Speech Planning and Prosodic Phrase Length

Jelena Krivokapic

Department of Linguistics, Yale University, USA

jelena.krivokapic@yale.edu

Abstract

A synchronous speech study investigates effects on pause duration of prosodic phrases of different length. The goal is to examine local and distant effects of prosodic phrase length on pause duration. Subjects read 24 English sentences varying along the parameters: a) length in syllables (long or short) of the intonation phrase immediately following a target pause and b) length in syllables (long or short) of the second, more distant, intonation phrase following the pause. We find both local and global effects of phrase length on pause duration, indicating that speakers have a large lookahead in speech production, mediated by prosodic structure.

1. Introduction

The occurrence and strength of prosodic boundaries, as manifested in pause duration, is systematically influenced by several factors, such as discourse structure [1-3], syntactic structure [4-12], prosodic structure [13-17], phrase length [8, 16-22], and speech rate [20, 23-26], to name the most prominent. Individual differences between speakers also play a large, though poorly understood, role in determining prosodic boundaries.

The syntactic, prosodic, and phrase length effects have mainly been examined for phrases immediately preceding and following a boundary. Greater complexity and length of these phrases have been found to lead to longer pause duration [4, 7-22], though there seem to be interactions between length and syntactic and prosodic structure [8, 16]. The observed post-boundary effects have been attributed to speech planning. It is thought that as the structural complexity of the upcoming, post-boundary phrase increases, the speaker needs more time to plan this more complex structure. As a consequence, the pause duration, during which speakers plan the upcoming phrase, increases [8, 18, 12, 4]. The same principle applies to phonological length, where longer upcoming phrases lead to longer pauses [8, 16-22].

While pause duration effects of phrases immediately adjacent to the boundary are well recognized, the effects of phrases at a distance have not been examined in detail, despite indications that phrases further away from a specific boundary might influence the production and perception of that boundary [27-31]. Frazier, Clifton and Carlson [31] 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. Similarly, research has shown that listeners' interpretation of boundary strength depends on the boundary strength of surrounding boundaries [27-30]. These studies indicate that speakers' production and listeners' interpretation are guided by both local and global properties of prosodic structure.

Studies investigating speech planning have also found evidence of distant effects. Studies at the word and phrasal level indicate that speakers plan more than one structural unit at a time [32, 16, 17]. Griffin [32] finds that in two-word sequences, speakers adjust the time of speech onset to allow enough time for the two word sequence to be produced fluently, taking into account both the first and the second word in the upcoming two word sequence. Krivokapic [16, 17] finds that speakers have a lookahead of at least one Intonation Phrase. Such effects, observed in phrasal planning, might well be the cause of the above mentioned effects between distant boundaries [27-31].

Given distant effects of prosodic structure and the fact that speakers plan more than one structural unit at a time, the question arises as to how different properties of phrases at a distance will affect a given boundary, and what this tells us about the architecture of speech processing.

The goal of the present study is to examine local and distant effects of prosodic phrase length on prosodic boundary strength (as instantiated in pause duration). An experiment is presented that investigates the effects on pause duration of post-boundary prosodic phrases (Intonation Phrases) of different length, both immediately at the target boundary and further away from it. The larger motivation of the study is to illuminate the speech planning processes, in particular to understand the incrementality in the production of prosodic structure (i.e., how far ahead speakers plan an utterance) and the role of prosodic structure in the planning process.

2. Methods

2.1. Stimuli

After the target boundary, all the sentences consisted of two Intonation Phrases (IP), varying along the following parameters: a) post-boundary length in syllables (short or long) of the intonation phrase immediately following a target pause and b) post-boundary length (short or long) in syllables of the second, more distant, intonation phrase following the pause. This yields 4 conditions (2 first IP x 2 second IP). There were six sentences for each condition (24 sentences), and twelve repetitions of each sentence, yielding a total of 288 sentences. The sentences were randomized in blocks of 24. Note that it was not possible to systematically vary phrase length while keeping overall post-boundary length constant. In order to keep the overall length of the post-boundary phrase partially constant, the duration of the second phrase varied, depending on whether the first phrase was 2 or 4 syllables. Thus if the first phrase was 2 syllables, the short second phrase was 10 syllables, and the long second phrase was 16 syllables. If the first phrase was 4 syllables, the short second phrase was 8 syllables, and the long second phrase was 14. The post-boundary conditions are given in Table 1. Since the focus was on planning effects, the pre-boundary sentence was always the same, to eliminate pre-boundary effects ("*Bob was buying books for Sam*").

Table 1. *The four experimental conditions. '#' marks the target boundary. The comma separates the two IPs (as verified in the ToBI transcription).*

Post-boundary
Short first phrase, short second phrase (2 + 10 syllables)
Zack sang, claiming that this would help him choose the books.
Mike slept, although he knew Bob wanted him to help.
Rob came, but did not want to help him with the task.
Sadly, Tom joined him for the shopping adventure.
Later, he would be selling books from a large shelf.
Nice books, he was murmuring to himself sadly.
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.
Mike slept, although he knew that Bob wanted him to help with choosing the books.
Rob came, but did not want to help with the tricky task of buying the books.
Sadly, Tom joined him for the expensive shopping adventures in the mall.
Later, he would be selling yellowish books from a large shelf in the store.
Nice books, he was quietly murmuring to himself the entire day.
Long first phrase, short second phrase (4 + 8 syllables)
Zack sang loudly, claiming that this would be helpful.
Mike slept soundly, although he knew Bob needed him.
Rob came with him, but did not want to help at all.8.
Sadly enough, Tom joined him for the adventure.
Later that day, he would be selling reddish books.
Nice books for kids, he was murmuring to himself.
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.
Mike slept soundly, although he knew that Bob wanted him to help with the books.
Rob came with him, but did not want to help with the task of buying the books.
Sadly enough, Tom joined him for the exciting adventure in the mall.
Later today, he would be selling red books from a large shelf in the store.
Nice books for kids, he was sadly murmuring to himself throughout the day.

Based on earlier phrase length studies, we predict that pause duration will vary with first IP length. Further, based on the word level effects observed in Griffin [32], we predict that pause duration will vary with second IP length. In both cases, long phrases are expected to lead to a longer pause than short phrases, as more syllables need to be processed.

2.2. The synchronous speech method

To reduce variability in production, the experiment was conducted using the synchronous reading paradigm [33-36, 21]. In this paradigm, two speakers (a dyad) are seated facing each other, and they read sentences together, simultaneously, starting at the prompt of the experimenter. This method has been found to reduce variability in pause placement and pause duration without introducing artificial temporal properties into speech [33-36, 21]. In a study examining changes in proportional durations between synchronous and solo speech, Cummins [35] shows that the relative duration of phrases, boundaries, syllable onsets, stressed and unstressed syllables and stressed vowels does not change using this paradigm relative to solo read speech, but the variability is reduced for boundary duration and phrase duration in the synchronous condition. Zvonik and Cummins [21] further find that in both solo and synchronous speech speakers pause longest following the longest phrase. Together, these studies indicate that synchronous speech does not change fundamental properties of speech timing. The distinct advantage of the synchronous speech paradigm is that the reduction in variability in pause placement and duration facilitates comparisons across speakers and examination of effects of prosodic structure, as these will be less obscured by individual variation.

2.3. Subjects and Recordings

Fourteen subjects (7 dyads) read the sentences. The subjects were undergraduates. They were naïve as to the purpose of the study, and they were paid for their participation. The two subjects of a dyad were seated facing each other. Before the recording, they familiarized themselves with the sentences. Once familiar with them, the subjects were asked to read the sentences 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 sentence again. Errors were rare.

Subjects were recorded on a DAT recorder, using two Shure head mounted unidirectional microphones. The recordings were made at a 44,100 Hz sampling rate. These recordings were transferred to a PC onto the right and left channels of a stereo file at a sampling rate of 22,050Hz.

2.4. Measurements

In order to verify that the intended prosody was used by the subjects—that is, to verify whether they produced the two Intonation Phrases intended to be elicited by the stimuli—the author examined the recordings using the ToBI conventions for prosodic transcription [37]. Post-boundary phrases were examined for whether they contained two Intonation Phrases, and the pre-boundary phrase was examined to make sure it was produced as one Intonation Phrase. Intonation Phrase boundaries were identified by phrase accent, final lengthening, and a boundary tone. All the sentences were produced as intended. Of the 4032 sentences produced by the 14 speakers, 6 were excluded due to experimental error.

Pause duration was measured from the end of voicing for the nasal stop closure (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.

In a large number of data points (52%), subjects were breathing at the target boundary. These sentences were kept in the analysis, following Grosjean & Collins [38] finding that breathing and non-breathing pauses show the same pattern. In addition, separate statistical analyses of the data showed that excluding the breathing data from the analysis or excluding the non-breathing data from the analysis did not change the overall pattern of results (see below for more details).

2.5. Statistical Analysis

For each sentence token, the dependent variable of pause duration was the average pause duration of the two speakers of each dyad. In order to pool the data across dyads, the averaged pause durations were converted to z-scores (calculated for each dyad separately). A two-factor ANOVA was performed on these data for each dyad separately and for the pooled data, testing the main effect and interaction 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). Significance was set at $p<0.05$.

3. Results

The results for the individual dyads were not significant, but the differences in means all go in the same direction, and the z-score durations for all dyads were pooled for further analysis. The results show an effect of both the first and the second post-boundary phrase (first post-boundary IP: F (1, 4022)=12.069, p= .0005; second post-boundary IP: F (1, 4022)=13.117, p=.0003). As predicted, in each case, longer phrases led to longer pauses. The results are shown in Figure 1. There was also a significant interaction (F (1, 4022)=7.222, p= .0072) showing that when both the first and the second post-boundary phrase were long, this lead to larger increase in pause duration than in other cases (Figure 2). Table 2 shows the pause duration means (in z-scores) for the first and second post-boundary phrases. Note that a lower *z*-score means shorter pause duration.

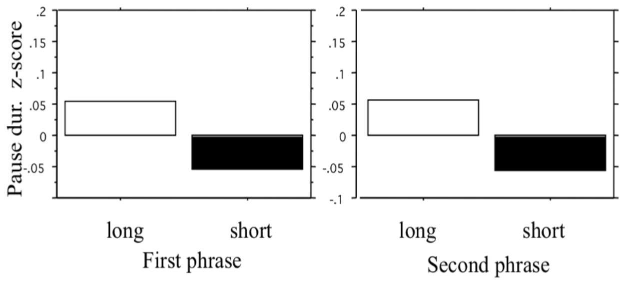

Figure 1: *Main effects of phrase length. A lower z-score means shorter pause duration.*

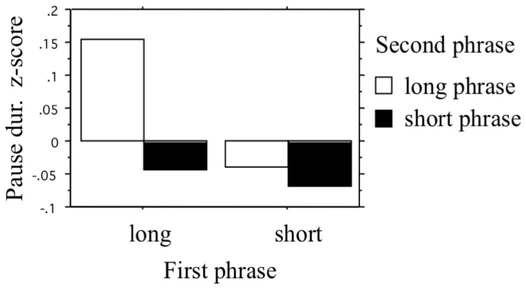

Figure 2: *Interaction of first and second phrase effects. A lower z-score means shorter pause duration.*

Similar results were obtained in the no-breathing and in the breathing subset of the data, which are reported here for completeness only. For the no-breathing subset, the pooled results show, as in the complete data set, for both the first and the second post-boundary phrase a significant effect, such that longer phrases lead to longer pre-boundary pauses (first post-boundary IP: F (1, 1918)=6.677, p= .0098; second post-boundary IP: F (1, 1918)=11.493, p= .0007) and no interaction effects. For the breathing only data, there was a significant interaction effect only (F (1, 2100)=5.240 p= .0222), with the same direction of the interaction effect as for the complete data set. The means for the first and second post-boundary phrase for the breathing subset, the non-breathing subset, and the complete data show overall the same pattern (see Table 2).

Table 2. *Pause duration (means, in z-scores) for the different phrase length factors. IP1 is the first post-boundary IP, IP2 the second post-boundary IP.*

	All data	No-breathing	breathing
IP 1 long	.054	-.122	.204
IP 1 short	-.055	-.239	-.127
IP 2 long	.057	-.1	.197
IP 2 short	-.057	-.263	-.136

4. Discussion

The results show that both the first and the second post-boundary phrase have an effect on pause duration. The longer first and the longer second phrase lead to longer pauses. Thus both local and more distant prosodic phrases have an effect on pause duration.

In terms of speech planning, the results of this study lend support to the idea that speech is planned quite far ahead. Speakers plan to some extent not just the first, but also the second phrase, adjusting speech onset to the planning needs of the first and second phrase in an utterance. Note that the findings do not mean that speakers, before they start articulating, have planned the full upcoming phrase. It does however mean that they have at least to a certain extent planned the two Intonation Phrases, probably only roughly (see [39] for a prosodic planning model that allows for such a process). The full phrases then presumably are planned out in more detail as the speaker proceeds with the utterance. Exactly how much a speaker plans ahead is likely to some extent speaker-dependent (see [40]). The results of the breathing subset of the data also might be interpreted in this light: there, the only effect observed was the interaction effect. Possibly, it is only when both the first and the second phrase were long, that the length effects combine to become noticeable—in other cases the planning of the post-boundary phrase could be accommodated during the long breathing time.

In the literature it is generally agreed that speech planning is incremental, meaning that different levels of speech processing occur at the same time (e.g., a speaker is at the same time planning speech and producing speech, rather than that the speaker first plans all aspect of speech and then produces the sentence). It is less clear however how long the planned stretch of speech is. Wheeldon and Lahiri [41] and Levelt and colleagues [42, 43] argue that the basic unit of planning is one phonological word at a time, with the possibility of planning longer stretches of time under certain circumstances. On the other hand, based on their findings, Griffin [32] and Ferreira & Swets [44] argue for a larger lookahead. Ferreira and Swets [44] argue that speech does not have to be strictly incremental, i.e., that incrementality is a strategy, not an architectural part of the processing mechanism. The presented findings lend support to the latter theories. Speakers can plan longer stretches of speech (two IPs in this case), and do not have to plan only the minimal amount.

Finally, while this study has not manipulated prosodic phrasing, the fact that the length of individual prosodic phrases has an effect on pause duration provides some evidence that prosodic structure is a relevant factor in speech planning processes (see also [13, 14] for such arguments), compatible with the notion of prosodic phrases as planning units [16, 17].

5. Conclusions

The effect of immediately adjacent prosodic phrases and phrases further away from the boundary was examined. Prosodic phrase length has an effect on boundary strength (as instantiated in pause duration) both locally and globally. Speakers plan their production at least two IPs ahead, supporting the idea that speech production is not architecturally (strongly) incremental, but can be planned quite far ahead. The results of the study offer further support for the role of prosodic structure in speech planning.

6. Acknowledgements

I am grateful to Dani Byrd and Jessica Hsieh for their help.

7. References

[1] Bannert, R., Botinis, A., Gawronska, B., Katsika, A. & E. Sandblom. 2003. Discourse structure and prosodic correlates. In *Proceedings of the XVth International Congress of Phonetic Sciences*. *Barcelona, Spain.* pp. 1229-1232.

[2] Smith, C. 2004. Topic transitions and durational prosody in reading aloud: production and modeling. *Speech Communication*, 42, 247-270.

[3] Swerts, M. & R. Geluykens. 1994. Prosody as a marker of information flow in spoken discourse. *Language and Speech*, 37, 21-43.

[4] Cooper, W. E. & J. Paccia-Cooper. 1980. *Syntax and Speech.* Cambridge, MA: Harvard University Press.

[5] Selkirk, E. 1981. On prosodic structure and its relation to syntactic structure. In: T. Fretheim (Ed.), *Nordic Prosody II*, pp. 111-140. Trondheim: Tapir.

[6] Nespor, M. & I. Vogel. 1986. *Prosodic Phonology.* Dordrecht, Holland/Riverton, USA: Foris Publications.

[7] Ferreira, F. 1988. Planning and timing in sentence production: The syntax-to-phonology conversion. Ph.D. dissertation, University of Massachusetts.

[8] Ferreira, F. 1991. Effects of length and syntactic complexity on initiation times for prepared utterances. *J. of Memory and Language*, 30, 210-233.

[9] Strangert, E. 1991. Pausing in texts read aloud. In: *Proceedings of the XIIth International Congress of Phonetic Sciences*, Aix-en-Provence, France, Vol 4, pp. 238-241.

[10] Terken, J. & R. Collier. 1992. Syntactic influences on prosody. In: Y. Tokhura, E. Vatikiotis-Bateson, & Y. Sagisaki.(Eds.), *Speech Perception, Production and Linguistic structure,* pp. 427-438. Amsterdam, Washington, Oxford: IOS Press and Tokyo, Osaka, Kyoto: Ohmsha,.

[11] Strangert, E. 1997. Relating prosody to syntax: Boundary signaling in Swedish. In: *Proceedings of the 5th European Conference on Speech Communication and Technology.* pp. 239-242.

[12] Watson, D. & E. Gibson. 2004. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19, 6, 713-755.

[13] Gee, J. P. & F. Grosjean. 1983. Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411-458.

[14] Ferreira, F. 1993. Creation of prosody during sentence production. *Psychological Review*, 100, 233-253.

[15] Horne, M., Strangert, E. & M. Heldner. 1995. Prosodic boundary strength in Swedish: Final lengthening and silent interval duration. In: *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm, Sweden, 1995. Vol. 1, pp. 170-173.

[16] Krivokapic, J. 2007a. Prosodic planning: Effects of phrasal length and complexity on pause duration. *J. Phonetics,* 35, 162-179.

[17] Krivokapic, J. 2007b. *The Planning, Production, and Perception of Prosodic Structure.* Ph.D. dissertation, University of Southern California.

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

[19] Whalen, D. H. & J. M. Kinsella-Shaw. 1997. Exploring the relationship of inspiration duration to utterance duration. *Phonetica*, 54, 138-152.

[20] Jun, S.-A. 2003a. The effect of phrase length and speech rate on prosodic phrasing. In: *Proceedings of the XVth International Congress of Phonetic Sciences*, Barcelona, Spain, pp. 483-486.

[21] Zvonik, E. & F. Cummins. 2003. The effect of surrounding phrase lengths on pause duration. In: *Proceedings of Eurospeech 2003*, Geneva, pp. 777-780. Geneva, Switzerland.

[22] Kentner, G. 2007. Length, ordering preference and intonational phrasing: Evidence from pauses. *Proceedings of Interspeech 2007*, Antwerp.

[23] Goldman Eisler, F. 1968. *Psycholinguistics. Experiments in Spontaneous Speech*. London and New York: Academic Press.

[24] Butcher, A. 1981. *Aspects of the Speech Pause: Phonetic correlates and communicative functions*. Kiel: Arbeitsberichte, 15. Institut fur Phonetik.

[25] Fletcher, J. 1987. Some micro and macro effects of tempo change on timing in French. *Linguistics*, 25, 951-967.

[26] Trouvain, J. & M. Grice. 1999. The effect of tempo on prosodic structure. In: *Proceedings of the XIVth International Congress of Phonetic Sciences*, San Francisco, California, August 1-7, 1999, Vol 2, pp. 1067-1070. Martinez: East Bay Institute for Research and Education, Inc.

[27] Schafer, A. J. 1997. *Prosodic Parsing: The role of Prosody in Sentence Comprehension*. Ph.D. dissertation. University of Massachussets, Amherst.

[28] Carlson, K., Clifton Jr., C. & L. Frazier. 2001. Prosodic boundaries in adjunct attachment. *J. of Memory and Language*, 45, 58-81.

[29] Clifton Jr., C., Carlson, K. & L. Frazier. 2002. Informative prosodic boundaries. *Language and Speech*, 45, 87-114.

[30] Jun, S.-A. 2003b. Prosodic phrasing and attachment preferences. *J. of Psycholinguistic Research*, 32, 219-249.

[31] Frazier, L., Clifton Jr., C. & K. Carlson. 2004. Don't break, or do: Prosodic boundary preferences. *Lingua*, 114, 3-27.

[32] 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.

[33] Cummins, F. 2002. On synchronous speech. *Acoustic Research Letter Online*, 3, 7-11.

[34] Cummins, F. 2003. Practice and performance in speech produced synchronously. *J. Phonetics*, 31, 139-148.

[35] Cummins, F. 2004. Synchronization among speakers reduces macroscopic temporal variability. *26th Annual Meeting of the Cognitive Science Society*, 304-309.

[36] Zvonik, E., & F. Cummins. 2002. Pause duration and variability in read texts. In: *Proceedings of the 2002 International Conference on Spoken Language Processing.* Denver, Colorado, pp. 1109-1112.

[37] Beckman, M. & G. Elam. 1997. Guidelines for ToBI labeling. (version 3.0). Unpub. ms. (online at: http://www.ling.ohio-state.edu/tobi/ame_tobi/labelling_guide_v3.pdf).

[38] Grosjean, F. & M. Collins. 1979: Breathing, pausing and reading. *Phonetica*, 36, 98-114.

[39] Keating, P. & S. Shattuck-Hufnagel. 2002. A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*, 101, 112-156.

[40] Swets, B., Desmet, T., Hambrick, D. Z. & F. Ferreira. 2007. The role of working memory in syntactic ambiguity resolution: A psychometric approach. *J. Exper. Psychology: General*, 64-81.

[41] Wheeldon, L. & A. Lahiri. 1997. Prosodic units in speech production. *J. Memory and Language*, 37, 356-381.

[42] Levelt, W. J. M. 1989. *Speaking. From Intention to Articulation.* Cambridge, MA: MIT Press.

[43] Levelt, W. J. M., Roelofs, A. & A. S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38.

[44] Ferreira, F. & B. Swets. 2002. How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *J. of Memory and Language*, 46, 57-84.