speakers not in the group. Each of the direct effects, mediated effects, and total model effects were estimated in those analyses through bootstrapping with 2000 Monte Carlo simulations. In the results reported below, ACME is the estimate of the average causal mediated effect (total effect – direct effect), ADE is the estimate of the average direct effect (total effect – mediated effect), and the Total Effect is the estimate of the ACME plus the ADE. Additionally, a measure of the Proportion Mediated is also reported characterizes the amount of the Total Effect that is being explained by the indirect mediated effect in the full model.

7.6. Results & Discussion

7.6.1. Predictor Variables

Table 2 provides summary statistics for each predictor variable included in the full moderated mediation models. On average the speakers produced clauses which contained approximately 6 words, but there was a good amount of variability in MLC across the individual speakers. The speaker with the lowest MLC produced clauses which only averaged about 4.5 words, compared to the speaker with the highest MLC whose clauses averaged almost 10 words. As expected, the mean MLT score was nearly twice as large as the mean MLC score, and there was a similar amount of difference between low and high MLT speakers. The mean score of discourse coherence was slightly higher than the mid-range score of 4, and likewise the overall ratings were slightly skewed toward the "very coherent" end of the scale (*max* = 5.525) as opposed to the "incoherent" end of the scale (*min* = 3.850). Finally, the mean score of WMC capacity when averaged across the Competing Language and Counting Span assessments was relatively high, although there was also a fairly large range of individual scores. Roughly one-third of the 40 participants scored below 0.800, while nearly one-fourth scored above 0.900 and the remaining scored somewhere between.

Table 2. Mean, SD and Range for each predictor variable included in the full model.

	Mean	***SD***	***Range***
Direct: **MLC**	5.862	0.738	3.527
Mediator: **MLT**	11.259	2.549	10.611
Mediator: **Coherence**	4.670	0.394	1.675
Moderator: **WMC**	0.842	0.065	0.246

Each of the predictor variables used in the models were also assessed for both normality of distributions and collinearity with other predictors prior to conducting the primary analyses. Shapiro-Wilk tests for each predictor variable indicated that all were more or less normally distributed, with *p*-values ranging from .34 to .53. To assess collinearity, each predictor in the full model was first regressed on the other predictors. A Variance Inflation Factor (VIF) for the regressed predictor was then calculated as $1/(1-R^2)$. As a rule, VIF values greater than 5 are typically interpreted as an indication of problematic multicollinearity, while values between 1-5 are taken to indicate various degrees of weaker multicollinearity (see, e.g., Sheather, 2009). Table 3 provides the VIF values for the predictors in each of the moderated mediation models constructed for the primary analyses, and there is no indication of multicollinearity in either model. Nonetheless, all predictors which were included in interaction terms in the models were first mean-centered following best-practice recommendations for linear regression analysis (see Aguinis & Gottredson, 2010; Afshartous & Preston, 2011; Robinson & Schumacker, 2009). It should also be noted that MLC and MLT were moderately correlated ($r = .46$).

Table 3. VIF measures of collinearity in the MLT and Coherence models.

	MLC	**WMC**	**MLT**	**Coherence**
MLT Model	1.21	1.04	1.30	---
Coherence Model	0.98	1.03	---	1.02

7.6.2. Mediator Models

The full results of the Mediator Model analyses are given in Tables 4 (MLT) and 5 (Coherence). Recall that these multiple linear regression models were constructed with the IV (MLC) and Moderator (WMC) as predictors of each Mediator respectively. For the speech planning model in Table 4, it can be seen that there was a significant effect of MLC on the MLT moderator. Neither WMC nor the interaction between MLC and WMC were predictive of MLT. Looking next at discourse planning model in Table 5, there were no effects of MLC, WMC, or their interaction on the mean Coherence ratings of the stories.

Table 4. Results of linear regression model predicting the MLT mediator from the MLC and WMC.

	Estimate	**Std. Error**	***T***	***p***
MLC	1.481	0.499	2.967	< .01**
WMC	9.791	5.921	1.654	.11
MLC*WMC	1.843	6.990	0.264	.79

Table 5. Results of linear regression model predicting the Coherence mediator from the MLC and WMC.

	Estimate	**Std. Error**	***T***	***p***
MLC	0.043	0.063	0.495	.62
WMC	1.685	1.027	1.641	.11
MLC*WMC	0.856	1.212	0.706	.48

Figure 18 shows the relationship between MLC and MLT among the speakers. There was a systematic trend for MLT to increase with increasing MLC. This is not surprising given that a T-unit must at contain a single clause at minimum. In addition to the visibly clear relationship between the measures, however, it is also apparent that there is a large degree of individual variability. An additional measure of clauses per T-unit (CperT) was calculated per speaker as a means of quantifying the individual differences shown in Figure 18. Overall, the speakers produced an average of nearly 2 clauses per T-unit (*M* = 1.945). On the other hand, the range of CperT measures across speakers points to meaningful differences between different groups of those speakers. A total of 8 speakers (20%) only produced 1.5 clauses or fewer per T-unit, while 21 of the speakers (~50%) produced at least 2 full clauses per T-unit. The remaining 11 speakers (~25%) produced between 1.5 and 2 clauses per T-unit. These differences suggest that some of the speakers adopted a mostly linear style of storytelling, where individual events were related successively as independent clauses. By contrast, a different group of speakers appeared to adopt a more complex hierarchical style of storytelling wherein clauses were more often embedded within other clauses. It is likely for this reason that MLC was not similarly predictive of the mean Coherence scores of the speakers. As a measure of more global discourse structure, it is reasonable to assume that the judged Coherence of a story is more closely associated with the relative degree of clausal embedding within T-units than simply with the average lengths in words of those T-units. It should be noted, however, that the correlation between judged Coherence scores and the CperT measure was relatively weak (*r* = .19).

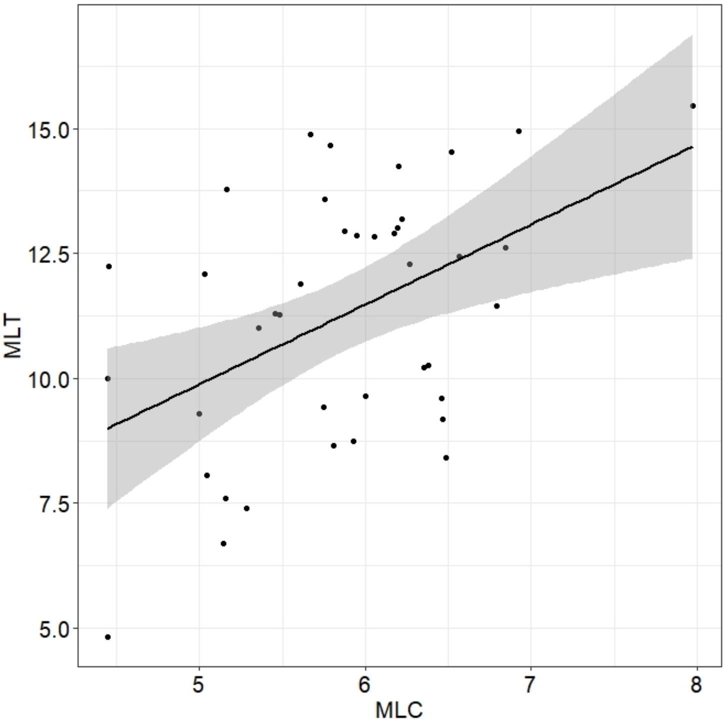


Figure 18. Relationship between MLC and MLT.

In terms of WMC, the results from the MLT Mediated Model are perhaps less surprising than those from the Coherence Mediated Model. The lack of relationship between WMC and MLT, as well as the lack of interaction between WMC and MLC, might be expected given that both MLC and MLT are measures of length related to speech planning which may be an automatic process as noted in Chapter 2. Whereas in the case of the more global discourse structure, production of a Coherent narrative requires that the speaker maintain various types of information in working memory over longer stretches of discourse time, such as the current status of referents, the overall themes or topics of the narrative, and the relationships between events embedded within those themes or topics. Thus, the lack of effect of WMC on the judged Coherence scores was unexpected.

In summary, the results presented here were only successful in confirming the hypothesized relationship between the Direct Effect of boundary strength and the Indirect Effect of speech planning domain; no relationship was found between boundary strength and discourse planning. Thus, there is the suggestion that the local language structure is influenced by the relative scope of a speaker's speech planning capabilities. However, the specific prediction was that clausal embedding would lead to shorter clauses overall (lower MLC) but also to longer "information" units (higher MLT). Independent clauses are expected to be longer than subordinate ones, but as stand-alone units they also constitute the entirety of their T-unit. To the contrary, the current results revealed a positive relationship between MLC and MLT. The additional analyses on clauses per T-unit may provide an explanation for this result. Approximately 75% of the speakers in the study produced on average at least 1.5 clauses per T-unit, and about two-thirds of those produced at least 2 more clauses per T-unit. This suggests that there was a strong tendency in general to embed subordinate clauses within main clauses across these speakers. In other words, more complex syntax was the rule rather than the exception. As such, there is likely not enough differentiation between speakers in terms of language structuring strategies for the expected relationship to emerge in the data. Regardless of the direction, however, the results nonetheless provide support for the hypothesized relationship between speech planning domain and the opportunity to pause (i.e. boundary strength).

The results failed to confirm the hypothesized relationship between boundary strength and the Indirect Effect of the participants' larger discourse planning capabilities. One possible explanation for the lack of a significant result is that while MLC is an objectively measured variable, the Coherence scores relied on the more subjective judgments of naïve raters. A more satisfying explanation would be that the relative Coherence levels of narratives are judged holistically, while the measure of MLC is intended to characterize but a single local aspect of those judgments. As such, MLC is likely only capturing a small amount of the information that is reflected in the broader Coherence judgment scores.

Finally, the was no evidence in the present analyses for the hypothesized moderation of the indirect planning effects by WMC. However, it worth noting that the relationship between WMC and MLT at least approached significance ($p = .11$). Indeed, when splitting speakers evenly into binary groups between the lowest and highest WMC scores, there is a suggestion of the expected relationship. The lower WMC speakers had a mean MLT value of 10.9 words compared the high WMC speakers whose mean MLT was 11.6. Despite differing in the expected direction, however, the overall difference between groups is still quite small. As noted, this small difference may be because speech planning is largely automatic, but a larger-scale study with more participants may help to shed light on the relationship.

7.6.2.1. Models Predicting Pause Frequency

MLT Mediated Model. Table 6 provides the results of the full multiple linear regression model predicting pause frequency with MLT as a mediating factor. There were no significant main effects of the Direct MLC predictor, Mediating MLT predictor, or Moderating WMC predictor on the frequency of pausing across speakers. There was also not a significant interaction between MLC and WMC in this regard. Of each of the predictor variables in the full model, MLC was alone in approaching significance. The relationship between MLC and Pause Frequency is shown in Figure 19. The regression line shows a general trend for a roughly 1-to-1 correspondence between clause length and pause-delimited utterance length. But as indicated by the overall fit of the model predicting Pause Frequency from MLC, there is a large amount of variability in the relationship between individual speakers [$F(1,38) = 4.80$, $p = .03$ $R^2 = .09$]. It is clear from the figure that while some of the speakers followed the regression trend, some of them also paused more frequently than once per clause and yet others produced nearly 2 clauses on average between pauses.

Table 6. Results of the full linear regression model predicting pause frequency with a speech planning mediator.

	Estimate	**Std. Error**	***t***	***p***
MLC	0.797	0.462	1.727	.09
WMC	-1.445	5.096	-0.284	.78
MLT	0.024	0.138	0.175	.86
MLC*WMC	-8.151	5.805	-1.404	.17

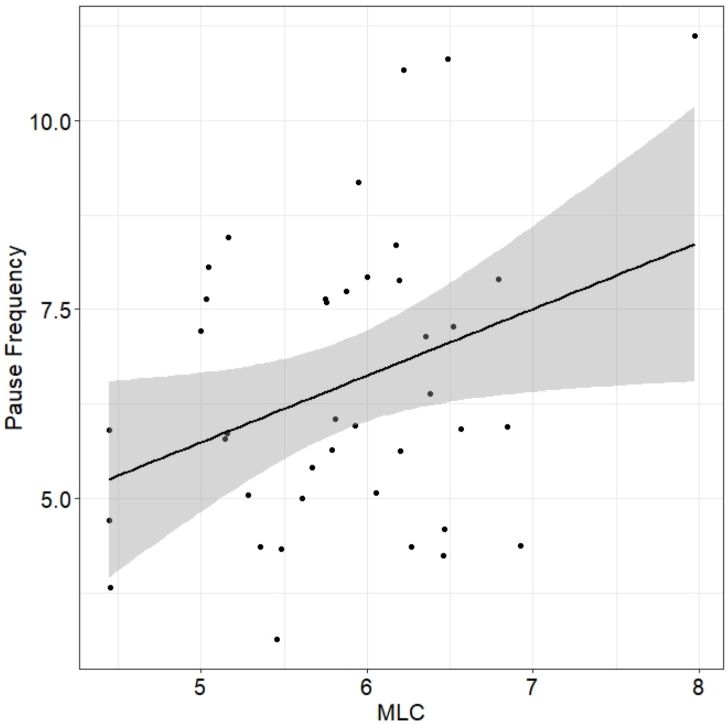


Figure 19. The relationship between MLC and Pause Frequency (= MLU).

Table 7 gives the full results of the Causal Mediation analyses, which investigated the conditional effects on pause frequency with the MLT moderator for low WMC and high WMC speakers, respectively. When comparing the low WMC speakers to the rest of the speakers there were no significant differences in the effects of MLC, MLT, or in the interaction between them. Likewise, a very similar picture emerges when comparing the high WMC speakers to the rest of the speakers. The lack of any significant effects in the full linear regression model predicting pause frequency holds across the speakers irrespective of individual WMC differences. Taken all together, the results from these analyses strongly suggest that at least within the framework of this preliminary model relative pause frequency is not meaningfully influenced by an individual speaker's WMC.

Table 7. Results of the Causal Mediation analyses of pause frequency with MLT as a mediator for low (top) and high (bottom) WMC speakers.

Low WMC				
	Estimate	**95%CI Lower**	**95%CI Upper**	***p***
ACME	0.018	-0.117	1.140	.81
ADE	-5.886	-15.850	4.280	.23
Total Effect	-5.8686	-15.629	4.430	.23
Prop. Mediated	-0.003	-0.150	0.240	.85
High WMC				
ACME	0.019	-0.137	0.960	.82
ADE	-6.306	-16.934	4.760	.21
Total Effect	-6.287	-16.601	4.850	.22
Prop. Mediated	-0.003	-0.013	5.590	.88

Among all the results from these analyses there was very little evidence to support the Direct Effect of boundary strength, the Indirect Effect of speech planning domain, or the Moderating effect of WMC on pause frequency. The only suggestion of a relationship was between MLC and pause frequency, which approached significance. As shown in Figure 19 (above), however, the direction of relationship was opposite the hypothesized one. Shorter clauses were predicted to result in less frequent pausing as function of more frequent clausal embedding. However, it is difficult to make inferences on the relationship given the very large amount of variability seen in Figure 19. The lack of effects from MLT or WMC are also counter to the model hypotheses but are perhaps unsurprising given the Mediator Model analyses which found neither measure to related to the MLC measure of boundary strength.

<u>*Coherence Mediated Model*</u>. Table 8 provides the results of the full multiple linear regression predicting pause frequency with judged narrative Coherence as a mediating factor. Contrary to the results from the MLT mediated model, the full Coherence

mediated model explains considerably more of the variance in pause frequencies overall [$F(4,35) = 4.184$, $p < .01$, $R^2 = .25$]. Furthermore, there was a significant effect of narrative Coherence on pausing frequency as shown in Figure 20. There is a clear trend for a decrease in pause frequency (i.e. increasing MLU) with an increase in judged Coherence levels. The most likely explanation for this relationship is that longer utterances correspond with more complex language production. A speaker who tends to elaborate on the topics of the narrative, and correspondingly tends to embed subordinate clauses within main clauses, is expected to produce longer utterances on average than a speaker who simply states the relevant events of the narrative in a linear fashion. An alternative possibility is that the relationship between pause frequency and judged narrative Coherence reflects differences in speech fluency. Pausing is expected to occur more frequently in disfluent speech, and by definition a disfluent narrative is less Coherent than a fluent narrative.

Table 8. Results of the full linear regression model predicting pause frequency with a discourse planning mediator.

	Estimate	**Std. Error**	***t***	***p***
MLC	0.744	0.373	1.996	.05
WMC	-4.719	4.572	-1.032	.31
Coherence	2.084	0.716	2.911	< .01**
MLC*WMC	-9.890	5.242	-1.887	.07

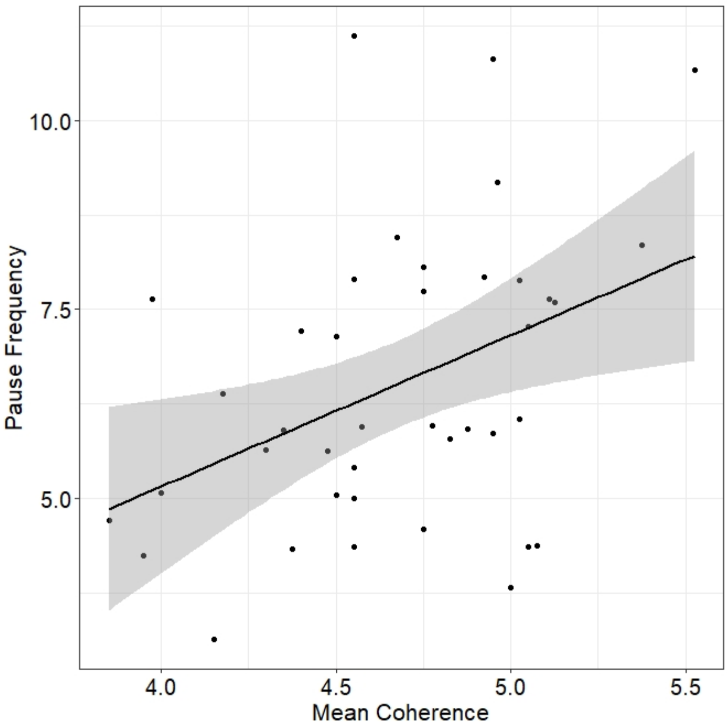


Figure 20. Pause frequency (= MLU) by mean narrative Coherence judgment scores.

Similar to the results from the MLT mediated model (see Figure 19), the effect of MLC on pause frequency again neared significance in the Coherence mediated model. In addition, there was a near significant effect of the interaction between MLC and WMC on pause frequency. This is shown in Figure 21 where speakers were split into binary low and high WMC groups. For the low WMC speakers, there appears to be a very similar 1-to-1 correspondence between MLC and Pause Frequency to that seen in Figure 19 (above) for all speakers. By contrast, the trend among the high WMC speakers is for relatively uniform frequency of pausing with more variable of clause lengths between those pauses. These patterns suggest that the low WMC speakers have a tendency to plan their speech production units more at the level of the clause, as opposed to the high WMC speakers who have a tendency to plan more at the level of multi-clausal utterances. In other words, it appears that low WMC speakers are more apt to plan communication of individual narrative events in a listwise fashion while high WMC speakers are more inclined to plan communication of more complete and elaborate ideas.

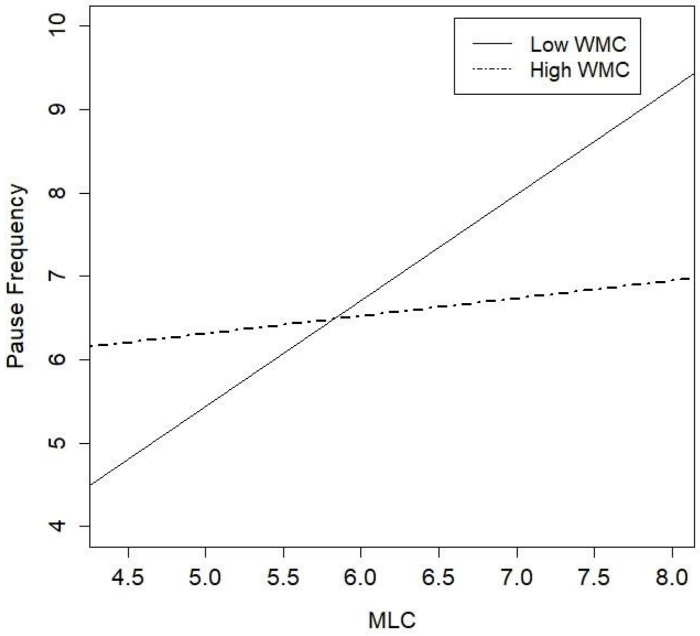


Figure 21. Relationship between MLC and Pause Frequency (= MLU) by low (solid) and high (dashed) WMC groups.

Table 9 gives the full results of the Causal Mediation analyses, which investigated the conditional effects on pause frequency with the Coherence moderator for low WMC and high WMC speakers, respectively. Once again, there were no significant differences in the effects of MLC, Coherence, or both together on pause frequency between the low WMC group and the rest of the speakers. Likewise, the analyses revealed no significant differences between the high WMC group and the rest of the speakers in these regards. This suggests that the significant mediating effect of narrative Coherence scores that was found in the full regression model is not being driven more strongly by a particular subgroup of WMC speakers. In other words, there appears to be a significant relationship between judged Coherence and pausing frequency that is independent of the individual cognitive processing differences.

Table 9. Results of the Causal Mediation analyses of pause frequency with Coherence as a mediator for low (top) and high (bottom) WMC speakers.

Low WMC				
	Estimate	**95%CI Lower**	**95%CI Upper**	***p***
ACME	1.552	-1.717	12.590	.47
ADE	-7.366	-21.959	1.600	.14
Total Effect	-5.814	-15.102	4.720	.24
Prop. Mediated	-0.267	-78.650	-0.140	.63
High WMC				
ACME	1.644	-1.633	12.530	.46
ADE	-7.875	-22.490	2.480	.15
Total Effect	-6.231	-15.416	5.770	.23
Prop. Mediated	-0.264	-1006.725	-0.45-	.60

The results from the full Coherence Mediated model provide more support for the model hypotheses overall. Most importantly, the judged Coherence measures significantly predicted the pause frequency, and in the hypothesized direction. More Coherent stories were characterized by less frequent pausing, likely reflecting more complex discourse structures and the corresponding decrease in opportunities for grammatical pausing. The hypothesized Direct Effect of MLC on pause frequency also approached significance, but the interpretation of this relationship is problematic for the reasons given above with reference to the MLT Mediated model. Finally, it was also found that the interaction between MLC and WMC neared significance in the full model. The suggestion from Figure 21 (above) is that low and high WMC speakers employ different strategies in the scope of language planning, with the latter planning more complex linguistic utterances than the former. Despite these results from the full model, the Causal Mediation analyses suggested that the moderating effect of WMC on pause frequency did not contribute significantly to either the Indirect or Total Effects in the model.

7.6.2.2. Models Predicting Mean Pause Duration

MLT Mediated Model. Table 10 provides the results of the full multiple linear regression predicting mean pause duration with MLT as a mediating factor. Neither the Direct MLC effect, Mediating MLT effect, or Moderating WMC effect were significantly predictive of the mean pause durations across speakers. There was also not a significant interaction between MLC and WMC in this regard. Accordingly, the overall fit of the full model was poor and it explained virtually none of the variance [$F(4,35) = 0.426$, $p = .79$, $R^2 < .01$]. Based on the experimental findings of the current study, the lack of relationships between the mean pause durations and boundary strength or speech planning domain was unexpected. However, there are some plausible explanations for these findings, which will be returned to in the Discussion below.

Table 10. Results of the full linear regression model predicting mean pause duration with a speech planning mediator.

	Estimate	**Std. Error**	***t***	***p***
MLC	0.043	0.034	1.266	.21
WMC	0.066	0.380	0.175	.86
MLT	-0.007	0.010	-0.694	.49
MLC*WMC	-0.042	0.433	-0.098	.92

Table 11 gives the full results of the Causal Mediation analyses, which investigated the conditional effects on mean pause duration with the MLT moderator for low WMC and high WMC speakers respectively. There were again no significant differences in the effects of MLC, MLT, or the combination of both on mean pause durations between the low WMC group and the rest of the speakers. The same lack of difference in effects holds also when comparing the high WMC to speakers to the rest of the speakers. In light of the lack of significant effects found in the full multiple linear regression model, the lack of meaningful differences between WMC groups with regards to each of these effects was foreseeable.

Table 11. Results of the Causal Mediation analyses of mean pause duration with MLT as a mediator for low (top) and high (bottom) WMC speakers.

Low WMC				
	Estimate	**95%CI Lower**	**95%CI Upper**	***p***
ACME	-0.021	-0.573	0.030	.74
ADE	0.009	-0.492	1.560	.91
Total Effect	0.012	-0.492	1.460	.97
Prop. Mediated	1.710	-0.012	0.010	.96
High WMC				
ACME	-0.022	-0.630	0.040	.76
ADE	0.007	-0.451	1.970	.87
Total Effect	-0.015	-0.433	2.040	.94
Prop. Mediated	1.440	-7.040	1.120	.95

Similar to the MLT Mediated Model for pause frequency, none of the hypothesized effects were supported in the similar model for pause duration. In fact, in this model there were not even any suggestions of effects in that none were near significant. The most likely explanation for these results lies in the nature of the analyses. Pause durations vary at a very local level but were investigated in the current model at a global speaker level. This is a point that will that be revisited below.

Coherence Mediated Model. Table 12 provides the results of the full multiple linear regression model predicting mean pause duration with Coherence as a mediating factor. Similar to the MLT mediated model, there were once again no significant relationships of the Direct MLC effect, Mediating Coherence effect, or Moderating WMC effect with the mean pause durations across speakers. The interaction between MLC and WMC was also not significant. As expected, the overall fit of the Coherence mediated model was poor and it also failed to explain any of the variance in the mean pause durations [$F(4,35) = 0.3138$, $p = .87$, $R^2 < .01$]. Once again, the apparent lack of relationship between mean pause durations and discourse structure was unexpected given the experimental findings reported in the current study.

Table 12. Results of the full linear regression model predicting mean pause duration with a discourse planning mediator.

	Estimate	**Std. Error**	***t***	***p***
MLC	0.033	0.031	1.077	.29
WMC	0.018	0.382	0.048	.96
Coherence	-0.013	0.060	-0.218	.83
MLC*WMC	-0.044	0.438	-0.101	.92

Table 13. Results of the Causal Mediation analyses of mean pause duration with Coherence as a mediator for low (top) and high (bottom) WMC speakers.

Low WMC				
	Estimate	**95%CI Lower**	**95%CI Upper**	***p***
ACME	-0.010	-0.371	0.020	.69
ADE	-0.003	-0.433	1.700	.92
Total Effect	-0.012	-0.450	1.550	.98
Prop. Mediated	0.778	-30.956	0.200	.85
High WMC				
ACME	-0.010	-0.367	0.030	.70
ADE	-0.005	-0.521	1.710	.95
Total Effect	-0.015	-0.506	1.630	1
Prop. Mediated	0.671	4.395	5840.240	.85

Table 13 gives the full results of the Causal Mediation analyses, which investigated the conditional effects on mean pause duration with the Coherence moderator for low WMC and high WMC speakers. Based on the results of the multiple linear regression given in Table 12 (above), it is not surprising that there were again no significant differences for either the low or high WMC groups on the effects of each predictor of mean pause duration.

Given the results of the MLT Mediated model on pause durations, the lack of Direct or Moderating effects of MLC or WMC on pause durations is not surprising. In addition to these, there was also no evidence for the hypothesized relationship between the Mediating Effect of narrative Coherence and pause durations. Moreover, the effect of Coherence on pause durations was found to be extremely weak ($p = .83$). In line with the full model results, the Causal Mediation analyses also revealed no suggestion of any of the effects on pause duration in terms of difference WMC groups.

7.6.2.3. Models Predicting Variability of Pause Durations

MLT Mediated Model. Table 14 provides the results of the full multiple linear regression model predicting pause duration variability (= CV_{dur}) with MLT as a mediating factor. The were no significant main effects of MLC, WMC, or MLT on the CV_{dur} measure, but there was a significant effect of the interaction between MLC and WMC. Figure 22 shows the significant interaction with speakers split evenly into binary low and high WMC groups. For the high WMC speakers, there was a trend for pause duration variability to increase with increasing MLC, while for the low WMC there does not appear to be a strong relationship between the 2 measures. This suggests that the prediction for less variability with increased language complexity holds true for the high WMC speakers but not the low WMC speakers. Overall, the MLT mediated model was not significantly predictive of pause duration variability and explained a relatively small amount of the variance in the CV_{dur} scores [$F(4,35) = 2.188$, $p = .10$, $R^2 = .10$].

Table 14. Results of the full linear regression model predicting pause duration variability with a speech planning mediator.

	Estimate	**Std. Error**	***t***	***p***
MLC	0.007	0.035	0.212	.83
WMC	0.526	0.382	1.379	.18
MLT	0.009	0.010	0903	.37
MLC*WMC	1.059	0.435	2.436	.02*

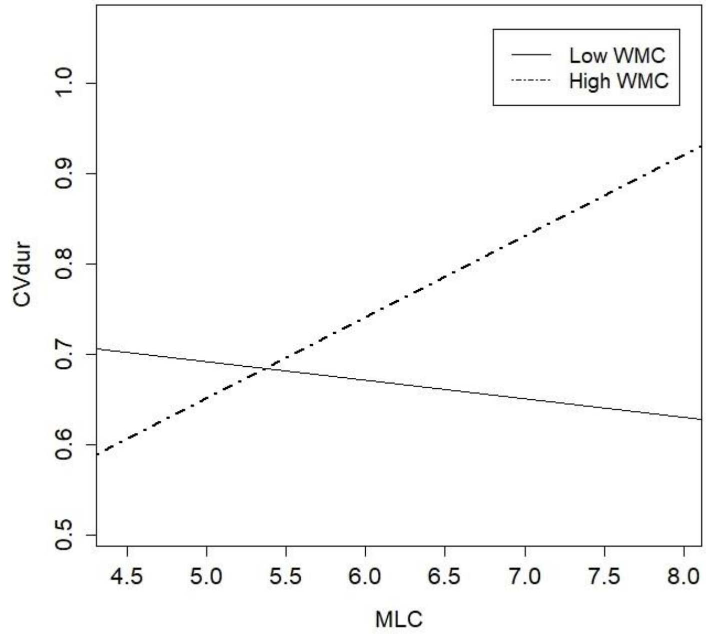


Figure 22. Relationship between MLC and CV_{dur} by low (solid) and high (dashed) WMC groups.

Table 15 gives the full results of the Causal Mediation analyses, which investigated the conditional effects on pause duration variability with the MLT moderator for low WMC and high WMC speakers respectively. When comparing both the low and high WMC capacity speakers to the remaining speakers, there were significant effects of both the direct effect of MLC and of the overall model, but not of the mediating MLT effect. The significant direct effect is not surprising given the significant relationship between MLC and WMC that was shown in Figure 22 (above). In light of the lack of a significant mediating effect in either analysis, the significant total effect in both analyses suggests that the significant direct effect of MLC alone is explaining enough the variance in CV_{dur} to make the full model significantly predictive.

Table 15. Results of the Causal Mediation analyses of CV_{dur} with MLT as a mediator for low (top) and high (bottom) WMC speakers.

Low WMC				
	Estimate	**95%CI Lower**	**95%CI Upper**	***p***
ACME	0028	-0.067	0.550	.79
ADE	0.876	-0.018	1.410	.04
Total Effect	0.904	0.048	1.430	.03
Prop. Mediated	0.031	-0.070	1.940	.79
High WMC				
ACME	0.029	-0.091	0.540	.77
ADE	0.931	-0.068	1.420	.04
Total Effect	0.959	-0.023	1.470	.03
Prop. Mediated	0.0301	-0.268	0.570	.78

In terms of pause duration variability, the results from the full MLT Mediated model revealed that only the hypothesized relationship between MLC and WMC was evident. The pattern shown in Figure 22 suggests that the prediction of decreasing CV_{dur} with increasing language complexity was upheld for the high WMC speakers but not the low WMC speakers. As the high WMC speakers produced longer clauses, assumed to index structurally simpler language, their CV_{dur} values increased. Besides this finding, none of the other hypothesized relationships between the Direct Effect of boundary strength, the Mediating Effect of speech planning domain, or the Moderating Effect of WMC were supported. In terms of the Causal Mediation analyses, the relationship between MLC and WMC on pause duration variability was further supported. When comparing either the low or high WMC groups to the rest of the speakers, there were significant differences in the effect of boundary strength on pause duration variability. More interestingly, these analyses also revealed differences between the low and high WMC speakers in terms of the total effect of the model on variability. This likely reflects that the relationship between MLC and duration variability is so much stronger for the high WMC speakers than the low WMC speakers that this difference alone is sufficient in accounting for variability in the CV_{dur} values across speakers.

Coherence Mediated Model. Table 16 provides the results of the full multiple linear regression model predicting pause duration variability (= CV_{dur}) with narrative Coherence as a mediating factor. The results of this model are very similar to the MLT mediated model, with no significant main effects of MLC, WMC, or judged narrative Coherence. Once again there was a significant effect of the interaction between MLC and WMC. Comparing the 2 mediated models, it also appears that neither speech planning domain (= MLT) or discourse planning (= Coherence) are mediating factors in determining the variability of pause durations within-speaker. This interaction can be seen in Figure 22 (above).

Table 16. Results of the full linear regression model predicting pause duration variability with a discourse planning mediator.

	Estimate	**Std. Error**	***t***	***p***
MLC	0.020	0.031	0.642	.52
WMC	0.477	0.385	1.499	.14
Coherence	0.024	0.060	0.400	.69
MLC*WMC	1.056	0.441	2.392	.02

Table 17 gives the full results of the Causal Mediation analyses using the Coherence moderator of pause duration variability for speakers with low and high WMC. The same overall patterns of results are visible as were found with MLT and the mediating factor. As with the full multiple linear regression for all speakers (Table 16), the mediating effect of narrative Coherence was not significantly more predictive of pause duration variability for either the low or high WMC groups of speakers. As before, there were significant differences between the groups in both the direct effect of MLC (see Figure 22) and the total effect of the model.

Table 17. Results of the Causal Mediation analyses of CV_{dur} with Coherence as a mediator for low (top) and high (bottom) WMC speakers.

Low WMC				
	Estimate	**95%CI Lower**	**95%CI Upper**	***p***
ACME	0.018	-0.092	0.440	.88
ADE	0.885	-0.027	1.450	.04
Total Effect	0.904	0.027	1.430	.03
Prop. Mediated	0.020	-2.699	0.070	.88
High WMC				
ACME	0.019	-0.101	0.410	.90
ADE	0.940	-0.061	1.490	.04
Total Effect	0.959	-0.013	1.480	.04
Prop. Mediated	0.020	-0.197	0.880	.90

The results from the Coherence Mediated model did not differ significantly from the MLT mediated one. The effects of boundary strength and WMC were expected given the results of the MLT model. Likewise, the significant interaction between MLC and WMC on pause duration variability was expected for the same reason. Most importantly, the hypothesized relationship between narrative Coherence and duration variability was not supported. The prediction was that more complex stories would provide fewer opportunities for grammatical pausing leading to decreased variability of pause durations, but no evidence of this relationship was found. As before, there is the possibility that the Coherence judgments obtained from naïve raters tapped into some aspect(s) of the narratives beyond simply their structural complexity. In other words, it is not entirely possible from the current results to dismiss the hypothesis that more coherent stories will be characterized by less variability in pause durations.

7.7. General Discussion

In summary, the preliminary results from the model analyses presented here suggest very little mediation from either speech or discourse planning on the various aspects of pausing patterns in narrative speech. The lone planning effect was a significant positive relationship between the narrative Coherence scores and the frequency of pausing, which was in the predicted direction. There was also the suggestion of an effect of clause boundaries on pause frequency, but not on the other features of pause patterns. In terms of WMC, there was some evidence that individual cognitive differences moderate effects on pausing behavior. These effects are worth exploring in future work with more participants. A final observation from these analyses which is important to note is that the mean pause duration measure was alone in not displaying relationships with any of the hypothesized effects in the model.

7.8. Model Limitations

An obvious limitation of the modeling approach employed in the current study is the heavy reliance on summary statistics to characterize the various effects within the model. The decision to use these measures was based largely on the goal of quantifying differences in pausing behavior at the speaker-level with the aim of characterizing the role of individual differences on the pausing patterns. However, this decision necessarily involves a trade-off between uniformity of the analyses and maintenance of meaningful information. Given that pause frequency and pause duration variability are more naturally global behavioral measures whereas pause durations themselves operate at the level of individual utterances, compromise was necessary in order to model all 3 aspects of pausing patterns within a single holistic framework. The clear downside to using summary measures is that a lot of the interesting and meaningful variability within stories or speakers will necessarily be obscured. For instance, it is entirely possible that 2 or more of the speakers produced very similar mean pause durations but very different distributions of pause distributions. It is more likely than not that such a difference in pause duration distributions could tell us something meaningful about more general differences between those speakers as opposed to simply representing random chance. The situation is similar for the predictor variables, in that the ways in which clause or T-unit lengths distribute within a narrative may be more informative than the mean lengths of the units.

Additionally, as with any attempt at formulating a theoretical model of behavioral processes, several assumptions must be made. In the case of the current model, the assumptions take 2 forms. The first major assumption is that the theorized components of relevance within the model are in fact the most important ones, not to mention whether they interact in the manner proposed. This is perhaps the less concerning assumption in the current model given that the included components were motivated both by findings from a vast array of previous studies as well as from the current one. The potentially more problematic assumption is that the measures included in the analyses accurately reflect the processes that they were assumed to index. For example, in the current study it was assumed that the naïve judgments of narrative Coherence provided an accurate characterization of the relative complexities of discourse structures across the stories. It is quite possible, however, that the use of a more rigorously defined standardized measure of discourse structure might yield more significant results than were found here. The use of this type of alternative measure would not be without limitations itself. Such measures typically require a large amount of hand-coding by the researcher, which often results in more limited sample sizes due to the time-consuming nature of the work. Of course, theoretical assumptions are not a problem in and of themselves so long as they are well-motivated, and the degree to which those assumptions are valid can only be properly assessed through empirical testing of the proposed model using naturalistic data. In that sense, the current study represents at least an important first step in attempting to arrive at an ecologically valid and holistic model of pausing behavior in narrative speech. If nothing else, the general lack of significant results in the above analyses at least provides a jumping-off point from which the theoretical assumptions can be re-assessed, and the model can be refined for future testing.

A final limitation in the testing of the proposed model is related to the issue of sample size. This is particularly problematic regarding assessment of the moderating effect of WMC on the direct effect of clause boundaries and the mediating effects of speech and discourse planning, which requires that participants to be split into subgroups. The most commonly employed method for designating low and high subgroups for the moderating factor is to set the cutoffs relative to the mean +/-1 SD. In doing this, the low subgroup speakers would be those whose WMC score was at least 1SD below the mean score while the high subgroup of speakers would be those whose WMC was at least 1SD above the mean score. In the current dataset, this method results in very low sample sizes for each of the WMC subgroups (low: $N = 5$; high: $N = 7$). By contrast, the most common suggestion for causal mediation analyses is that each subgroup have a sample size of at least 30 to ensure reliable estimation of the coefficients (see e.g., Sekaran, 2003). It is likely for this reason that a meta-analysis of the use of moderating effects in previously published studies found that on average the sample sizes ranged from 100 at lowest end to more than 1000 at the highest end (Aguinis, Beaty, Boik, & Pierce, 2005). Thus, the sample of only 40 speakers in the current study was well below the norm. An attempt was at least made at alleviating this issue in the current study with the methodological decision to split the WMC subgroups according to the 15 lowest and highest scores, respectively. It has also been argued that the use of bootstrapping methods in moderation analyses is typically an effective strategy for obtaining more reliable estimation of coefficients from smaller sample sizes (e.g. Preacher, Rucker, & Hayes, 2007), and so it is not entirely clear to what extent the current analyses were hindered by the small sample size of the data. Nonetheless, caution should still be exercised in drawing any strong inferences about the role of individual WMC differences on pausing behavior from the results of the current study alone.

VIII. CONCLUSIONS AND FUTURE WORK

The experimental goals of the current study were largely successful. The primary aim of Experiment 1 was to quantify the relative weighting of perseveratory respiratory recovery effects and anticipatory speech planning effects on pausing. The speakers were found to pause more frequently and for longer when both the preceding and following sentences were long as opposed to short. This finding confirms that both low-level recovery and planning effects act to constrain the pausing patterns that emerge during speech production. Further investigation of the interaction between the 2 effects also provided insight into relative weighting of each in this regard. There was a systematic pattern of more frequent and longer pauses following long sentences than following short sentences that was independent of the following sentence length. This finding provides evidence that the respiratory recovery effects are stronger constraints on pausing behavior than are the speech planning ones.

The primary aim of Experiment 2 was to confirm and quantify the effect of linguistic cohesion between successive clauses on pausing. The speakers were found to pause with increasing frequency as the degrees of referential coherence and embeddedness between the clauses increased. This finding confirms that at least in terms of frequency, the local syntactic structure also contributes systematically to variations in the pausing patterns that emerge during speech production. There was less conclusive evidence of similar effects on the measures related to pause durations, but there was likewise not conclusive evidence that such effects should be ruled out.

The primary aim of Experiment 3 was to confirm and quantify the effect of global discourse planning on pausing. This effect was assessed in terms of the relative degree of linguistic coherence in narrative speech productions. The speakers were found to pause more frequently and for longer when telling a less coherent narrative. There was also a suggestion that the variability of pause durations increases with decreasing coherence. These findings served to confirm that discourse planning processes also constrain the various features of pausing patterns which emerge in more naturalistic speech productions.

In contrast to the experimental findings, it is clear that more work remains to be done toward constructing a satisfactorily explanatory model of pausing behavior in narrative speech. The results from both the experimental analyses and the preliminary model results provide plenty of suggestions on the possible directions that this future work should take. The remainder of the conclusions are devoted to a discussion of these future research directions.

8.1. Indexing of the Model Components

As previously discussed with regard to the limitations of the current study, the results from the proposed model leave open the question of whether other measures might be better suited to indexing the various factors hypothesized to systematically influence narrative pausing behavior. For example, while it is reasonable to expect the measure of MLC to capture some level of information about the relative boundary strengths within speech samples, it is obviously not capable of providing more specific information about the exact nature of relationships that hold between the utterances in the sample. As a summary measure, it is sure to obscure some sorts of meaningful information about the linguistic structure of the discourse. As demonstrated by the results from Experiment 2, it is not enough to look simply in terms of embedded vs unembedded in characterizing the effect of syntactic cohesion. The relative degree of cohesion is also an important factor, and MLC is not particularly well-suited at providing such a characterization. A measure based on more careful and theory-bound coding of syntactic cohesion from clause-to-clause within the narratives is likely to yield more useful information on the role that this factor plays. For much the same reasons, the measure of MLT as an index of the scope of speech planning is potentially problematic in that it can only provide a very general characterization of the planning process.

By contrast, the narrative Coherence measure was at least based on human judgments, and so is more likely to be reflective of the factor that is was intended to index. Nonetheless, the extent to which judgments from naïve raters actually reflect the features that they were intended to measure can only be assumed. For this reason, it would be worthwhile to also take a more hands-on approach to the coding of discourse structure within the narratives to be analyzed. Numerous frameworks already exist for conducting discourse structure analyses, such as Story Grammar analysis (see Stein & Glenn, 1979) or Rhetorical Structure Theory (see Mann, Matthiessen, & Thompson, 1989), and adoption of one of these preexisting frameworks would allow for a characterization of discourse structure that is more grounded in theory. The drawback to such an approach is that it is much more time-consuming than the crowd-sourced collection of ratings, but the results would likely be more informative to a model of pausing behavior.

The WMC measure that was used in the current model was based on assessment procedures that are well-established in the literature, and so their use in characterizing WMC in the current study was supported by previous research. That notwithstanding, these 2 assessment tests alone do not provide a full characterization of the various aspects of cognitive functioning which may be relevant to narrative speech production. In future work it would be useful to incorporate a complete assessment of speakers' cognitive functioning abilities through use of more comprehensive assessment tools such as provided by the NIH Toolbox Cognition Battery (see Zelazo et al., 2013). The collection of measures related to attention and executive functioning, episodic and working memory, processing speed, etc., would likely yield more meaningful and differentiating results on the role of individual differences in pausing behavior.

8.2. Additional Model Components

In addition to arriving at more appropriate indexing measures of hypothesized model components in the current study, it is certain that a final model of pausing behavior must include additional components which were not investigated in the preliminary model analyses. The most obvious of these is suggested by the results of Experiment 1. The speech planning component of the model was motivated by the finding that the length of an upcoming sentence has a systematic effect on the frequency and duration of pausing. However, it was also found that there was an effect of the length of the preceding sentence on those pausing features, and there was a suggestion that this effect was even stronger that the planning effect. The analyses from Experiment 1 provided strong evidence that this effect was driven largely by respiratory recovery, and any future attempt at improving a model of pausing behavior must account for the role of respiratory functioning. This was not possible in the current study due to the lack of a reliable measure. Future work should be aimed at the collection of respiratory kinematic data from speakers performing similar narrative storytelling tasks in order to further account for role of speech breathing in pausing behavior.

Additionally, it is well-demonstrated in the speech production literature that articulation rate plays a major role in shaping pausing patterns. In particular, numerous studies have found that articulation rate is negatively correlated with pause duration such that faster speech results in shorter pauses (see Brubaker, 1972; Fougeron & Jun, 1998; Kendall, 2013). Also, if measuring pause frequency in terms of number of words as in the current study, then it would also be expected that faster articulation rates would be associated with less frequent pausing. This factor was not investigated in the current study given the experimental focus on the structural aspects of narrative speech production but it should be included in future models.

8.3. Alternative Methods of Analyses

If nothing else, it should be clear by this point that naturalistic pausing behavior is highly complex along both systematic and idiosyncratic dimensions. Attempting to fully explain the underlying processes of the system within a unified model is accordingly no small feat. It is no doubt for this reason that such attempts are conspicuously absent in the existing literature on pausing. As an integral feature of speech production, however, a more thorough understanding of pausing behavior more specifically will in turn lead to a more complete understanding of the speech production system more broadly. The preliminary model proposed in the current study represents but one step toward this end. It is clear from the preliminary results that future work in this domain will need to explore the use of analysis methods better suited at extracting the meaningful patterns from the noise in the speech pausing signal. Given the limitations of the preliminary model as laid out in the preceding chapter, the best approach may be to combine multiple approaches aimed alternatively at investigating pausing patterns in terms of their more global characteristics, as in the current study, as well as the patterns of variation that emerge from a more local sequential analysis. The challenge will be in synthesizing the approaches into a single model to create a more unified account of speech pausing behavior than currently exists.

APPENDIX A: EXPERIMENT 1 STIMULI

The following sentences were used for the Short condition:

1) *The science teacher gave her students directions.*

2) *Her husband made all of us some materials.*

3) *Their daughter sent her professor many questions.*

4) *A local company bought the town a building.*

5) *The woman showed her family the new picture.*

6) *The current director brought them information.*

7) *Your little brother built his friend a computer.*

8) *The officer read somebody a good story.*

The following sentences were used for the Long condition:

1) *The American expert gave the foreign official a significant opportunity.*

2) *The recent administration bought each of the former authorities the available homes.*

3) *The traditional president built her financial business a general management team.*

4) *The serious individual showed the popular manager several new solutions.*

5) *The respected institution read their likely employees the economic analysis.*

6) *The important medical doctor brought the public hospital some much needed experience.*

7) *The international player sent the various couples a lot of personal resources.*

8) *A private community member made his college a more modern education policy.*

APPENDIX B: EXPERIMENT 2 STIMULI

The following clause pairs were used for the Relative clause condition:

1) *I have yet to read any of the science books*
   that my professor gave me for personal use.
2) *We did not remember the special examples*
   that the likely candidate presented to us.
3) *The company determined the activities*
   that the other people were allowed to enjoy.

The following clause pairs were used for the Coordinated clause condition:

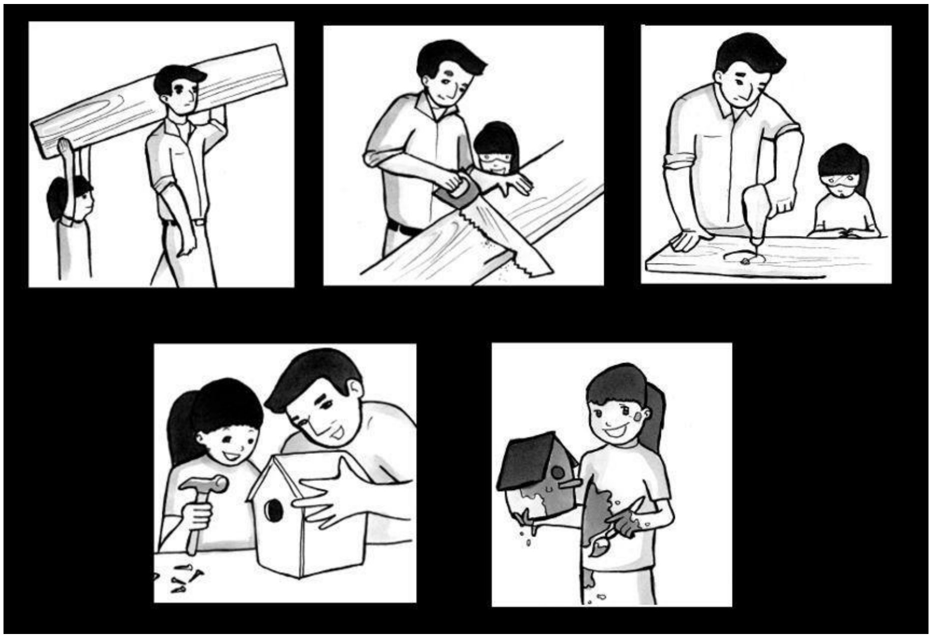
1) *My friend's father is a language authority*
   and he often teaches us about the subject.
2) *The executive understood the benefits*
   and he identified the most simple response.
3) *The television star recognized the artist*
   and they established a social relationship.

The following clause pairs were used for the Separate clause condition:

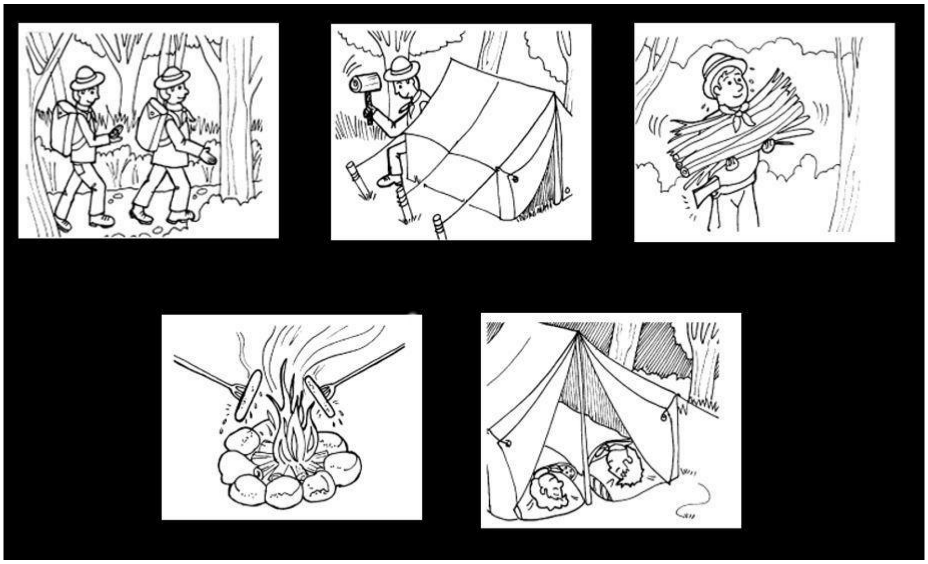
1) *My brother and I visited the property*
   and then the manager showed us many units.
2) *The politician considered the evidence*
   then the other leaders developed a plan.
3) *The government created the technology*
   and the industry people imagined its use.

APPENDIX C: EXPERIMENT 3 STIMULI

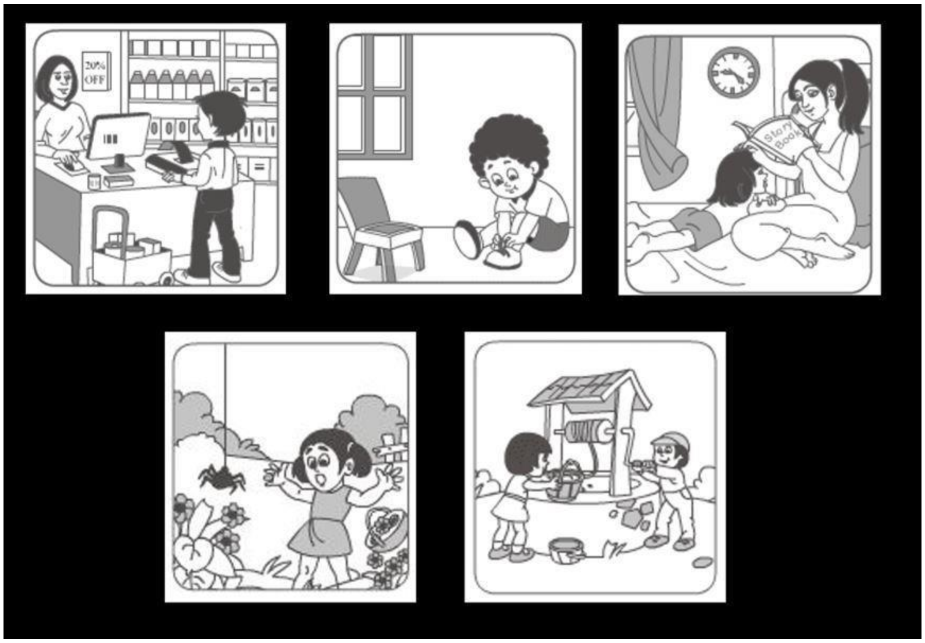
The following picture sets were for the High Coherence condition:



The following picture sets were for the Moderate Coherence condition:



The following picture sets were for the Incoherent Coherence condition:



REFERENCES CITED

Acheson, D.J., & McDonald, M.C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*(1), 50-68. doi: 10.1037/a0014411

Afshartous, D., & Preston, R.A. (2009). Key results of interaction models with centering. *Journal of Statistics Education*, *19*(3), 1-24. doi: 10.1080/10691898.2011.11889620

Aguinis, H., Beaty, J.C., Boik, R.J., & Pierce, C.A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology*, *90*(1), 94-107. doi: 10.1037/0021-9010.90.1.94

Aguinis, H., & Gottfredson, R.K. (2010). Best-practice recommendations for estimating interaction effects using moderated multiple regression. *Journal of Organizational Behavior*, *31*, 776-786. doi: 10.1002/job.686

Ahmadian, M.J. (2013). Working memory and task repetition in second language oral production. *Asian Journal of English Language Teaching*, *23*, 37-55.

Alario, F.-X., Costa, A., Ferreira, V.S., & Pickering, M.J. (2006). Architectures, representations and processes of language production. *Language and Cognitive Processes*, *21*(7-8), 777-789. doi: 10.1080/016909600824112

Bailly, G., & Gouvernayre, C. (2012). Pauses and respiratory markers of the structure of book reading. In *Proceedings of the 13th Annual Conference of the International Speech Communication Association (INTERSPEECH2012)* (pp. 2218-2221). Portland, OR, USA.

Barrouillet, P., & Camos, V. (2001). Developmental increase in working memory span: Resource sharing or temporal decay? *Journal of Memory and Language*, *45*, 1-20. doi: 10.1006/jmla.2001.2767

Beattie, G.W., & Butterworth, B.L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, *22*(3), 201-211. doi: 10.1177/002383097902200301

Berman, R.A., & Slobin, D.I. (Eds.). (1994). *Relating events in narrative: a crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.29, retrieved from http://www.praat.org/

Braga, D., & Marques, M.A. (2004). The pragmatics of prosodic features in the political debate. In *Proceedings of the 2nd International Conference on Speech Prosody 2004* (pp. 321-324). Nara, Japan.

Brotherton, P. (1979). Speaking and not speaking: Processes for translating ideas into speech. In A.W. Siegman & S. Feldstein (Eds.), *Of Speech and Time: Temporal Speech Patterns in Interpersonal Contexts* (pp. 179-209). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brubaker, R.S. (1972). Rate and pause characteristics of oral reading. *Journal of Psycholinguistic Research*, *1*(2), 141-147. doi: 10.1007/bf01068103

Burnham, B.R. (2015). *Fundamental Statistics for the Behavioral Sciences (v. 2.1)*. Retrieved from: https://sites.google.com/site/fundamentalstatistics/home

Butcher, A. (1981). *Aspects of the Speech Pause: Phonetic Correlates and Communicative Functions*. Doctoral Dissertation, University of Kiel.

Chafe, W. (1987). Cognitive constraints on information flow. In R.S. Tomlin (Ed.), *Coherence and Grounding in Discourse* (pp. 21-51). Philadelphia, PA: John Benjamins Publishing.

Clark, M. (2013). *An Introduction to Machine Learning with Applications in R*. Retrieved from: http://web.ipac.caltech.edu/staff/fmasci/home/astro_refs/ML_inR.pdf

Clay, O.J., Edwards, J.D., Ross, L.A., Okonkwo, O., Wadley, V.G., Roth, D.L., & Ball, K.K. (2009). Visual function and cognitive speed of processing mediate age-related decline in memory span and fluid intelligence. *Journal of Aging and Health*, *21*(4), 547-566. doi: 10.1177/0898264309333326

Colozzo, P., & Whitely, C. (2014). Keeping track of characters: Factors affecting referential adequacy in children's narratives. *First Language*, *34*(2), 155-177. doi: 10.1177/0142723714522164

Cooper, P.V. (1990). Discourse production and normal aging: Performance on oral picture description tasks. *Journal of Gerontology*, *45*(5), 210-214. doi: 10.1093/geronj/45.5.p210

Cooper, W., & Paccia-Cooper, J. (1980). *Syntax and Speech*. Cambridge, MA: Harvard University Press.

Costello, A.L., & Warrington, E.K. (1989). Dynamic aphasia: The selective impairment of verbal planning. *Cortex*, *25*(1), 103-114. doi: 10.1016/S0010-9452(89)80010-3

Danahy, K., Windsor, J., & Kohnert, K. (2007). Counting span and the identification of primary language impairment. *International Journal of Language and Communication Disorders*, *42*(3), 349-365. doi: 10.1080/13682820600940083

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research*, *20*, 445-464. doi: 10.1007/bf01067637

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*(1), 1-18. doi: 10.1016/0749-596x(86)90018-5

Davies, M. (2008-present). *The Corpus of Contemporary American English (COCA): 600 million words, 1990-present.* Available at https://www.english-corpora.org/coca/

de Jong, N.H. (2016). Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 113-132: doi: 10.1515/iral-2016-9993

van Dijk, T.A. (1982). Episodes as units of discourse analysis. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 177-195). Washington, DC: Georgetown University Press.

van Dijk, T.A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.

van Donzel, M.E. (1999). *Prosodic Aspects of Information Structure in Discourse*. Den Haag: Holland Academic Graphics.

van Donzel, M.E., & Koopmans-van Beinum, F.J. (1996). Pausing strategies in discourse in Dutch. In *Proceeding of the 4th International Conference on Spoken Language Processing (ICSLP 96)* (pp. 1029-1032). Philadelphia, USA. doi: 10.1109/icslp.1996.607779

Duez, D. (1982). Silent and non-silent pauses in three speech styles. *Language and Speech*, *25*(1), 11-28. doi: 10.1177/002383098202500102

Engle, R.W., Tuholski, S.W., Laughlin, J.E., & Conway, A.R.A (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309-331. doi: 10.1037/0096-3445.128.3.309

Fairchild, A.J., & MacKinnon, D.P. (2009). A general model for testing mediation and moderation effects. *Prevention Science*, *10*(2), 87-99. doi: 10.1007/s11121-008-0109-6

Fant, G., Kruckenberg, A., & Ferreira, J.B. (2003). Individual variations in pausing. A study of read speech. *Phonum, 9*, 193-196.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30*, 210-233. doi: 10.1016/0749596x(91)90004-4

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review, 100*(2), 233-253. doi: 10.1037/0033-295x.100.2.233

Fougeron, C., & Jun, S.-A. (1998). Rate effects on French intonation: prosodic organization and phonetic realization. *Journal of Phonetics, 26*, 45-69. doi: 10.1006/jpho.1997.0062

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics, 31*, 931-952. doi: 10.1016/s0378-2166(98)00101-5

Fuchs, S., Petrone, C., Krivokapic, J., & Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics, 41*, 29-47. doi: 10.1016/j.wocn.2012.08.007

Garnham, A., Oakhill, J., & Johnson-Laird, P. N. (1982). Referential continuity and the coherence of discourse. *Cognition, 11(1)*, 29–46. doi: 10.1016/0010-0277(82)90003-8

Gaulin, C.A., & Campbell, T.F. (1994). Procedure for assessing verbal working memory in normal school-age children: Some preliminary data. *Perceptual and Motor Skills, 79*(1), 55-64. doi: 10.2466/pms.1994.79.1.55

Gayraud, F., Lee, H.-R., & Barkat-Defradas, M. (2011). Syntactic and lexical context of pauses and hesitations in the discourse of Alzheimer's patients and healthy elderly subjects. *Clinical Linguistics and Phonetics, 25*(3), 198-209. doi: 10.3109/02699206.2010.521612

Gee, J.P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15*, 411-458. doi: 10.1016/0010-0285(83)90014-2

Gee, J.P, & Grosjean, F. (1984). Empirical evidence for narrative structure. *Cognitive Science, 8*, 59-85. doi: 10.1207/s15516709cog0801_3

Gillam, R.B., & Pearson, N.A. (2004). *Test of Narrative Language*. Austin, TX: PRO-ED.

Givon, T. (1983). Topic continuity in discourse: The functional domain of switch reference. In J. Haiman & P. Munro (Eds.), *Switch Reference and Universal Grammar* (pp. 51-82). Philadelphia, PA: John Benjamins Publishing.

Goldman-Eisler, F. (1961a). The distribution of pause durations in speech. *Language and Speech*, *4*(4), 232-237. doi: 10.1177/002383096100400405

Goldman-Eisler, F. (1961b). Continuity of speech utterance, its determinants and its significance. *Language and Speech*, *4*(4), 220-231. doi: 10.1177/002383096100400404

Goldman-Eisler, F. (1961c). A comparative study of two hesitation phenomena. *Language and Speech*, *4*(1), 18-26. doi: 10.1177/002383096100400102

Goldman-Eisler, F. (1967). Sequential temporal patterns and cognitive processes in speech. *Language and Speech*, *10*(2), 122-132. doi: 10.1177/002383096701000205

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech*. London: Academic Press.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, *15*(2), 103-113. doi: 10.1177/002383097201500201

Grosjean, F., & Collins, M. (1979). Breathing, pausing and reading. *Phonetica*, *36*, 98-114. doi: 10.1159/000259950

Grosjean, F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, *11*, 58-81. doi: 10.1016/0010-0285(79)90004-5

Gustafson-Capkova, S., & Megyesi, B. (2001). A comparative study of pauses in dialogues and spontaneous speech. In *Proceedings of EUROSPEECH 2001* (pp. 931-935). Aalborg, Denmark.

Gustafson-Capkova, S., & Megyesi, B. (2002). Silence and discourse context in read speech and dialogues in Swedish. In *Proceedings of the 1st International Conference on Speech Prosody 2002* (pp. 363-366). Aix-En-Provence, France.

Hawkins, P.R. (1971). The syntactic location of hesitation pauses. *Language and Speech*, *14*(3), 277-288. doi: 10.1177/002383097101400308

Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1965). Temporal patterns of cognitive activity and breath control in speech. *Language and Speech*, *8*(4), 236-242. doi: 10.1177/002383096500800405

Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1966). Sequential temporal patterns in spontaneous speech. *Language and Speech*, *9*(4), 207-216. doi: 10.1177/002383096600900402

Hieke, A.E., Kowal, S., & O'Connell, D.C. (1983). The trouble with "articulatory" pauses. *Language and Speech, 26*(3), 203-214. doi: 10.1177/002383098302600302

Hirschberg, J., & Nakatani, C.H. (1996). A prosodic analysis of discourse segments in direction-giving monologues. In *Proceeding of the 34th Annual Meeting on Association for Computational Linguistics* (pp. 286-293). Santa Cruz, CA. doi: 10.3115/981863.981901

Hoit, J.D., & Lohmeier, H.L. (2000). Influence of continuous speaking on ventilation. *Journal of Speech, Language, and Hearing Research, 43*(5), 1240-1251. doi: 10.1044/jslhr.4305.1240

Huber, J.E., & Darling, M. (2011). Effect of Parkinson's disease on the production of structured and unstructured speaking tasks: Respiratory physiologic and linguistic considerations. *Journal of Speech, Language, and Hearing Research, 54*, 33-46. doi: 10.1044/1092-4388(2010/09-0184)

Hunt, K.W. (1965). *Grammatical structures written at three grade levels*. Urbana, IL: National Council of Teachers of English.

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. In *Lund Working Papers in Linguistics, 53* (pp. 61-79).

Jose, P.E. (2013). *Doing Statistical Mediation and Moderation*. NY: The Guilford Press.

Joshi, A., Kale, S., Chandel, S., & Pal, D.K. (2015). Likert scale: Explored and explained. British *Journal of Applied Science and Technology, 7*(4), 396-403. doi: 10.9734/bjast/2015/14975

Kallay, J.E., Mayr, U., & Redford, M.A. (2019). Characterizing the coordination of speech production and breathing. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceeding of the 19th International Congress of Phonetic Sciences* (pp. 1412-1416). Melbourne, Australia.

Kasl, S.V., & Mahl, G.F. (1965). Relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology, 1*(5), 425-433. doi: 10.1037/h0021918

Kendall, T. (2013). *Speech Rate, Pause, and Sociolinguistic Variation: Studies in Corpus Sociophonetics.* New York: Palgrave Macmillan.

Kintsch, W. (1995). How readers construct situation models for stories: The role of syntactic cues and causal inferences. In M.A. Gernsbacher & T. Givon (Eds.), *Coherence in Spontaneous Text* (pp. 139-160). Philadelphia, PA: John Benjamins Publishing.

Kintsch, W., & van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363-394.

Klatt, H. (1980). Pauses as indicators of cognitive functioning in aphasia. In H.W. Dechert & M. Raupach (Eds.), *Temporal Variables in Speech* (pp. 113-120). The Hague: Mouton.

Kolly, M.-J., Leemann, A., Boula de Mareuil, P., & Dellwo, V. (2015). Speaker-idiosyncrasy in pausing behavior: evidence from a crosslinguistic study. In *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 1-5). Glasgow, Scotland.

Koopmans-van Beinum, F.J., & van Donzel, M.E. (1996). Discourse structure and its influence on local speech rate. In *Proceedings of the Institute of Phonetic Sciences* (pp. 1-11). Amsterdam, NL.

Krause, F., & Lindemann, O. (2014). Expriment: A Python library for cognitive and neuroscientific experiments. *Behavior Research Methods*, *46*(2), 416-428. doi:10.3758/s13428-013-0390-6

Krivokapic, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics*, *35*, 162-179.

Krivokapic, J. (2012). Prosodic planning in speech production. In S. Fuchs, M. Weihrich, D. Pape, & P. Perrier (Eds.), *Speech Planning and Dynamics* (pp. 157-190). Munchen, Germany: Peter Lang.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307-322. doi: 10.1093/applin/16.3.307

Lay, C.H., & Paivio, A. (1969). The effects of task difficulty and anxiety on hesitations in speech. *Canadian Journal of Behavioural Science*, *1*(1), 25-37. doi: 10.1037/h0082683

Lee, J., & Paek, I. (2014). In search of the optimal number of response categories in a rating scale. *Journal of Psychoeducational Assessment*, *32*(7), 663-673. doi: 10.1177/0734282914522200

Linck, J.A., Osthus, P., Koeth, J.T., & Bunting, M.F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin and Review*, *21*, 861-883. doi: 10.1037/e572422013-022

Lively, S.E., Pisoni, D.B., van Summers, W., & Bernacki, R.H. (1993). Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *Journal of the Acoustical Society of America*, *93*(5), 2962-2973. doi: 10.1121/1.405815

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36-62. doi: 10.5054/tq.2011.240859

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, *96*, 190-208. doi: 10.1111/j.1540-4781.2011.01232_1.x

Mahl, G.F. (1956). Disturbances and silences in the patient's speech in psychotherapy. *The Journal of Abnormal and Social Psychology*, *53*, 1-15. doi: 10.1037/h0047552

Mandler, J.M. (1984). *Stories, Scripts, and Scenes: Aspects of Schema Theory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mandler, J.M., & Johnson, N.S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*, 111-151. doi: 10.1016/0010-0285(77)90006-8

MacLarnon, A., & Hewitt, G. (2004). Increased breathing control: Another factor in the evolution of human language. *Evolutionary Anthropology*, *13*, 181-197. doi: 10.1002/evan.20032

Maclay, H., & Osgood, C.E. (1959). Hesitation phenomena in spontaneous English speech. *WORD*, *15*(1), 19-44. doi: 10.1080/00437956.1959.11659682

Manard, M., Carabin, D., Jaspar, M., & Collette, F. (2014). Age-related decline in cognitive control: the role of fluid intelligence and processing speed. *BMC Neuroscience*, *15*(1), *Article 7*. doi: 10.1186/1471-2202-15-7

Mann, W.C., Matthiessen, C.M.I.M, & Thompson, S.A. (1989). Rhetorical Structure Theory and text analysis. In W.C. Mann and S.A. Thompson (Eds.), *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text* (pp. 39–78). Amsterdam and Philadelphia, PA: John Benjamins.

Mathy, F., Chekaf, M., & Cowan, N. (2018). Simple and complex working memory tasks allow similar benefits of information compression. *Journal of Cognition*, *1*(1), 1-12. doi: 10.5334/joc.31

McClain, R., & Goldrick, M. (2018). The neurocognitive mechanisms of speech production. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuropsychology* (pp. 1-28). NY: John Wiley & Sons.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zerembka (Ed.), *Frontiers in Econometrics* (pp. 105-142). New York: Academic Press.

McFadden, D. (1978). Quantitative methods for analyzing travel behaviour of individuals: some recent developments. In D.A. Hensher & P.R. Stopher (Eds.), *Behavioural Travel Modelling* (pp. 279-318). Croom Helm London.

Megyesi, B., & Gustafson-Capkova, S. (2001). Pausing in dialogues and read speech in Swedish: Speakers' production and listeners' interpretation. In *Proceedings of ITRW on Prosody in Speech Recognition and Understanding*. Red Bank, NJ.

Mundt, J.C., Snyder, P.J., Cannizzaro, M.S., Chappie, K., & Geralts, D.S. (2007). Vocal acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of Neurolinguistics*, *20*, 50-64. doi: 10.1016/j.jneuroling.2006.04.001

Mundt, J.C., Vogel, A.P., Feltner, D.E., & Lenderking, W.R. (2012). Vocal acoustic biomarkers of depression severity and treatment response. *Biological Psychiatry*, *72*(7), 580-587. doi: 10.1016/j.biopsych.2012.03.015

Nariyama, S. (2004). Subject ellipsis in English. *Journal of Pragmatics*, *36*, 237-264. doi: 10.1016/s0378-2166(03)00099-7

NIST/SEMATECH e-Handbook of Statistical Methods. Retrieved from: http://www.itl.nist.gov/div898/handbook/ . doi: https://doi.org/10.18434/M32189

Oliveira, M. (2002). Pausing strategies as means of information processing in spontaneous narratives. In *Proceedings of the International Conference on Speech Prosody 2002* (pp. 539-542). Aix-en-Provence, France.

Panek, D.M., & Martin, B. (1959). The relationship between GSR and speech disturbances in psychotherapy. *The Journal of Abnormal and Social Psychology*, *58*(3), 402-405. doi: 10.1037/h0045347

Pope, B., Blass, T., Siegman, A.W., & Raher, J. (1970). Anxiety and depression in speech. *Journal of Consulting and Clinical Psychology*, *35*(1), 128-133. doi: 10.1037/h0029659

Pope, B., & Siegman, A.W. (1962). The effect of therapists verbal activity level and specificity on patient productivity and speech disturbance in the initial interview. *Journal of Consulting Psychology*, *26*(5), 489. doi: 10.1037/h0046821

Preacher, K.J., Rucker, D.D., & Hayes, A.F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, *42*(1), 185-227. doi: 10.1080/00273170701341316

Preston, C.C., & Coleman, A.M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*, 1-15. doi: 10.1016/s0001-6918(99)00050-5

Rand, M.K. (1984). Story schema: Theory, research and practice. *The Reading Teacher*, *37*(4), 377-382. doi: 10.1037/h0029659

Ransdell, S., & Hecht, S. (2003). Time and resource limits on working memory: Cross-age consistency in counting span performance. *Journal of Experimental Child Psychology*, *86*, 303-313. doi: 10.1016/j.jecp.2003.08.002

Reynolds, A., & Paivio, A. (1968). Cognitive and emotional determinants of speech. *Canadian Journal of Psychology*, *22*(3), 164-175. doi: 10.1037/h0082757

Robinson, C., & Schumacker, R.E. (2009). Interaction effects: Centering, Variance Inflation Factor, and interpretation issues. *Multiple Linear Regression Viewpoints*, *35*, 6-11.

Rumelhart, D.E. (1975). Notes on a schema for stories. In D.G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 211-236). New York: Academic Press.

Salami, A., Eriksson, J., Nilsson, L.-G., & Nyberg, L. (2012). Age-related white matter microstructural differences partly mediate age-related decline in processing speed but not cognition. *Biochimica et Biophysica Acta*, *1822*, 408-415. doi: 10.1016/j.bbadis.2011.09.001

Schank, R.C., & Abelson, R.P. (1977). *Scripts, Plans, Goals and Understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schiffrin, D. (2001). Discourse markers: Language, meaning and, context. In D. Schiffrin, D. Tannen, & H.E. Hamilton (Eds.), *The Handbook of Discourse Analysis* (pp. 54-75). Malden, MA: Blackwell Publishers.

Schneider, P., & Hayward, D. (2010). Who does what to whom: Introduction of referents in children's storytelling from pictures. *Language, Speech, and Hearing Services in Schools*, *41*, 459-473. doi: 10.1044/0161-1461(2010/09-0040)

Schroder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., & Gielen, S. (2001). Acoustic correlates of emotion dimensions in view of speech synthesis. In *Proceedings of the 7th European Conference on Speech Communication and Technology* (*EUROSPEECH 2001*) (pp. 87-90). Aalborg, Denmark.

Sekaran, U. (2003). *Research Methods for Business: A Skill Building Approach*. NY: John Wiley and Sons.

Shapiro, L.R., & Hudson, J.A. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental Psychology, 27*(6), 960-974. doi: 10.1037/0012-1649.27.6.960

Sheather, S. (2009). *A modern approach to regression with R*. New York: Springer.

Siegman, A.W., & Pope, B. (1965). Effects of question specificity and anxiety-producing messages on verbal fluency in the initial interview. *Journal of Personality and Social Psychology, 2*(4), 522-530. doi: 10.1037/h0022491

Smith, A. (2006). Speech motor development: Integrating muscles, movements, and linguistic units. *Journal of Communication Disorders*, *39*, 331-349. doi: 10.1016/j.jcomdis.2006.06.017

Smith, C.L. (2004). Topic transitions and durational prosody in reading aloud: production and modeling. *Speech Communication*, *42*, 247-270. doi: 10.1016/j.specom.2003.09.004

Smith, A., & Zelaznik, H.N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology, 45*(1), 22-33. doi: 10.1002/dev.20009

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*, 103-130. doi: 10.1016/s0749-596x(02)00519-3

Solomon, N.P., & Hixon, T.J. (1993). Speech breathing in Parkinson's disease. *Journal of Speech and Hearing Research*, *36*, 294-310. doi: 10.1044/jshr.3602.294

Stein, N.L, & Glenn, C.G. (1979). An analysis of story comprehension in elementary school children. In R.O. Freedle (Ed.), *New Directions in Discourse Processing: Advances in Discourse Processes (Vol. 2)* (pp. 32-120). Norwood, NJ: Ablex.

Strangert, E. (2004). Speech chunks in conversation: Syntactic and prosodic aspects. In *Proceedings of the International Conference on Speech Prosody 2004* (pp. 305-308). Nara, Japan.

Strangert, E. (2005). Prosody in public speech: analyses of a news announcement and a political interview. In *Proceedings of the 9th European Conference on Speech Communication and Technology* (pp. 3401-3404). Lisbon, Portugal.

Swerts, M. (1997). Prosodic features of discourse boundaries of different strength. *Journal of the Acoustical Society of America*, *101*(1), 514-521. doi: 10.1121/1.418114

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, *37*(1), 21-43. doi: 10.1177/002383099403700102

Szabadi, E., Bradshaw, C.M., & Besson, J.A.O. (1976). Elongation of pause time in speech: A simple, objective measure of motor retardation in depression. *The British Journal of Psychiatry*, *129*(6), 592-597. doi: 10.1192/bjp.129.6.592

Thompson, S.A., & Mann, W.C. (1987). Rhetorical structure theory: A framework for the analysis of texts. *IPrA Papers in Pragmatics*, *1*(1), 79-105. doi: 10.1075/iprapip.1.1.03tho

Thorndyke, P.W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, *9*(1), 77-110. doi: 10.1016/0010-0285(77)90005-6

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5), 1-38. doi: 10.18637/jss.v059.i05

Towse, J.N., & Hitch, G.J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity. *The Quarterly Journal of Experimental Psychology Section A*, *48*(1), 108-124. doi: 10.1080/14640749508401379

Towse, J.N., Hitch, G.J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory and Language*, *39*, 195-217. doi: 10.1006/jmla.1998.2574

Trouvain, J., Fauth, C., & Mobius, B. (2016). Breath and non-breath pauses in fluent and disfluent phases of German and French L1 and L2 read speech. In *Proceedings of the 8th International Conference on Speech Prosody* (pp. 31-35). Boston, MA.

Trouvain, J., & Grice, M. (1999). The effect of tempo on prosodic structure. In *Proceedings of 14th International Congress of Phonetic Sciences* (pp. 1067-1070). San Francisco, CA.

Unsworth, N., & Engle, R.W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038-1066. doi: 10.1037/0033-2909.133.6.1038

Vaissiere, J. (1983). Language-independent prosodic features. In A. Cutler & D.R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 53-65). Berlin: Springer-Verlag

Whalen, D.H., & Kinsella-Shaw, J.M. (1997). Exploring the relationship of inspiration duration to utterance duration. *Phonetica*, *54*, 138-152. doi: 10.1159/000262218

Wlodarczak, M., & Heldner, M. (2017). Respiratory constraints in verbal and non-verbal communication. *Frontiers in Psychology, 8*, 1-11. doi: doi.org/10.3389/fpsyg.2017.00708

Yang, L.-C. (2004). Duration and pauses as cues to discourse boundaries in speech. In *Proceedings of the 2nd International Conference on Speech Prosody 2004*. Nara, Japan.

Yunusova, Y., Graham, N. L., Shellikeri, S., Phuong, K., Kulkarni, M., Rochon, E., Tan-Wai, D.F., Chow, T.W., Black, S.E., Zinman, L.H., & Green, J. R. (2016). Profiling speech and pausing in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). *PLOS ONE*, *11*(1), e0147573. doi: 10.1371/journal.pone.0147573

Zelazo, P.D., Anderson, J.E., Richler, J., Wallner-Allen, K., Beaumont, J.L., & Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): measuring executive function and attention. *Monographs of the Society for Research in Child Development*, *78*(4), 16-33.

Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), *Fundamentals of Speech Synthesis and Speech Recognition* (pp. 41-62). Chichester, UK: John Wiley.

Zvonik, E., & Cummins, F. (2003). The effect of surrounding phrase lengths on pause duration. In *Proceedings of the 8th European Conference on Speech Communication and Technology (EUROSPEECH2003)* (pp. 777-780). Geneva, Switzerland.