J. Krivokapić. (2012). Prosodic planning in speech production. In Susanne Fuchs, Melanie Weihrich, Daniel Pape, Pascal Perrier (eds.) Speech planning and dynamics. Peter Lang, pp. 157-190.

Prosodic planning in speech production

Jelena Krivokapić

Abstract: Theories of language production need to account for how speakers produce prosodic structure. Very few models have addressed the role of prosodic phrasing in speech planning, and little experimental work examining this question has been conducted so far. Four studies designed to investigate the effect of phrase length and prosodic phrasing on speech planning are presented. The first experiment examines the effect of prosodic structural complexity on pause duration using very long phrases (28 syllables). In a second experiment the effect of prosodic structural complexity using phrases of different length (6, 10, 14 syllables) is investigated. Experiment 3 examines the effect of prosodic complexity when the utterances are strings of numbers and Experiment 4 examines local and distant effects of prosodic phrase length on pause duration. The results show that speakers have a large scope of planning and that both local and distant prosodic phrases have an effect on the speech planning process. It is argued that prosodic structure determines the chunk to be planned by speakers. Implications for models of speech production are discussed.

1 Introduction

The goal of this study is to investigate the role of prosodic structure, in particular prosodic phrasing, in speech planning.

A large body of literature has investigated how linguistic structure influences speech planning processes, in particular focusing on syntactic structure, discourse structure, and phrase length. Studies on the effects of syntactic structure on pause duration have found that more complex structure leads to longer pauses, leading investigators to the conclusion that syntactically complex phrases are more demanding on the production system, and that longer pauses indicate the time speakers need to plan the more complex structure (e.g., Cooper and Paccia-Cooper, 1980; Ferreira, 1991; Strangert, 1997).

The effect of prosodic phrasing on pause duration is less well understood. In a set of studies we examine local and distant effects of phrase length and effects of prosodic structure on speech planning, as evident in pause duration. The goal of these studies is to gain an understanding into the nature of prosodic planning and to examine the role of prosodic structure in speech planning (particularly focusing on planning units).

2 Background

2.1 Prosodic structure

In the last couple of decades, the relevance of a hierarchically organized prosodic structure in speech production has become increasingly clear (see Shattuck-Hufnagel and Turk, 1996, for an overview). Prosodic structure refers to the suprasegmental structure of an utterance that encodes prominence and phrasing. Prominence serves to highlight important or new information (e.g., Bolinger, 1972; Halliday, 1967) while phrasing serves to chunk the stream of speech for processing purposes (Cutler et al., 1997; Frazier et al., 2006; Krivokapić, 2007a,b). The focus of this study is on the phrasing aspect of prosody, and particularly on the temporal aspects.

Prosodic structure is hierarchically organized into units, with higher units dominating lower units. In addition to tonal events, prosodic boundaries are characterized by systematic variation in the acoustic and articulatory temporal properties of segments. Research has shown that at boundaries segments increase in duration in the acoustic signal (e.g., Oller, 1973; Klatt, 1975; Shattuck-Hufnagel and Turk, 1998; Wightman et al., 1992). In articulation, it has been found that boundary adjacent gestures become temporally longer and that this lengthening increases cumulatively for larger prosodic boundaries (e.g., Byrd and Saltzman, 1998; Byrd, 2000; Cho, 2005; Edwards et al., 1991).

Depending on the theoretical framework, prosodic units can be largely inferred based on syntactic structure (e.g., Selkirk, 1984; Nespor and Vogel, 1986) or on intonational properties (e.g., Beckmann and Pierrehumbert, 1986; Pierrehumbert, 1980; Jun, 1993, with the latter providing a comparison of the previous two approaches). While there are different proposals in the literature regarding the number and precise definition of categories, in general researchers for English agree on the need for at least one minor and one major category above the word level. In this study we will use the terminology of Beckmann and Pierrehumbert (1986) and refer to these categories as the Intonational Phrase (IP) and the Intermediate Phrase (ip) respectively (for an overview of the various models see Shattuck-Hufnagel and Turk, 1996). The IP is the largest unit, defined as the domain of a coherent intonational contour that has at least a nuclear pitch accent, a phrase accent, and a boundary tone. The IP branches into intermediate phrases, which include at least a pitch accent and a phrase accent. Both IP and ip exhibit final lengthening, but IPs are lengthened more than intermediate phrases. Intermediate phrases branch further into words, and these into syllables. The *Tone and Break Indices* (ToBI) intonation transcription system (Beckman and Elam, 1997), which will be used in this study to examine prosodic structure, is based on this model for intonation and on the work of Price et al. (1991) and Wightman et al. (1992) for the break indices. The distinct prosodic break indices (signaling the perceived boundary strength) in ToBI correspond to the three distinct prosodic categories: word (break index 1), ip (break index 3) and IP (break index 4), and one more break index signaling the within a word boundary, i.e., a clitic boundary (break index 0).[1] A schematic representation of the constituents that are relevant for this study are given in Figure 1.

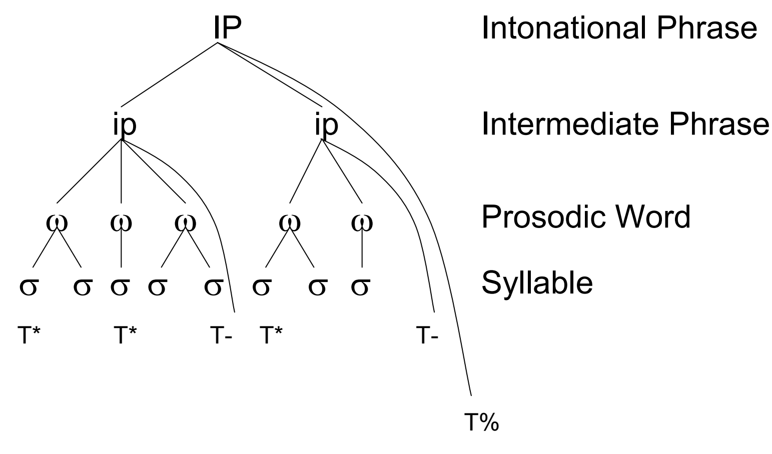

Figure 1: Prosodic constituents (adapted from Beckman, 1989). 'T*' stands for different types of pitch accents, 'T-' for phrase accents and 'T%' for boundary tones. Only the immediately relevant aspects of prosodic structure are shown.

1 Break index 2 signals a mismatch between tonal properties and the perception of the prosodic break.

It is typically assumed that prosodic structure is nonrecursive, as stated in Selkirk's Strict Layer Hypothesis (SLH, see Selkirk, 1984; Nespor and Vogel, 1986). According to this view, a prosodic category always dominates a constituent that is hierarchically immediately lower than itself, thus "in a prosodic tree, any domain at a given level of the hierarchy consists exclusively of domains at the next lower level of hierarchy" (Ladd, 1996, p.238; thus, for example, an ip cannot be dominated by another ip). While this view is generally accepted (but see for opposing views Ladd, 1988, 1996; Itô and Mester, 1992, 2010; Dresher, 1994; Wagner, 2005), there does not exist any experimental support for it, and some experimental evidence argues against this view. For example, acoustic and articulatory studies have found that IP boundaries can systematically be produced with different strength reflecting their depth of the embedding (Ladd, 1988; Byrd and Krivokapić, 2008). Listeners in turn have been found to be sensitive to these distinctions (Byrd and Krivokapić, 2008), and are generally able to distinguish between a larger number of categories than predicted under a nonrecursive view (Ladd, 1996). Based on this kind of evidence, here the view that prosodic recursion is possible is adopted (for a discussion and further arguments against recursion, see Ladd, 1996; Wagner, 2005). For a less strict version of the SLH, see Itô and Mester (1992); Selkirk (1995), where the hypothesis is stated as a set of violable constraints.

2.2 Structural effects in planning

Speech planning involves processes from conceptualization to articulation. It includes, among others, lexical selection, the planning of syntactic structure, prosodic structure, and articulation.
A large body of literature on planning has examined the occurrence and duration of pauses as an indicator of speech planning processes. The most important factors for the present discussion are the structural factors affecting planning (in particular syntactic structure and prosodic structure) and phrase length. These have been found to affect both pause occurrence and pause duration.
A number of studies have investigated the effect of syntactic complexity (syntactic branching) on pause duration and have found that syntactically complex phrases lead to longer pauses compared to syntactically

simpler phrases (e.g., Cooper and Paccia-Cooper, 1980; Grosjean et al., 1979; Ferreira, 1991; Strangert, 1991, 1997; Sanderman and Collier, 1995; Terken and Collier, 1992; Butcher, 1981). Generally, it is assumed that the reason for more complex phrases to be preceded by longer pauses is that the speaker needs more time to plan a syntactically complex utterance, compared to a syntactically simpler phrase, so the time to initiation of the post-boundary phrase increases, allowing planning to occur during this interval (e.g. Ferreira, 1991; Strangert, 1997). Particularly interesting is the study of Ferreira (1991), reporting that pause duration between a subject noun phrase and an object noun phrase (i.e., within a verb phrase) increases with the complexity of the object noun phrase, but only if the subject noun phrase is also complex. She interprets this finding as indicating that when the subject noun phrase is complex, speakers are not able to plan the object noun phrase during the production of the subject, so a pause occurs between the subject and verb phrase, which gives the speaker the time needed for planning the upcoming phrase. In other words, she suggests that planning proceeds in chunks, over structurally defined units. This notion will be particularly relevant in the discussion of the findings presented here (see also Krivokapić, 2007a,b). While these studies have shown that syntactic structure influences pause duration, a large body of research indicates that prosodic structure might be a better predictor of pause duration. A study by Gee and Grosjean (1983) examines the data collected by Grosjean et al. (1979) and finds that pause duration can be better predicted if both syntactic and prosodic structure is used, rather than just syntactic structure. Additionally, Grosjean et al. (1979) examine pause occurrence and show a tendency for speakers to divide phrases into smaller chunks of equal length, even if syntactic structure would lead to a different phrasing, indicating that rhythmical aspects of prosodic structure might be the determining factor of pause occurrence in such cases (see also Krivokapić, 2007a, 154ff on rhythmic effects in prosodic phrasing). Examining the perception of pauses in spontaneous speech Martin (1970) finds that prosodic structure (final lengthening) overrides syntactic structure. Similarly, Ferreira (1993) finds that final lengthening and pause duration are determined by prosodic rather than syntactic structure. These studies lead to the conclusion that prosodic structure might be the critical factor in determining pause duration. The effects of prosodic structure on pause duration are also found in a study by Horne et al. (1995) where it was found that pause duration increases with prosodic boundary strength (see also Choi, 2003).

Phrase length has also been found to have an impact on pause duration. Sternberg et al. (1978) have found that an increase in the number of words in a string of words leads to longer initiation times of the utterance. They suggest that the longer initiation times reflect the time speakers need to retrieve motor programs for the words. In sentences, Zvonik and Cummins (2003) show that the length of prosodic phrases before and after the boundary has an effect on pause duration. Pauses less than 300ms almost exclusively occur with phrases of ten or less syllables before or after the boundary, and the likelihood of a pause being less than 300ms increases when both the prosodic phrase before and after the boundary are less than ten syllables (Zvonik and Cummins, 2003). A study by Ferreira (1991) also finds a length effect, in that she shows that sentence initiation time increases with the number of phonological words in a sentence. Similarly, Watson and Gibson (2004) show that pause occurrence is influenced by phrase length, in that the number of phonological phrases within a preceding and a following syntactic phrase (and thus, the length of the syntactic phrase) is a good indicator of the likelihood of an IP boundary (and therefore of the likelihood of pause occurrence, since IP boundaries are often marked by pauses).

In summary, previous work has provided evidence of the influence of phrase length and even more robust evidence of the influence of syntactic structure on pause duration and speech planning. While some evidence indicates that prosodic structure might play a role in speech planning as well, much less is known about this factor. Furthermore, the process of planning of prosodic structure is also not understood well. We turn to these questions in the remainder of the article.

2.3 Theories of prosodic planning

In contrast to research examining the planning of syntactic structure, and word production models, the planning of larger prosodic phrases has not received much attention in the literature. In what follows we examine two views on prosodic planning, focusing in particular on the question of how incrementaly prosodic structure is planned and on the question of when in the speech production process prosodic structure is built. A non-controversial assumption in models of planning is that speech is produced incrementally, that is, speakers do not plan the whole utterance before speech onset, but instead plan and produce speech simultaneously (e.g., Ferreira, 1996; Ferreira and Swets, 2002; Keating and Shattuck-Hufnagel, 2002; Meyer et al., 2007; Levelt, 1989; Roelofs, 1998; Wheeldon and Lahiri, 1997). Thus speakers plan an utterance in stages (e.g., the stage of planning the conceptual content of a sentence, the stage of its syntactic encoding), and the planning of different fragments of an utterance can proceed in parallel, at different stages. For example, as soon as one fragment is ready, speakers can start articulating, without having entirely planned the rest of the utterance, and the planning of the rest of the utterance proceeds while the speaker is articulating. This allows for efficient and fluent speech production.

It is less agreed upon whether the speech production system is "architecturally incremental" (Ferreira and Swets, 2002), in the sense that incrementality is automatic and a part of the architecture of the language production system. The consequence of such a design of the production system would be that speakers obligatorily start speaking the moment a minimal unit is encoded. Ferreira and Swets (2002) specifically address this question by examining subjects' production when asked to calculate sums of two-digit numbers at their own pace and under time pressure. They show that speakers are able to produce speech in a highly incremental manner, but given the option, they plan further ahead and don't start articulating immediately when encoding is finished for a chunk of the utterance. Ferreira and Swets (2002) argue that the speech production system is not architecturally, but strategically incremental, in the sense that it is the speaker's choice how incrementally they produce speech, and they can choose to plan further than just the smallest unit necessary (Ferreira and Swets, 2002; Wagner et al., 2010). They further show that speakers, even when under time pressure, do plan to some extent up to the end of the sentence (see also Griffin, 2003, for further evidence of a sentential planning scope). The implication of this finding is that speakers are able to plan more than one planning unit at a time, and plan quite large chunks of speech before speech onset.

A second question of interest is how prosodic structure - prosodic phrasing in particular - is planned, and what role in the speech planning process it plays. While the planning of prosodic structure in its entirety is still a largely unexplored field, two approaches have emerged in the literature: a prosody-first approach, developed in Keating and Shattuck-Hufnagel (2002), and a more incremental, prosody-last approach, developed by Levelt (Levelt, 1989; Levelt et al., 1999, see also a related proposal by Ferreira, 1993).

Levelt (1989) is the first model to incorporate prosodic structure into a model of speech production. The approach taken in this model is that speech production is strictly incremental, ideally without any lookahead. Prosodic phrases are created in the Prosody Generator component of the model by scanning the syntactic structure. Thus ip boundaries are placed after a phonological word that is the head of a syntactic constituent (although an adjunct can optionally be included into the ip) and IPs result from the speaker's decision to break at a certain point in an utterance. While the possible locations of the IP boundary are constrained by the syntactic structure, it is the speaker's choice to create a boundary. Whether a speaker does so depends on a number of factors, for example the length of the phrase at that point, the availability of new syntactic material (in the sense that if no further material is available for processing, the speaker is forced to pause), speech rate, prominence of a word, and the speaker's desire to be intelligible. Crucially, in this model prosodic boundaries are determined by looking at just one word at the time, thus with very little lookahead. While Levelt (1989) points out that speakers can occasionally have a larger lookahead (for example in very carefully produced speech) generally the assumption is that there is no lookahead.

Another view is presented in Keating and Shattuck-Hufnagel (2002). They argue that a body of evidence suggests that speakers have a large lookahead in speech production. Regarding phrasing they note that it is influenced, among other things, by the length of not just the preceding but also of the following prosodic phrase (see e.g., Gee and Grosjean, 1983; Watson and Gibson, 2004). This indicates a larger lookahead than the one suggested by Levelt (1989, see Keating and Shattuck-Hufnagel, 2002 for further evidence for a larger lookahead, such as speech errors, stress clash, resyllabification and assimilation patterns). It is also known that speakers use information about prosodic structure during phonetic encoding (e.g., temporal properties of gestures will depend on whether a syllable is stressed or not). This kind of evidence suggests, Keating and Shattuck-Hufnagel argue, that prosodic structure is an essential part of the process of phonological and phonetic encoding, and that it therefore needs to be available early on. Keating and Shattuck-Hufnagel (2002) suggest a "prosody-first" approach where prior to phonological encoding a rough outline of the prosodic structure (a "skeletal default prosody", Keating and Shattuck-Hufnagel, 2002, p.139) is built based on syntactic information. This representation contains the prosodic phrasing, the marking of the relative prominence of individual constituents, pitch accent marking, and possibly some tonal marking of boundaries. The representation is then restructured in the course of the encoding process based on word form and non-syntactic information. With prosodic structure being available early, prosodic information can inform the encoding process. Keating and Shattuck-Hufnagel's (2002) model necessarily has a large lookahead, as it requires that prosodic structure is built before individual segments can be encoded. The model does not hinge on the exact size of the syntactic chunk available for further planning, it is only important that "the increments ... [are] large enough to account for the facts of phonological and phonetic segmental sensitivity to prosodic structure" (Keating and Shattuck-Hufnagel, 2002, p.139).

In a series of four experiments, we investigate how prosodic phrasing and phrase length affect pause duration. The goal is 1) to examine the role prosodic structure plays in speech planning, and 2) to examine how far ahead speakers plan their utterances and in that way contribute to the discussion of incrementality and to the question whether prosodic structure is build in a "prosody-first" (Keating and Shattuck-Hufnagel, 2002) or a "prosody last" (Levelt, 1989) manner.

Foreshadowing the conclusions of this study, an outline of our view of the role of prosodic structure in speech planning follows: We will assume Keating and Shattuck-Hufnagel (2002)'s "prosody-first" approach, for which we will present additional evidence. Thus based largely on the syntactic structure, speakers build the prosodic structure, as suggested in their model. It will be argued in Experiment 1 that prosodic structure has a critical role in speech planning, namely that it determines the chunk of speech to be phonologically and phonetically encoded at a time. This means that speech planning proceeds over prosodic units (Krivokapić, 2007b). It will be suggested that either the ip or the IP - any sufficiently large prosodic phrase - can serve as a planning unit. Which prosodic phrase is selected depends on the cognitive load of the upcoming phrase, with hierarchically lower units being selected for cognitively more demanding material. The following four experiments present arguments for this role of prosodic structure and examine how cognitive load affects which prosodic phrase in the hierarchy does in the end serve as the planning unit.

3 Experimental evidence

We examine the effect of prosodic phrasing and how it interacts with phrase length in four experiments. The first experiment (see Krivokapić, 2007b, Experiment 2) examines the effect of prosodic structural complexity (the number of prosodic phrases in an utterance) on pause duration using long phrases (28 syllables before and after the pause). To examine possible interactions of prosodic complexity and phrase length, the second experiment examines the effect of prosodic complexity on pause duration using phrases of different length (6, 10, 14 syllables before and after the pause). The effect of prosodic complexity on pause duration using strings of numbers is investigated in Experiment 3 and Experiment 4 examines the effect of phrase length of phrases immediately after and further away from a boundary. Given that the goal of this article is to examine speech planning, in reporting the results we focus on post-boundary effects.

The four studies are conducted using the synchronous speech paradigm introduced by Cummins (Cummins, 2002, 2003, 2004; Zvonik and Cummins, 2002). In this paradigm, two subjects (one dyad) are seated facing each other and they read sentences simultaneously, at the prompt of the experimenter. Synchronous speech has been shown to reduce inter-speaker variability in pause placement and pause duration without introducing artificial properties into a subject's speech. For example, Cummins (2004) compares the ratio of boundary duration to phrase length in subjects' synchronous and solo productions and finds that the ratios are similar for the two speech styles, but that the variability is reduced in synchronous speech. He finds similar results for the subjects' ratios of phrase length of two phrases. Zvonik and Cummins (2003) fur-

ther find that both in solo and in synchronous speech speakers had the longest pause following the longest phrase, again indicating that while reducing variability, synchronous speech does not change fundamental properties of speech timing. Given these properties of synchronous speech, and given the large variability speakers exhibit in both duration and placement of pauses (e.g., Goldman Eisler, 1968; Butcher, 1981), this experimental paradigm provides a good way to examine pause duration, as it will minimize individual variation and facilitate comparisons across speakers.
It is worth noting that the experimental design used here - as is the case with the majority of laboratory speech studies - involves certain processes not typical of spontaneous speech, such as reading, repetition, and, in this case, simultaneous speaking with another speaker. Results of such laboratory experiments might not represent the most typical behavior in spontaneous speech processing, however, such studies have a long history of informing us usefully as to properties of the speech production system (see also Ferreira, 1991, 2007, for a discussion of the merits and concerns of laboratory speech).

3.1 Experiment 1

3.1.1 Goal

The first experiment examines the effect of prosodic complexity on pause duration in very long phrases (for a detailed description of the study see Krivokapić, 2007b, Experiment 2). Prosodic complexity is taken to mean, similar to syntactic complexity, the number of prosodic phrases in an utterance. Thus if for example a post-boundary phrase branches into two Intonational Phrases it is prosodically more complex than a nonbranching post-boundary phrase that consists of only one IP (see Figure 2).

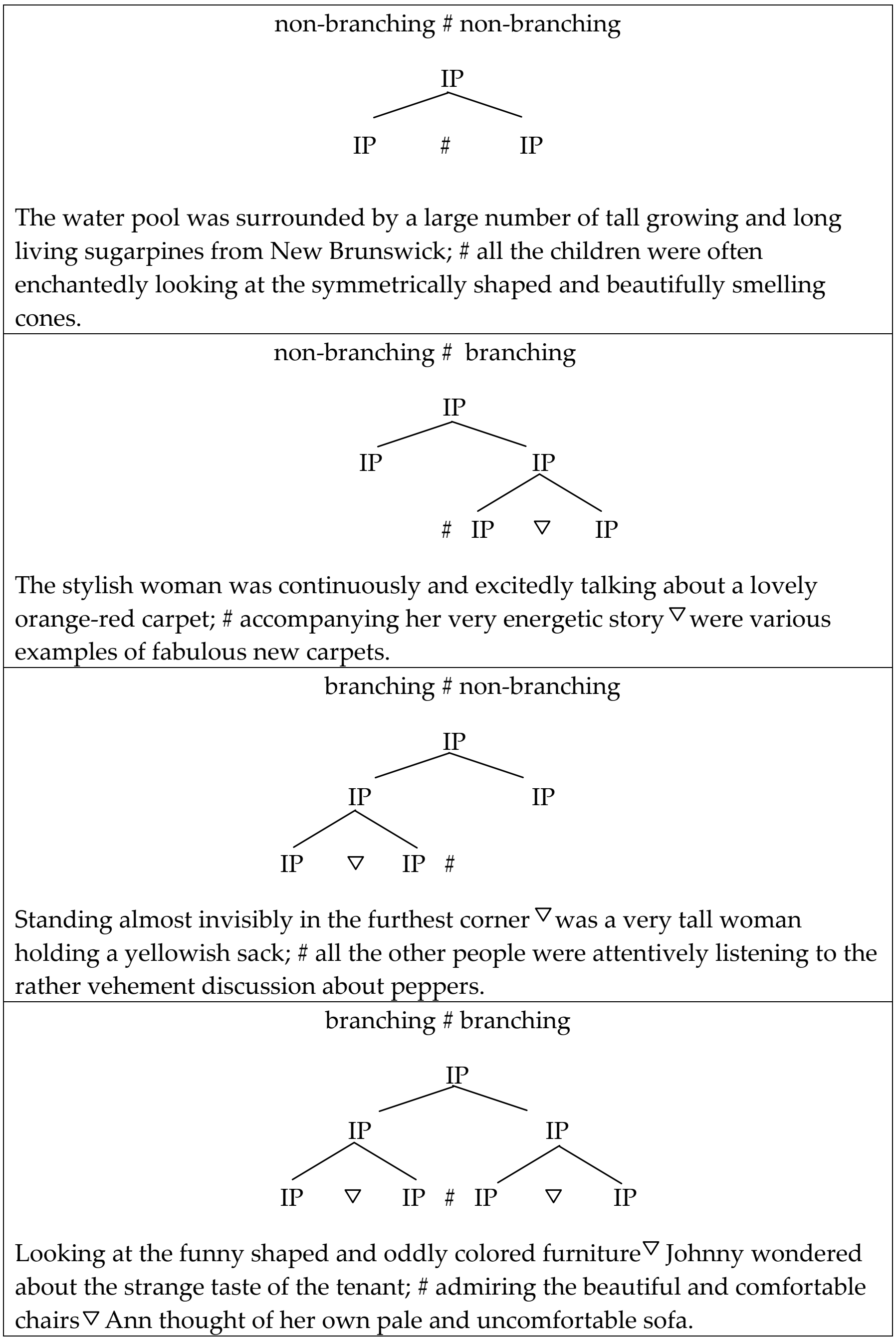

Figure 2: Structural conditions for Experiments 1, 2, and 3 and sample experiment stimuli for Experiment 1. "Branching" indicates a complex, and "non-branching" a simple structure. The measured pause is indicated with “#”, and the branching is indicated with the triangle. These symbols were not present in the sentences the subjects read.

3.1.2 Design and materials

The independent variables in the experiment were: a) complexity of the pre-boundary phrase (whether the pre-boundary phrase was branching into two prosodic phrases or not) and b) complexity of the post-boundary phrase (whether the post-boundary phrase was branching into two prosodic phrases or not). These factors were crossed, yielding four conditions (branching # branching, branching # non-branching, non-branching # branching, non-branching # non-branching, shown in Figure 2). The pause duration between the pre- and post-boundary phrase was the dependent variable. The whole pre-boundary phrase was always 28 syllables, with branching targeted at the 14th syllable, and the same was true for the post-boundary phrase. The targeted prosodic structure was elicited by varying the syntactic structure and the length of syntactic constituents.

There were three utterances for each condition, for a total of twelve utterances. An example for each of the conditions is given in Figure 2. The stimuli for this experiment were combined with stimuli for another experiment (see Krivokapić, 2007b, for randomization procedures). Each item was read three times by each subject, yielding 36 items per subject.

3.1.3 Subjects, recording, and analyses

Data from sixteen speakers (eight dyads) were collected. One dyad was unable to complete the experiment. The results of the remaining seven dyads are reported here. For each dyad, the two subjects were seated facing each other. Before the recording, they familiarized themselves with the stimuli. Once familiar with them, the subjects were asked to read the utterances aloud, at the prompt of the experimenter, together with their co-speaker, as if reading a story to someone. In cases of errors, they were asked to read the utterances again.

Subjects were recorded on a DAT recorder, using two Shure head mounted unidirectional microphones and recordings were made onto the left and right channels of a stereo file.

In order to verify that speakers produced the intended branching and nonbranching structures, the sentences were prosodically analyzed using the Tone and Break Indices labeling system (ToBI Beckman and Elam, 1997). Overall, the branching structures were produced as two IPs, the non-branching ones as one IP. IPs were identified by final lengthening and the presence of a phrase accent and a boundary tone. Only the tokens where both subjects of a dyad produced the targeted structure were

included in the analysis. In all stimuli, the pre-boundary phrase ended in a voiced segment followed by a voiceless stop, and the post-boundary phrase started with a vowel. The duration of the pause was measured from the end of periodic voicing (i.e., the beginning of the voiceless stop closure) to the beginning of periodic voicing for the vowel of the post-boundary phrase. At the end of the pause, occasionally there was a glottal pulse preceding the regular periodic voicing for the vowel, which was taken to be evidence for glotalization of the phrase initial vowel (see Redi and Shattuck-Hufnagel, 2001). In these cases, the pause was taken to end at the onset of the glottal pulse. In all other cases, the pause was taken to start with the end of periodic voicing and to end with the onset of periodic voicing. The threshold for silent intervals to be counted as pauses was 200ms. This threshold was set since all silent intervals included a voiceless stops and shorter intervals could not with certainty be assumed to be pauses. All silent intervals were longer than 200ms.
The duration of the pause was averaged between the two speakers of each dyad. A two-factor ANOVA was performed on these data for each dyad separately testing the effect of the pre-boundary prosodic complexity (with the two levels branching and nonbranching) and the post-boundary prosodic complexity factor (with the two levels branching and non-branching) and their interaction. If the results of the individual dyads patterned in the same manner, the dyads were pooled. To pool the data across the dyads, z-scores were calculated over all tokens of each dyad separately. Note that a z-score above 0 represents pause durations longer than the dyads' average duration, and a z-score below 0 durations shorter than the average duration. Significance for all tests was set at $p<0.05$. In this, as in the other experiments, in reporting the results we focus on the post-boundary effects.

3.1.4 Results

For the pooled data, post-boundary complex phrases lead to significantly ($F(1,185)=23.076$, $p<.001$) shorter preceding pauses than post-boundary simple phrases, as shown in Figure 3. The same pattern obtained for each dyad separately, though the effect reached significance only for two dyads. The general view of pause duration as it relates to upcoming phrases is that it reflects effects of speech planning processes, in that longer pauses occur in conjunction with longer planning time for the upcoming phrase (e.g., Cooper and Paccia-Cooper, 1980; Ferreira, 1991; Griffin, 2003; Smith and Wheeldon, 1999).

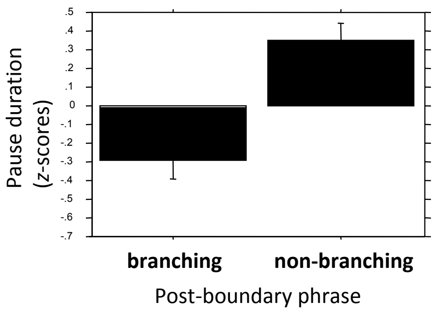

Figure 3: Experiment 1: Effect of post-boundary branching on pause duration (*z*-scores and standard errors) for all dyads pooled. *Z*-scores above 0 represent durations longer than the dyads' average pause duration, and *z*-scores below 0 durations shorter than the average duration.

Based on this, the prediction for this study was that more complex prosodic phrases would lead to longer preceding pauses. Such a result would also correspond to previous studies examining the effect of syntactic structure on pause duration (e.g., Cooper and Paccia-Cooper, 1980; Ferreira, 1991; Strangert, 1997). However, contrary to this prediction, the results showed that shorter pauses precede branching phrases, i.e., more complex prosodic structure leads to shorter preceding pauses. To account for these surprising findings, following Ferreira's (1991) suggestion that speech encoding proceeds over structurally defined units, it was suggested in Krivokapić (2007b) that in the case of the post-boundary branching structure, speakers may be planning only up to the branching node (i.e., planning the first post-boundary IP) before starting to speak the post-boundary phrase. The second post-boundary IP is then planned during the production of the first IP. In the case of the nonbranching post-boundary phrase, speakers plan the complete upcoming phrase. Thus speakers were, in all instances, planning the first upcoming IP, but in the case of branching structures, the first post-boundary IP was 14 syllables long, while in the case of non-branching structures, the first IP was 28 syllables long. The pause duration, on this assumption, reflects the time it takes for the speakers to plan the upcoming prosodic phrase. The

implication of this interpretation is that prosodic structure participates in determining the size of the chunk to be phonologically/phonetically encoded at a time. In other words, phonological/phonetic encoding proceeds over prosodic units. Depending on the size of the unit, the pause will be shorter or longer. Note that the suggestion is not that speakers encode the complete upcoming phrase before they start articulating. Rather, following Keating and Shattuck-Hufnagel (2002), the suggestion is that speakers encode the prosodic structure to a certain extent, and continue with the planning process as they are articulating.
Another point to note is that the result of the experiment implies that speakers have a large scope of planning, as they were aware, at the pause, of a branching structure 14 syllables away. This provides some evidence against the view (in Levelt, 1989; Levelt et al., 1999) that speakers produce speech with very little lookahead.

3.2 Experiment 2

3.2.1 Goals

To further examine the scope of planning and the idea that planning proceeds over prosodic units, a second experiment was conducted (as reported in Krivokapić, 2007a). While the first experiment examined the effect of prosodic structural complexity on pause duration using very long phrases (28 syllables), Experiment 2 examined the effect of prosodic structural complexity on pause duration using phrases of different length. The goal is to examine whether a decrease in cognitive load (such as a decrease in phrase length) will affect how prosodic structure affects the planning process.

3.2.2 Design and materials

This study had a 3-factor design, examining the effects of pre-boundary complexity (with the levels branching and non-branching), post-boundary complexity (with the post-boundary phrase being branching or non-branching) and surrounding phrase length, with the levels short (six syllables before and six syllables after the boundary), medium (ten syllables before and ten syllables after the boundary) and long (14 syllables before and 14 syllables after the boundary). The factors were crossed, for a total of 12 conditions (2 pre-boundary x 2 post-boundary x 3 length; Note that the prosodic structures of the stimuli were the same as is Experiment 1, shown in Figure 2). In the complex phrases, branching was targeted at the middle of the phrases (i.e., at the third, fifth and seventh syllable for the short, medium and long sentences respectively). An example utterance (for the short branching # branching condition), is "Although mad, she rang Chap. # Abe picked up, but called him", with branching targeted at the third syllable before and after the boundary (at the comma), and the examined pause being between "Chap" and "Abe". In all utterances, each pre-boundary phrase ended in "Chap" and each post-boundary phrase started with "Abe". There was one utterance per condition. To avoid memorization of the utterances, there were also twelve filler items, matching the experimental items in prosodic structure and length but with different lexical content. The utterances were randomized in blocks of 24 (12 test plus 12 filler utterances). Seven dyads were recorded. In total, 984 test utterances were recorded for analysis (12 utterances x 12 repetitions x 6 dyads, and for one dyad 12 utterances x 10 repetitions). Note that while this study had only one utterance per condition, a follow-up study was conducted, with 4 utterances per condition, varying in syntactic structure and segments surrounding the pause. Only results which were replicated in the verification study are presented here.

3.2.3 Subjects, recording, and analyses

Data from 14 speakers (7 dyads) were collected. The synchronous speech paradigm, described earlier, was used in this study as well. The recording procedure and the pause duration measurements were the same as in Experiment 1. Prosodic labeling (described in Experiment 1) verified that the branching structures were produced as two IPs, the nonbranching ones as one IP. A three-factor ANOVA was performed on the data of each dyad separately, testing the effect of the three factors (pre-boundary complexity, post-boundary complexity, phrase length) and their interactions. If the results of the individual dyads patterned in the same manner, the dyads were pooled, as described in Experiment 1. Significance for all tests was set at $p<0.05$.

3.2.4 Results

For phrase length, the predicted outcome was that with longer phrases, pause duration would increase. Length effects have been reported in the literature and it is assumed that more length will lead to longer pauses as speakers will need more time to process the upcoming phrase. For the pooled dyads, the results showed that phrase length had the expected effect (Figure 4): An increase in the number of syllables led to an increase in pause duration ($F(2, 918) = 52.671$, $p<.0001$). Fisher's PLSD shows a significant difference between all 3 levels of length, such that surrounding longer phrases lead to longer pauses (long compared to medium $p=.0011$, long compared to short $p<.0001$, medium compared to short $p<.0001$). The results indicate that more phonological length (syllables) in an upcoming phrase requires longer planning times, which leads to longer pauses.

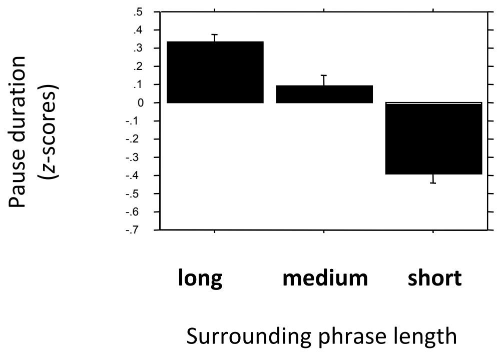

Figure 4: Experiment 2: Effect of length on pause duration (z-scores and standard errors) for all dyads pooled. Z-scores above 0 represent durations longer than the dyads' average pause duration, and z-scores below 0 durations shorter than the average duration.

For post-boundary complexity, there are two possible outcomes, based on previous work: The first possible outcome, based on the results in Experiment 1, is that prosodically complex (branching) structures will induce shorter pauses than non-branching phrases. The second possibility, in line with findings of studies examining the effect of syntactic structure on pause duration, is that phrasal complexity could lead to longer pauses. The reason to expect this latter effect is that the phrases examined in this study were considerably shorter than the ones in Experiment 1 (Krivokapić, 2007b) in which pre-boundary and post-boundary phrases were each 28 syllables. The phrases in Experiment 2 are 6, 10 and 14 syllables. In this case prosodic structure might not be chunking the upcoming phrase into smaller units for encoding, as the upcoming phrase might "as is" be short enough for the processing system to manage in one chunk. The two predictions could combine, in that for shorter phrases, prosodic complexity may increase pause duration, while for longer phrases complexity may decrease pause duration. In other words, as the length of the phrase increases and the load of the upcoming phrase increases as well, prosodic structure may chunk the upcoming phrase, causing pause duration to become shorter for branching/complex than for non-branching/simple phrases. On this scenario, in the long branching condition pauses would be shorter than in the long non-branching condition, while in the short branching condition pauses would be longer than in the short non-branching condition.

For all dyads pooled, the results show that complex phrases induced longer pauses ($F(1,918)=4.805$, $p=.0286$), as shown in Figure 5. Individual dyads patterned in the same manner, and the effect reached significance for one dyad.

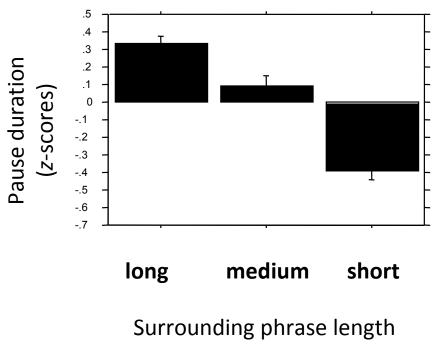

Figure 5: Experiment 2. Effect of post-boundary branching on pause duration (*z*-scores and standard errors) for all dyads pooled. *Z*-scores above 0 represent durations longer than the dyads' average pause duration, and *z*-scores below 0 durations shorter than the average duration.

This indicates that, like in previous, syntactic studies, structural complexity increased pause duration as more structure for the same length needs to be planned. These findings differ however from the findings in Experiment 1 (Krivokapić, 2007b), where more complex structure led to shorter pauses. Both findings can be accounted for by assuming that prosodic structure participates in determining the size of the chunk to be planned. Depending on the size of a potential chunk, speakers might plan only up to the branching node rather than the whole post-boundary phrase, or, for smaller utterances, they might plan the whole post-boundary phrase. In other words, depending on the size of the chunk (and probably other factors) a hierarchically higher or lower IP (presumably ip as well) might be the chunk to be planned (Krivokapić, 2007a). Note that this interpretation assumes prosodic recursion, but we suggest that the crucial point is that a hierarchically higher or lower category can determine the chunk to be encoded, while the exact type of the category is not important, as any prosodic category above the word level could potentially be the chunk to be planned.[2]

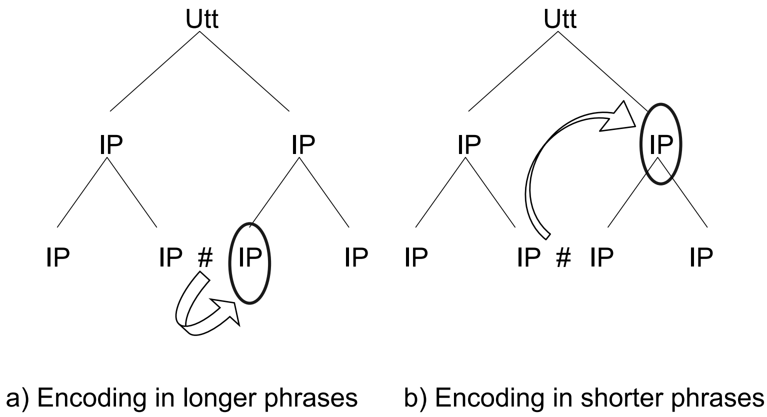

Figure 6: Prosodic chunking. The arrow points to the prosodic chunk that is planned during the pause (marked with #).

2 In addition to the size of the potential chunk, other factors are likely to influence whether a hierarchically higher or lower prosodic phrase will be the chunk to be planned by the speaker. A likely factor is individual differences. Swets et al. (2007) for example showed that individual differences influence prosodic phrasing in silent reading. Readers with a low working memory had a greater tendency to break up larger chunks of text into smaller chunks than readers with a high working memory. These findings could be due to a difference in planning scope. Petrone et al. (2011) show that working memory span affects the scope of planning, as evidenced in phrase initial F0, again indicating an effect of individual differences. Simultaneous cognitive demands on the speaker could also have an impact such that the chunks to be planned at a time become smaller.

Thus when phrases are shorter (as in Experiment 2) speakers plan the whole post-boundary phrase and do not chunk it into smaller planning units. In this case then, the prosodically more complex phrases lead to longer planning time as there is more structure to be planned. On this analysis, prosodic structure interacts with phrase length in heterogeneous ways. For very long phrases, speakers use hierarchically lower phrases to determine the planning chunk, but when the phrases are shorter, hierarchically higher phrases determine the chunk to be planned (Figure 6).

3.3 Experiment 3

3.3.1 Goals

This study examines how pause duration is affected by prosodic structure in strings of numbers. The goal is to examine the scope of planning in a string that is more difficult to plan due to the near absence of semantic meaning and syntactic structure.

3.3.2 Design and materials

Four strings were constructed to measure the effect of prosodic structure on pause duration.[3] As in Experiment 1, the independent variables were a) the complexity of the pre-boundary phrase (whether the pre-boundary phrase was branching into two prosodic phrases or not) and b) the complexity of the post-boundary phrase (whether the post-boundary phrase was branching into two prosodic phrases or not). As in the first study, the branching and non-branching strings consisted of 28 syllables, and the branching was targeted at the 14th syllable. One number string was constructed for each condition. Subjects saw each number string both as a number (as in 1a) and spelled out, as shown in (1b). It was assumed that reading the numbers just by themselves would be too difficult, therefore the subjects were asked to look at the numbers (as in 1a) first, but to read aloud the numbers written as words (as in 1b). In order to induce the prosodic branching, the branching number strings were visually grouped, and when spelled out, they were connected with

3 I would like to thank Jean-Roger Vergnaud for suggesting this experiment.

"and". Thus for example the 'branching # non-branching' condition was as given in (1a/1b). In (1a) the first line is the pre-boundary branching condition, and the second line is the post-boundary nonbranching condition. The branching occurs before "772 308" in the string in (1a), and in the spelled version in (1b) before "and". To avoid additional branching, the complete pre-boundary line, and the complete post-boundary condition on the second line (not as represented below, where the pre- and post-boundary condition didn't fit on one line each). The measured pause was, as in the previous experiments, between the pre- and the post-boundary condition. In the example below, the pause was between the number "eight" of the pre-boundary phrase and the number "one" of the post-boundary phrase. All pre-boundary phrases ended in a voiceless stop, and all post-boundary phrases started with a vowel. Each utterance was repeated three times and the data were randomized in blocks of four utterances, for a total of 12 utterances per subject.

(1a) Branching # non-branching utterance presented as numbers

952 469 772 308
177 977 797

(1b) Branching # non-branching utterance presented as words

nine hundred fifty two thousand four hundred sixty nine *and*
seven hundred seventy two thousand three hundred eight

one hundred seventy seven million nine hundred seventy seven
thousand seven hundred ninety seven

3.3.3 Subjects, recording, and analyses

Data for 14 subjects (seven dyads) were collected. The synchronous speech paradigm was used in this study as well. The recording procedure, the pause duration measurements, the data pooling process, and the statistical analysis were the same as in Experiment 1.

As in the previous experiments, a prosodic analysis (the ToBI labeling system, see Beckman and Elam, 1997) was used to verify that subjects produced the targeted prosodic structure. The analysis showed that all utterances were produced with the anticipated prosody (branching structures were produced as two IPs, the non-branching ones as one IP), but, in addition, all utterances had IPs branching into intermediate phrases. The nonbranching utterances were thus produced as branching structures as well.

3.3.4 Results

The results showed that there were no effects of prosodic complexity. This leads to the question whether number strings are planned differently than other utterances, such that prosodic structure does not have an effect on planning. We suggest that the difference between strings of numbers and regular utterances is not as surprising as it might seem at first: A string of unrelated numbers not connected conceptually, and with an impoverished syntactic structure, is cognitively more demanding than a coherent sentence. In this case then, speakers might be producing speech more incrementally, i.e., they might use smaller chunks for planning. If we assume that prosodic constituents are not qualitatively different for purposes of planning, any prosodic constituent could be used to chunk speech into planning units. In the case of cognitively demanding tasks, as numbers most likely are, speakers might choose to plan an ip, and not an IP. Since all utterances in Experiment 3 were produced with ip-branching, if subjects planned ip rather than IP phrases, the lack of effect can be accounted for.

3.4 Experiment 4

3.4.1 Goals

It is known that boundary strength is affected by properties of phrases immediately adjacent to the boundary (e.g., by phonological length, syntactic structure, and, as has been established in the previous studies, by prosodic structure). What is not well understood is the influence that phrases further away from the boundary have on pause duration. The goal of the last study is to examine local and distant effects of prosodic phrase length on boundary strength (see Krivokapić, 2010, for further details of the experiment). The larger goal is to examine incrementality in speech production, as it relates to prosodic planning.

In a study examining word planning effects, Griffin (2003) found that in a two word sequence, when a short word precedes a long word, speech onset occurs later than when the first word is also long. Griffin (2003) interprets these findings as indicating that in the short-long sequence the first word is not immediately articulated when planned, but buffered while the second word is being sufficiently prepared (enough to avoid disfluency). In the long-long sequence, the first word is immediately articulated, as there will be enough time for the preparation of the second word during the production of the first (cf. also Ferreira, 1991).
A second line of research finds that distant prosodic boundaries affect boundary strength in both production and perception (e.g., Wagner, 2005; Schafer, 1997; Carlson et al., 2001; Clifton et al., 2002; Jun, 2003). A study by Frazier et al. (2004) has found that the naturalness in prosodic boundary production (as judged by listeners) depends not just on the strength of a specific boundary but also on the strength of surrounding boundaries. Studies by Schafer (1997); Carlson et al. (2001); Clifton et al. (2002); Jun (2003) have shown that listeners' interpretation of boundary strength depends on the boundary strength of surrounding boundaries, indicating that global prosodic structure guides listeners' interpretation. Given such distant effects, and the fact that speakers may plan more than one structural unit at a time, the question arises whether more than one prosodic phrase might be planned as well - at least to some degree - before speech onset. Experiment 4 examines the effect of the length of the first and the second post-boundary IP on pause duration.

3.4.2 Design and materials

The study was a two factor design: a) post-boundary length (short or long) of the first IP following a target pause and b) post-boundary length (short or long) of the second, more distant IP following the pause. The factors were crossed, yielding four conditions. The dependent variable was the pause before the first post-boundary phrase (see Table 1). Note that it was not possible to systematically vary phrase length while keeping overall post-boundary length constant. In order to control for the overall length of the post-boundary phrase, the duration of the second phrase varied, depending on whether the first phrase was two or four syllables, as seen in Table 1. To eliminate pre-boundary effects, the pre-boundary sentence was always the same ("Bob was buying books for Sam"), and was seven syllables long. There were six utterances for each of the four conditions and twelve repetitions of each utterance, for a total of 288 utterances per subject. An example for each of the conditions is given in Table 1.

Table 1: Experimental conditions with one example utterance. The comma separates the two post-boundary IPs (as verified in the ToBI transcription). The pre-boundary phrase was always "Bob was buying books for Sam". # indicates the measured pause.

Post-boundary phrase (IP_1, IP_2)
Short first phrase, short second phrase (2 + 10 syllables) # Zack sang, claiming that this would help him choose the books.
Short first phrase, long second phrase (2 + 16 syllables) # Zack sang, claiming that this would help him choose the best children's book in the store.
Long first phrase, short second phrase (4 + 8 syllables) # Zack sang loudly, claiming that this would be helpful.
Long first phrase, long second phrase (4 + 14 syllables) # Zack sang loudly, claiming that this would help him choose the best book in the store.

3.4.3 Subjects, recording, and analyses

Data from 14 subjects (seven dyads) were collected. The synchronous speech paradigm was used in this experiment as well. The recording procedure and prosodic analysis were the same as described in Experiment 1. Pause duration was measured from the end of voicing for the nasal stop closure of the pre-boundary phrase (pause onset) to the beginning of voicing or frication for the phrase-initial segment (pause end), depending on what the first post-boundary segment was (the post-boundary phrase started with one of the following: [z, m, ɹ, s, l, n]). The duration of the pause was averaged between the two speakers of each dyad.

A two-factor ANOVA was performed on these data for each dyad separately testing the effect of the two factors: 1) length of the first post-boundary IP (with the two levels: short and long) and 2) length of the second post-boundary IP (with the two levels: short and long) and their interaction. If the results of the individual dyads patterned in the same manner, the dyads were pooled, as described in Experiment 1. Significance for all tests was set at $p<0.05$.

3.4.4 Results and discussion

For the pooled data, the results show that both the first and the second post-boundary phrase had an effect, such that longer phrases lead to longer pause duration (first post-boundary IP: $F(1,2009)= 9.043$, $p=.0027$; second post-boundary IP: $F(1,2009)= 10.521$, $p=.0012$), shown in Figure 7. Thus both the local and the more distant phrase had an effect on pause duration. The same pattern obtained for the individual dyads, though the effects did not reach significance.

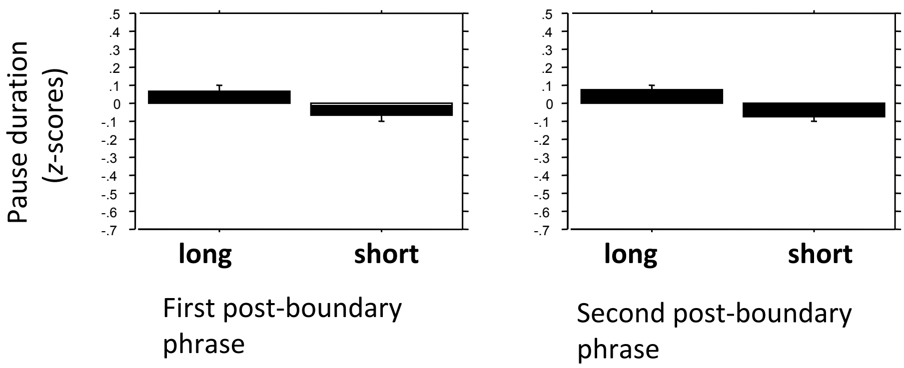

Figure 7: Effects of length of the first and second post-boundary IP on the duration of the pause (z-scores and standard errors) for all dyads pooled. Z-scores above 0 represent durations longer than the dyads' average pause duration, and z-scores below 0 durations shorter than the average duration.

The results indicate that speakers plan quite far ahead, since at speech onset they have planned, at least to some extent, the two post-boundary IPs. Note that we cannot say that in this case speakers select the hierarchically higher IP (rather than the hierarchically lower IP) and are thus planning the whole post-boundary phrase, as we suggested for the results in Experiment 2. If that were the case we would not expect to see the effects of the individual post-boundary phrases. Thus, the experiment provides further evidence that prosodic phrases determine the chunk to be planned by the speaker.

4 Discussion

The preceding studies examined the influence of phrase length and prosodic structure on pause duration in a variety of contexts. Phrase length was found to have an effect such that longer phrases lead to longer pauses. For prosodic structure it was found that depending on the length of the post-boundary phrase, prosodic structure impacts pause duration differently. For longer phrases, a complex structure leads to shorter pauses (Experiment 1), while for shorter phrases, a more complex structure leads to longer pauses (Experiment 2). No effect of IP branching was found in Experiment 3. Experiment 4 found both local and global effects of prosodic structure on pause duration.

The results of the experiments show that speakers have a fairly large lookahead in speech planning. At the time of the post-boundary phrase initiation, speakers were aware of the prosodic branching occurring at the third, fifth, seventh, and even 14th syllable after the boundary, and Experiment 4 provided evidence that speakers plan not only the first but also the second post-boundary phrase before speech onset. Note that the experimental paradigm used in these studies reduces the amount of planning needed for the sentences, in that reading and repetition most likely ease the planning load (see also Ferreira, 1991, 2007). Thus speakers in these experiments likely have a larger lookahead than speakers usually do. But importantly, the results show that speakers are capable of a large lookahead.

This large lookahead provides further evidence that speech production is not, in Ferreira and Swets (2002)' terminology, architecturally incremental. Speakers do not start speaking as soon as a minimal production unit is ready. Rather, speakers plan a large chunk–to a certain degree at least–before they start articulating.

This large lookahead is compatible with Keating and Shattuck-Hufnagel's (2002) "prosody-first", but not with Levelt's (1989) "prosody-last" approach. Thus we will assume Keating and Shattuck-Hufnagel's (2002) model, in which speakers, based largely on the syntactic structure, build a rough prosodic structure before phonological encoding. We suggest that prosodic structure then chunks the upcoming material into phrases that will be planned (phonologically and phonetically encoded) by the speaker before speech onset. The phonological length of the prosodic phrase (in number of syllables, or some other indicator of length), will then determine the time it will take the speaker to encode the upcoming phrase. The time needed will be reflected in the duration of the pause preceding the phrase. Any prosodic category (ip or IP) can be the chunk to be planned, and which category is selected will depend on the cognitive load of the upcoming material. For very long utterances a hierarchically lower category might be selected, thus chunking the utterance into smaller planning units, as was seen in Experiment 1. For shorter utterances a hierarchically higher category might be selected, since the utterance is already short enough for planning purposes, as in Experiment 2. In line with this suggestion, there was no effect of prosodic structure when the utterances were strings of numbers (Experiment 3). It was argued that these semantically and syntactically impoverished strings are cognitively demanding and that for that reason a hierarchically lower phrase (an ip) is selected for encoding. Since all utterances were branching into ip-s, there was no effect of prosodic branching. Also, as mentioned before, speakers under different circumstances might have a smaller scope of planning than the participants in these studies. Consequently, speakers might typically use hierarchically lower prosodic categories to determine the chunk to be processed (for example a lower IP, or an intermediate phrase).

5 Conclusion

The preceding studies have shown that phrase length has an effect on speech planning times (as instantiated in pause duration) both locally and globally, such that longer phrases lead to longer planning time. Speakers have been found to be able to plan their production at least two Intonation Phrases ahead, supporting the idea from Ferreira and Swets (2002) that speech production is not architecturally (strongly) incremental and supporting a prosody-first approach (as in Keating and Shattuck-Hufnagel, 2002). Prosodic complexity was found to have an effect on speech planning time, such that for longer phrases, prosodically complex structures lead to shorter pauses, while for shorter phrases, a complex structure leads to longer pauses. It was argued that prosodic phrasing serves to chunk the upcoming material into planning units, and that prosodic constituents of any category can serve this purpose. Whether a hierarchically higher or lower prosodic phrase is selected as a planning unit will depend on the length and difficulty of the upcoming material. The phonological length (in syllables) of that phrase will determine how long the time to encode it will be.

6 Acknowledgements

I am grateful to Sun-Ah Jun, Sungbok Lee, Jean-Roger Vergnaud, Stefanie Shattuck- Hufnagel, Jessica Hsieh, Kathy Cooke, Tine Mooshammer, and especially to Dani Byrd for their generous help with this study. I would like to thank two anonymous reviewers and Pascal Perrier, who provided very thoughtful and insightful reviews of the first draft of the manuscript. I would also like to thank the organizers of the Summer school "Cognitive and physical models of speech production, speech perception and the production-perception interaction" and the audience there for their feedback. I gratefully acknowledge the support of the NIH (Grant No. DC03172 to Dani Byrd) and the USC and Yale Phonetics Laboratory.

References

Beckman, M. E. (1989). Notes on intonation. In *Course reader of Ling 694B (Intonation), winter 1989, Department of Linguistics, Ohio State University*.

Beckman, M. E. and Elam, G. A. (1997). Guidelines for ToBI labelling. Version 3.0, unpublished ms. *Available online at: http://www.ling.ohio-state.edu/ tobi/ame_tobi/labelling_guide_v3.pdf*.

Beckmann, M. E. and Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3:255–309.

Bolinger, D. (1972). Accent is predictable (if you're a mind reader). *Language*, 48:633–644.

Butcher, A. (1981). Aspects of the speech pause: Phonetic correlates and communicative functions. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel*, 15.

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, 57:3–16.

Byrd, D. and Krivokapić, J. (2008). Prosodic variation: Understanding scope, categoricity, and recursion in speech production and perception. *Journal of the Acoustical Society of America*, 123:3423.

Byrd, D. and Saltzman, E. (1998). Intragestural dynamics of multiple phrasal boundaries. *Journal of Phonetics*, 26:173–199.

Carlson, K., Clifton, C., and Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, 45:58–81.

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a, i/ in English. *Journal of the Acoustical Society of America*, 117:3867–3878.

Choi, E. (2003). Pause length and speech rate as durational cues for prosody. *Journal of the Acoustical Society of America*, 114:2395.

Clifton, C., Carlson, K., and Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech*, 45:87–114.

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

Cummins, F. (2002). On synchronous speech. *Acoustic Research Letter Online*, 3:7–11.

Cummins, F. (2003). Practice and performance in speech produced synchronously. *Journal of Phonetics*, pages 139–148.

Cummins, F. (2004). Synchronization among speakers reduces macroscopic temporal variability. In *26th Annual Meeting of the Cognitive Science Society*, pages 304–309.

Cutler, A., Dahan, D., and Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40:141–201.

Dresher, E. B. (1994). The prosodic basis of the tiberian hebrew system of accents. *Language*, 70:1–52.

Edwards, J., Beckman, M. E., and Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America*, 89:369–382.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, 30:210–233.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100:233–253.

Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes*, 22:1151–1177.

Ferreira, F. and Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46:57–84.

Ferreira, V. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language*, 35:724–755.

Frazier, L., Carlson, K., and Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, 10:244–249.

Frazier, L., Clifton, C., and Carlson, K. (2004). Dont break, or do: Prosodic boundary preferences. *Lingua*, 114:3–27.

Gee, J. P. and Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15:411–458.

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

Griffin, Z. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin and Review*, 10:603–609.

Grosjean, F., Grosjean, L., and Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, 11:51–58.

Halliday, M. A. K. (1967). *Intonation and Grammar in British English.* Mouton de Gryuter.

Horne, M., Strangert, E., and Heldner, M. (1995). Prosodic boundary strength in Swedish: Final lengthening and silent interval duration. In *Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm*, pages 170–173.

Itô, J. and Mester, A. (1992). Weak layering and word binarity. Technical report, Linguistic Research Center, LRC-92-09, University of California, Santa Cruz.

Itô, J. and Mester, A. (2010). *Phonological Domains: Universals and Derivations.*, chapter The extended prosodic word, pages 135–194. Mouton de Gruyter.

Jun, S. A. (1993). *The phonetics and phonology of Korean prosody.* PhD thesis, PhD dissertation, The Ohio State University, Columbus, Ohio.

Jun, S. A. (2003). Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research*, 32:219–249.

Keating, P. and Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*, 101:112–156.

Klatt, D. (1975). Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics*, 3:129–140.

Krivokapić, J. (2007a). *The planning, production, and perception of prosodic structure.* PhD thesis, PhD dissertation, University of Southern California.

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

Krivokapić, J. (2010). Speech planning and phrase length. In *Proceedings of the 5th Conference on Speech Prosody, Chicago, paper number 100311*.

Ladd, D. R. (1988). Declination "reset" and the hierarchical organization of utterances. *Journal of the Acoustical Society of America*, 84:530–544.

Ladd, D. R. (1996). *Intonational Phonology.* Cambridge University Press.

Levelt, W. J. M. (1989). *Speaking. From Intention to Articulation.* MIT Press, Cambridge MA.

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Brain and Behavioral Sciences*, 22:1–38.

Martin, J. (1970). On judging pauses in spontaneous speech. *Journal of Verbal Learning and Verbal Behavior*, 9:75–78.

Meyer, A. S., Belkeb, E., Häckera, C., and Mortensen, L. (2007). Use of word length information in utterance planning. *Journal of Memory and Language*, 57:210–231.

Nespor, M. and Vogel, I. (1986). *Prosodic Phonology.* Foris Publications, Dordrecht, Riverton.

Oller, K. D. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, 54:1235–1247.

Petrone, C., Fuchs, S., and Krivokapić, J. (2011). Consequences of working memory differences and phrasal length on pause duration and fundamental frequency. In *Proceedings of the 9th International Seminar on Speech Production, Montréal, Canada*, pages 393–400.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation.* PhD thesis, MIT, Cambridge, Massachusetts.

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

Redi, L. and Shattuck-Hufnagel, S. (2001). Variation in realization of glottalization in normal speakers. *Journal of Phonetics*, 29:407–429.

Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb–participle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24:904–921.

Sanderman, A. and Collier, R. (1995). Prosodic phrasing at the sentence level. In Bell-Berti, F. and Lawrence, R., editors, *Producing Speech: Contemporary Issues. For Katherine Safford Harris*, pages 321–332. American Institute of Physics, New York.

Schafer, A. J. (1997). *Prosodic parsing: The role of prosody in sentence comprehension.* PhD thesis, PhD dissertation, University of Massachusetts, Amherst, Massachusetts.

Selkirk, E. (1984). *Phonology and Syntax: The Relation Between Sound and Structure.* MIT Press, Cambridge MA.

Selkirk, E. (1995). *Papers in Optimality Theory.*, chapter The prosodic structure of function words, pages 439–470. GLSA Publications.

Shattuck-Hufnagel, S. and Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25:193–247.

Shattuck-Hufnagel, S. and Turk, A. (1998). The domain of phrase-final lengthening in English. In *Proceedings of the 16th International Congress on Acoustics and 135th Meeting Acoustical Society of America*, pages 1235–1236.

Smith, M. and Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, 73:205–246.

Sternberg, S.and Monsell, S., Knoll, R. L., and Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In Stelmach, G., editor, *Information Processing in Motor Control and Learning*, pages 117–152. Academic Press, New York.

Strangert, E. (1991). Pausing in texts read aloud. In *Proceedings of the XIIth International Congress of Phonetic Sciences, Université de Provence, Aix-en-Provence, France*, volume 4, pages 238–241.

Strangert, E. (1997). Relating prosody to syntax: Boundary signaling in Swedish. In Kokkinakis, G., Fakotakis, N., and Dermatas, E., editors, *Proceedings of the 5th European Conference on Speech Communication and Technology*, pages 239–242.

Swets, B., Desmet, T., Hambrick, D. Z., and Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experiment Psychology: General*, 136:64–81.

Terken, J. and Collier, R. (1992). Syntactic influences on prosody. In Vatikiotis-Bateson, E., Sagisaka, Y., and Tohkura, Y., editors, *Speech Perception, Production and Linguistic Structure*, pages 427–438. IOS Press, Amsterdam, Washington, Oxford, Tokyo, Osaka, Kyoto.

Wagner, M. (2005). *Prosody and recursion.* PhD thesis, PhD dissertation, Massachusetts Institute of Techonology.

Wagner, V., Jescheniak, J. D., and Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36:423–440.

Watson, D. and Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19:713–755.

Wheeldon, L. and Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37:356–381.

Wightman, C. W., Stefanie, S. H., Ostendorf, M., and Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91:1707–1717.

Zvonik, E. and Cummins, F. (2002). Pause duration and variability in read texts. In *Proceedings ICSLP02, Seventh International Conference on Spoken Language Processing*.

Zvonik, E. and Cummins, F. (2003). The effect of surrounding phrase lengths on pause duration. In *Proceedings of Eurospeech, Geneva, Switzerland*, pages 777–780.