



Speech rate and syntactically conditioned influences on prosodic boundaries

Seung-Eun Kim, Sam Tilsen

Department of Linguistics, Cornell University, Ithaca, USA

sk2996@cornell.edu, tilsen@cornell.edu

Abstract

Previous studies of the interaction between syntactic structure and prosodic organization have not resolved whether there is articulatory and acoustic evidence for categories of prosodic boundaries or phrase types. One of the major problems in interpreting past studies is that the effects of speech rate have not been thoroughly examined. In the current study, we used a visual analogue cue to elicit continuous variation in speech rate for the production of two types of relative clauses. We hypothesized that if syntactic structure is mapped to categorical differences in prosodic organization, then measures of articulator kinematics and acoustic durations at phrase boundaries should differ or should scale differently with rate. Articulatory and acoustic data were collected from four English speakers. Analyses of gestural timing and movement range revealed strong differences in the effects of rate at boundaries before vs. after the relative clauses. Interaction effects between relative clause type and rate were found for some speakers. For acoustic measurements, both effects of boundary and relative clause type were observed. These findings are important because they show that articulatory kinematic and acoustic variables are more sensitive to rate variation at some phrasal boundaries than others.

Index Terms: syntax-prosody interface, speech rate, articulation, prosodic boundary

1. Introduction

A number of researchers have argued that prosodic structure is closely related to syntactic structure (e.g., [1], [2], [3]). Specifically, it has been asserted that the boundaries of prosodic units are mapped to the left and right edges of syntactic constituents. However, there have been few systematic investigations of how phonetic measurements vary as a function of phrasal boundary and syntactic structure.

In this context, the current study conducted a production experiment on two types of English relative clauses, non-essential relative clauses (NERC, also “non-restrictive” relative clauses) and essential relative clauses (ERC, also “restrictive” relative clauses), examples of which are shown in (1).

- (1) *Non-essential* relative clause (NERC)
There is one Mr. Hodd. He knows Mr. Robb.
A Mr. Hodd, $/_{ip}$ who knows Mr. Robb, $/_{ip}$ often plays tennis.
- Essential* relative clause (ERC)
There are two Mr. Hodds. Only one knows Mr. Robb.
The Mr. Hodd $/_{pwd}$ who knows Mr. Robb $/_{ip}$ often plays tennis.

In the NERC example, the relative clause does not contribute to identifying Mr. Hodd, and is similar to a parenthetical. In the ERC, the relative clause identifies a particular Mr. Hodd from a set of Mr. Hodds and therefore has a stronger semantic relation with the argument. It is therefore sensible that the

prosodic phrase associated with the ERC is more tightly integrated with the preceding phrases. According to theories of the syntax-prosody interface, and in particular [4], these should have a different prosodic structure. The prosodic categories predicted by syntactic analyses are indicated in (1) as well. The boundary labels *IP*, *ip*, *pwd* represent Intonational phrase, Intermediate (or Phonological) phrase, and Prosodic word boundary respectively.

On the basis of native speaker intuitions, it is reasonable to suspect that these different types of relative clauses are associated with different prosodic boundary categories, but such intuitions are generally derived from a conscious effort to “perform” the contrast. Empirically we do not know whether phonetic measures associated with these boundaries are categorically different or vary continuously.

Effects of speech rate on prosodic phrasing and the phonetic realization of prosodic boundaries have been found in several languages (e.g., for English: [5]; for French: [6]; [7]; for Korean: [8]). One problem with previous studies is that they have varied speech in a categorical fashion, i.e. eliciting fast vs. normal vs. slow speech. This is problematic for several reasons. First, different speakers are likely to interpret instructions to speak “quickly” or “slowly” etc. in different ways; this creates additional variance in phonetic measures and makes it more difficult to conduct analyses across speakers. Second, imposing rate categories in a production task may encourage speakers to adopt categorically different behaviors which they otherwise might not. Lastly, it is possible that syntactic/prosodic effects on phonetic variables might be apparent at some rates but not others, or that the phonetic effects of syntactic/prosodic categories might be best analyzed as differences in how phonetic variables change as speech rate varies.

To avoid these problems, we elicited continuous variation in speech rate with a visual analogue cue that moved across a computer screen at different speeds. The cue did not impose a small set of rate categories; instead, it induces a wide range of rate variation in each speaker. Given this aspect of the design, we test the following hypotheses by assessing how articulatory kinematics and acoustic measures vary with speech rate.

Hypothesis 1. Positional boundary differences: prosodic boundaries before and after a relative clause are different. **Predictions:** kinematic and acoustic measures at prosodic boundaries before and after a relative clause will differ or will scale differently with rate.

Hypothesis 2. Syntactically conditioned prosodic organization: prosodic boundaries will differ depending upon the type of a relative clause. **Predictions:** kinematic and acoustic measures associated with NERCs, due to their lesser degree of syntactic/semantic integration, will be stronger than those associated with ERCs, or will scale more strongly with rate.

We note here regarding Hyp. 1 that pre-clause and post-clause boundaries can be interpreted as different for syntactic reasons but are also different simply because they occur ear-

lier or later in a sentence; hence we remain agnostic regarding whether observed differences between pre- and post-clause boundaries are syntactic in nature. Also note that we refer to an abstract conception of boundary “strength” under the assumption that longer durations and larger movement ranges are associated with stronger boundaries.

2. Methods

2.1. Participants and task

Four native speakers of English – coded below as JR, CH (Male), and EJ, CJ (Female) – participated in the experiment. Articulatory and acoustic data were collected with an NDI Wave Electromagnetic Articulograph (EMA).

There were six blocks of 40 trials in each experimental session. Participants produced the same type of RC throughout a block, and blocks alternated between the two types of RCs. In each trial, two sentences appeared on the screen as in (1). The first sentence (plain font) provided context to the second sentence (bold font – target). Participants were instructed to read both sentences silently when they first appeared. After 1.5 seconds, the visual analogue rate-cue appeared. This cue was a red box that moved from left to right across the screen. The period of time it took for the rate-cue to move across the screen was varied in ten steps from 0.8 to 4.1s (this was based upon the ranges observed in pilot data). Participants were instructed to vary the speed of their production of the target sentence in a way that reflected variation in the speed of the visual analogue. Importantly, participants were required to wait until the cue had disappeared before initiating their production. Thus, the cue functioned to indirectly elicit variation in rate, without prescribing any specific timeframe for sentence production and without requiring participants to interpret categories such as “quickly”, “normally”, etc. The speed of the cue was varied randomly from trial to trial. Post hoc analyses of effective speech rate (measured as durations of produced sentences) showed that the rate manipulation strategy was successful in eliciting a wide range of variation in rate within each speaker. Overall, the utterance duration increased as the cue duration (0.8 to 4.1s) increased (mean coefficient: 0.35, mean R-squared: 0.745).

The target words in the experiment were the names that follow *Mr.*, and specifically, the targets of the articulatory analyses were the coda consonants in each target word. For acoustic analyses, the rime portion of the name and the duration of a following pause (if present) were measured. All the names had the same vowel /a/ and either a /b/ or /d/ coda (120 each per speaker). Note that in order to prevent the participants from focusing on the target words, they were explicitly instructed not to emphasize the names. All names started in /h/, /r/, or /l/ to avoid repeating consonants within the name.

2.2. Data collection and processing

Kinematic data were collected at 400 Hz. Articulator sensors were located midsagittally on the tongue tip and body (TT, TB), gum below the lower incisors (JAW), and upper/lower lips (UL, LL). Reference sensors for head movement correction were located on the nasion and left and right mastoid processes. Acoustic data were collected at 22050 Hz with a shotgun microphone located 1.5 m from the participant. A total of 240 trials were collected for each of the four participants. Out of the 960 trials, 82 trials were discarded due to speech errors, disfluencies, or problems in data collection, and this left 878 trials in total (91.5%). In order to locate boundaries in the acoustic signal, 6

trials were manually labeled for each speaker and used to train HMMs in the Kaldi speech recognition toolkit. A forced alignment was conducted for all remaining trials.

Kinematic variables were extracted as follows. First, after head movement correction, horizontal and vertical positions of articulator sensors were resampled to 1000 Hz. A lip aperture signal (LA) was then defined as the Euclidean distance between the UL and LL. Using the acoustic segmentation as a reference point, relevant velocity extrema were detected for consonantal constriction and release gestures at the ends of phrases. Gestural onsets and offsets were then identified in relation to the velocity extrema. For target words with an alveolar closure (e.g., *Hodd*, *Rodd*, etc.), the vertical position of the TT sensor was analyzed. For target words with a bilabial closure (e.g., *Robb*, *Lobb*, etc.), the LA signal was analyzed. The kinematic measures that were extracted for each trial and boundary along with temporal and spatial dependent measures are shown in Figure 1. Note that the onsets and targets of closure/release phases are defined as the points in time when the velocity signal rises above or falls below 20% of the maximum velocity. 35 trials (4%) with ambiguous kinematic landmarks were excluded from analyses.

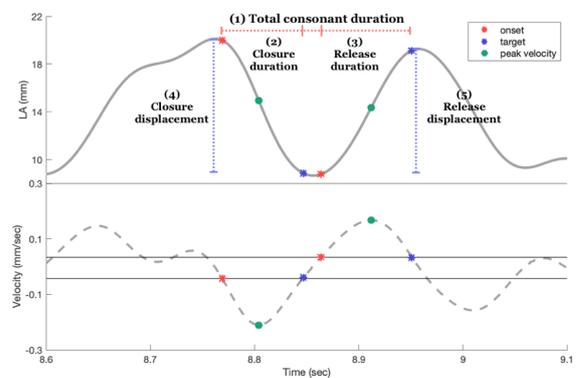


Figure 1: An example token showing LA trajectory (top panel) and the accompanying velocity trajectory (bottom panel) with markings of kinematic landmarks and dependent measures.

2.3. Data analysis

For each measure at each boundary, a mixed effects regression with speech rate as a fixed effect and speaker as a random intercept was conducted to identify outliers. Datapoints whose standardized residuals were > 2.807 (99.5%) were excluded from subsequent analyses. The remaining data were then analyzed with stepwise regressions using mixed effects models, with random intercepts for speaker. To test Hyp. 1, the full model included speech rate, boundary position, and their interaction. To test Hyp. 2, the full model included speech rate, relative clause type, and their interaction. The same procedure was applied to the data from each speaker in order to assess inter-speaker variation.

3. Results

In articulatory data, the results showed a strong effect of boundary position (pre-clause $<$ post-clause), but mixed evidence of syntactically conditioned differences. In acoustic data however, both positional as well as syntactic effects on boundaries were

observed.

3.1. Articulatory results

Regarding Hyp. 1, we found evidence that there were positional effects on boundaries for both relative clause types. Specifically, there were interaction effects between boundary position and speech rate for the total and release duration ($p < 0.001$). For closure duration, both clause types showed a significant main effect of rate ($p < 0.001$), but only NERCs exhibited a significant boundary effect ($p < 0.01$) (See Figure 2).

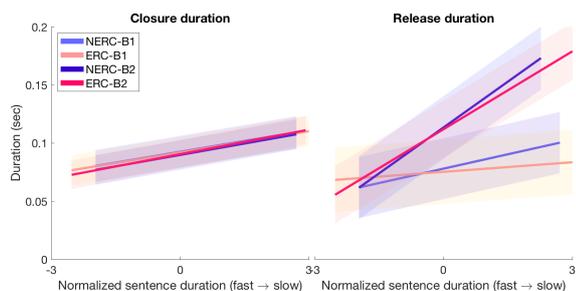


Figure 2: Temporal measurements (closure, release duration) of all speakers combined. The empirical rate (sentence duration) was normalized in order to compare the data across speakers. Lines represent the polynomial fitting of the data, and the bands indicate ± 1 standard deviation.

The by-speaker analyses showed similar results (See Figure 3): all speakers had a significant interaction between rate and boundary position in total and release duration (JR, CH, CJ, EJ (except release dur): $p < 0.001$; EJ (release) $p < 0.05$) in both types of sentences. In the closure duration, at NERC where the boundary effect was found, one speaker (JR) showed an interaction effect and two speakers (CH, EJ) showed a significant boundary effect (all p 's < 0.01). In ERCs, however, only one speaker (CH) showed a marginal effect of boundary. If the boundary effect exists, post-clausal boundary was produced more strongly than pre-clausal boundary.

The above analyses support the boundary position hypothesis: there was either boundary position-rate interaction effect, such that kinematic measures of the post-clause boundary change more strongly with rate than those associated with the pre-clause boundary, or the main effect of boundary, where post-clause had a longer duration than pre-clause boundary. The difference conditioned by the boundary position is particularly large in the release phase of the consonant. Note that boundary positions are generally not distinguished by temporal articulatory measures at fast rates.

With respect to Hyp. 2, we found mixed evidence that there are syntactically conditioned differences in boundaries, both before and after the relative clause. At the pre-clausal boundary, there was evidence for syntactically conditioned differences in some speakers, which supports Hyp. 2; but at the post-clausal boundary, we found that the effect directions sometimes differed by speaker. Across speakers, there was a significant interaction effect between rate and RC type on total duration at both boundaries and release duration at pre-clausal boundary (all p 's < 0.01); the release duration at post-clausal boundary showed a marginal interaction effect. In the closure duration, however, there was a significant main effect of rate ($p < 0.001$), but no effect of RC type or interaction.

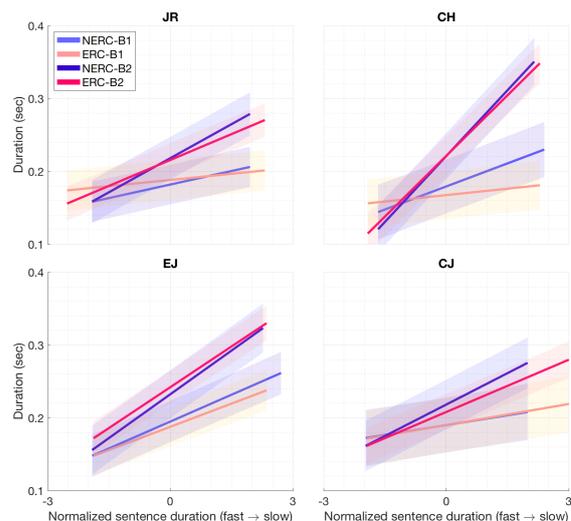


Figure 3: Temporal measurements (total duration) of each speaker.

The by-speaker analysis showed that at pre-clausal boundary, one speaker (CH) showed an interaction effect ($p < 0.01$) for both total and release duration; another speaker (EJ) showed a significant effect of RC type in the release duration ($p < 0.05$), but marginal effect in the total duration. In cases where there was an effect of RC type, NERC was produced more strongly than ERC. JR, however showed a marginal interaction effect in both total and release duration, and in his case, the duration of ERC had a stronger scaling effect with rate than that of NERC. The closure duration at pre-clausal boundary showed no effect of RC type or interaction in all speakers.

At post-clause, two speakers showed an effect of RC type. EJ showed a significant RC effect in all measurements (i.e., closure, release, and total duration), showing longer duration at ERC than at NERC ($p < 0.05$). On the other hand, CJ showed a significant effect of RC at the release ($p < 0.001$) and total duration ($p < 0.05$), and in her case, the NERC had a longer duration than ERC. Since the two speakers who showed an RC effect had a different pattern regarding which RC is produced more strongly, the RC effect was not found in post-clause in the combined analysis (See Figure 2).

Overall, this shows that the RC effect exists in two speakers at pre-clausal boundary (NERC $>$ ERC) especially at the release phase of the consonant. Two speakers showed an effect of RC at post-clause as well, but they did not show a uniform pattern. The results in post-clausal boundary suggest that speakers produce the same type of boundary in a different way.

Analyses on the spatial measurements were similar to the temporal measurements (See Figure 4). In both NERC and ERC, there was a significant interaction between speech rate and boundary position in the release displacement ($p < 0.001$), but only a significant main effect of rate ($p < 0.001$) was found in the closure displacement.

These results were supported in the by-speaker analysis; only one person showed a significant boundary effect in the closure phase specifically at NERC ($p < 0.05$). In the release phase, one speaker (JR at NERC $p < 0.05$ and CH at ERC $p < 0.001$) showed an interaction effect and three other speakers showed a main effect of boundary ($p < 0.001$). This shows that

Table 1: Summary of the results. ✓ represents the effect exists. ✗ shows that the effect does not exist in the majority of speakers, while ✕ shows that the effect exists in some speakers, but they have a different pattern.

		Rate effect	Position effect		Clause type effect	
			NERC	ERC	pre-	post-
Temporal	total	✓	✓	✓	2 speakers	✕
	closure	✓	✓	✗	✗	✗
	release	✓	✓	✓	2 speakers	✕
Spatial	closure	✓	✗	✗	✗	✗
	release	✓	✓	✓	2 speakers	✕
Acoustic	pause	✓	✓	✓	✓	✓
	rime	✓	✓	✓	✓	✗

the boundary effect exists only at the release phase, and if there is an effect, the post-clause was produced more strongly than the pre-clause boundary. Note that the release displacement of pre- and post-clause does not merge even at the fast rate, which is different from the temporal results.

Regarding rate and RC type, both boundary positions only showed a strong effect of rate at the closure phase ($p < 0.001$); in the release phase however, an interaction effect was found at pre-clause ($p < 0.01$) but only the main effect of rate was found at post-clause ($p < 0.001$).

In the closing phase, except CJ, all speakers showed only a main effect of rate (cf. one speaker showed an interaction effect) ($p < 0.05$), but no speakers showed an RC effect. In the release phase at pre-clause, CH showed an interaction between rate and RC ($p < 0.05$) and EJ showed a main effect of RC ($p < 0.001$), both of which showed a stronger effect in NERC than in ERC. At post-clause, EJ and CJ showed an effect of RC ($p < 0.05$), but again they showed a different pattern: EJ showed a larger displacement at ERC while CJ showed a larger displacement at NERC. For JR, a marginally significant interaction was found at pre-clause and a significant interaction was found at post-clause ($p < 0.05$). For the latter case, JR showed a stronger scaling effect at ERC compared to NERC.

The results on the spatial measurements show that boundary position as well as RC type do not affect the closure displacement, but in the release phase, boundary position causes a significant difference (pre- < post-), while the RC type causes difference in pre-clause but inter-speaker variation in post-clause.

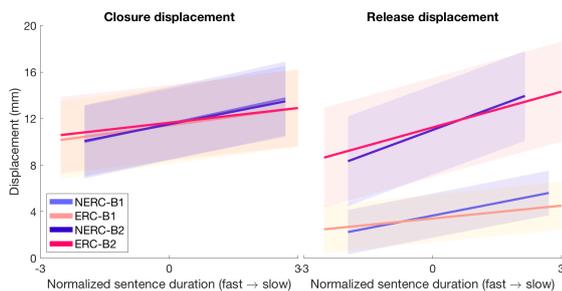


Figure 4: Spatial measurements (closure, release displacement) of all speakers combined.

3.2. Acoustic results

In acoustic analyses, we found evidence that there are positional as well as syntactically conditioned effects on boundaries, which support both Hyp. 1 and Hyp. 2. Analyses on the pause

duration showed a strong interaction effect between speech rate and boundary position at both relative clause types ($p < 0.001$), while the rime duration showed an interaction effect at NERC ($p < 0.05$) but a main effect of rate and boundary at ERC ($p < 0.001$). In both structures, post-clause scaled more strongly with rate or had a longer duration than pre-clause. Regarding RC type, the pause duration at pre-clause and post-clause and the rime duration at pre-clause showed a strong interaction effect, but the rime duration at post-clause only showed the effect of rate but not RC (all p 's < 0.001). In cases where there was an interaction effect, NERC scaled more strongly than ERC.

4. Discussion and conclusion

As summarized in Table 1, most of the measurements showed a strong effect of boundary: post-clausal boundary was produced more strongly than pre-clausal boundary. This result confirms Hyp. 1. However, it should be noted that the difference between pre- and post-clause disappears in fast rate in temporal articulatory measurements and acoustic data. The difference between the two boundaries was constant at all speech rates in spatial measurements.

On the contrary, Hyp. 2, which examines the effect of RC, had mixed support; the effect was found across speakers in the acoustic measurements but a mixed result was found in the articulatory data. In articulatory measures, two speakers were sensitive to RC type at pre-clause, exhibiting a stronger effect at NERC than at ERC, but at post-clause, there was either no effect of RC type or two speakers showed an effect, but they showed an opposite pattern. Interestingly, the RC type effect was found only at the release phase of the consonant.

One of the novel contributions of this study is that we elicited continuous variation in speech rate and examined how the measurements change within this continuum. By adopting this method, it is found that the speech rate is highly correlated with all the dependent measures: all the articulatory as well as acoustic measurements increased as the speech rate slowed down. Furthermore, each measurement showed a gradient change rather than a categorical (non-linear) change along the continuum.

In sum, we found that pre- and post-clause boundaries differed in strength. These differences may be interpreted as syntactically conditioned or attributable to their position in the sentence. Furthermore, there is substantial variability across speakers with respect to differences conditioned on relative clause type. These findings are important because they call into question any theoretical accounts which prescribe an invariant mapping of syntactic structure to prosodic organization. They also demonstrate the importance of fine-grained variation in speech rate to assess prosodic variation.

5. References

- [1] M. Nespov and I. Vogel, *Prosodic Phonology*. Dordrecht, Holland: Foris Publications, 1986.
- [2] E. Selkirk, "On derived domains in sentence phonology," *Phonology*, vol. 3, pp. 371–405, 1986.
- [3] H. Truckenbrodt, "On the relation between syntactic phrases and phonological phrases," *Linguistic inquiry*, vol. 30, no. 2, pp. 219–255, 1999.
- [4] E. Selkirk, "Comments on intonational phrasing in english," *Prosodies: With special reference to Iberian languages*, vol. 9, pp. 11–58, 2005.
- [5] D. R. Ladd, D. Faulkner, H. Faulkner, and A. Schepman, "Constant "segmental anchoring" of f 0 movements under changes in speech rate," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1543–1554, 1999.
- [6] C. Fougerson and S.-A. Jun, "Rate effects on french intonation: Prosodic organization and phonetic realization," *Journal of Phonetics*, vol. 26, no. 1, pp. 45–69, 1998.
- [7] A. Michelas and M. d'Imperio, "When syntax meets prosody: Tonal and duration variability in french accentual phrases," *Journal of Phonetics*, vol. 40, no. 6, pp. 816–829, 2012.
- [8] S.-A. Jun, "The effect of phrase length and speech rate on prosodic phrasing," in *proceedings of the XVth international congress of phonetic sciences*, 2003, pp. 483–486.