Journal of Phonetics 93 (2022) 101152

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

An investigation of functional relations between speech rate and phonetic variables

Seung-Eun Kim *, Sam Tilsen

Department of Linguistics, Cornell University, 203 Morrill Hall, Ithaca, NY 14853, USA

ARTICLE INFO

Article history:
Received 9 July 2021
Received in revised form 27 April 2022
Accepted 28 April 2022

Keywords:
Speech rate
Prosody
Functional relations
Non-linearity
Attenuation effects
Bayesian analysis
Syntactic context

ABSTRACT

It is well known that speech rate is correlated with many phonetic variables. The current study aims to obtain a more precise characterization of how phonetic measures covary with speech rate. Specifically, we assess whether there is evidence for linear and/or non-linear relations with rate, and how those relations may differ between phrase boundaries. Productions of English non-restrictive (NRRCs) and restrictive relative clauses (RRCs) were collected using a method in which variation in speech rate is cued by the speed of motion of a visual stimulus. Articulatory and acoustic variables associated with phrase boundaries were analyzed; for each variable, Bayesian regression was used to obtain posterior parameter distributions for a set of generalized linear models. Analyses of posterior predictions showed that phonetic variables associated with a phrase boundary that follows the relative clause (post-RC boundary) were more susceptible to rate variation than those at a boundary that precedes the relative clause (pre-RC boundary). Phonetic variables at the post-RC boundary also showed evidence for non-linear relations with rate, which suggest floor or ceiling attenuation effects at extreme rates. On the other hand, substantial differences between syntactic contexts were found primarily at the pre-RC boundary. A high degree of participant-specificity was observed in F0-related variables.

1. Introduction

Speech rate has been found to covary with most articulatory measures (i.e. gestural timing, speed, and magnitude) and acoustic measures (i.e. segmental durations and F0). However, most previous studies on speech rate have used qualitative instructions to elicit rate variation, with just a few rate categories (e.g. “speak slowly”, “speak normally”, “speak quickly”). Such instructions may elicit multimodal rate distributions which, although useful for establishing that rate effects exist, are not ideal for precisely characterizing those effects. Indeed, many analyses have treated speech rate as a categorical factor. Our goal in this paper is to draw more detailed inferences regarding how phonetic variables near phrase boundaries vary with rate; specifically, we are interested in (i) whether there is evidence for linear and/or non-linear relations between speech rate and various phonetic measures, and (ii) whether such relations differ between phrase boundaries and/or syntactic contexts.

To accomplish this, we used a recently developed rate elicitation method in which a moving visual stimulus signals participants to vary their speech rate from one experimental trial to another. The speed of the moving stimulus iconically represents rate of speech. Using this approach, we were able to elicit variation in rate without imposing qualitative categories such as “slow”, “normal”, and “fast”. Moreover, to obtain better inferences regarding how phonetic measures vary with rate, we used an empirical measure of speech rate which is based on utterance duration rather than stimulus parameters. The utterances that we elicited were sentences which included either a restrictive relative clause (RRC), e.g. *The Mr. Hodd who knows Mr. Robb often plays tennis*, or a non-restrictive relative clause (NRRC), e.g. *A Mr. Hodd, who knows Mr. Robb, often plays tennis.* These clause types are generally understood to differ both syntactically and semantically (see Arnold, 2007) and have been hypothesized to be associated with different prosodic organizations (Nespor & Vogel, 1986; Selkirk, 2005). Articulatory and acoustic variables in the vicinity of phrase boundaries were examined; we specifically examined the phrase boundary that occurs before the RC (i.e. between *Mr. Hodd* and *who*, henceforth *pre-RC boundary* or

* Corresponding author.
E-mail address: sk2996@cornell.edu (S.-E. Kim).

https://doi.org/10.1016/j.wocn.2022.101152
0095-4470/

B1) and the boundary that follows the RC (i.e. between *Mr. Robb* and *often*, henceforth *post-RC boundary* or *B2*).

The main analyses were conducted using Bayesian statistical inference. Linear and non-linear models were fit on each variable at each boundary with speech rate as a continuous predictor; models were compared using an information-based criterion, and rate effects were interpreted using posterior predictions. The analyses presented here are exploratory in that we intend to explore the functional relations between phonetic variables and speech rate, rather than testing specific hypotheses regarding whether certain variables have linear or non-linear relations. Our motivation for adopting this approach derives from taking to heart the recommendations of various contributions to a recent special issue of this Journal – *Emerging Data Analysis in Phonetic Sciences*, eds. Roettger, Winter, and Baayen (2019) – and specifically distinguishing between exploratory and confirmatory hypothesis testing as described by Nicenboim et al. (2018). We further examined whether the functional relations between speech rate and phonetic variables differ by phrase boundaries that vary in strength and/or by syntactic structure (i.e. by RC type). The main findings are (i) that some phonetic variables exhibit non-linear relations with speech rate, and (ii) that these non-linear relations are more likely to arise in the phonetic measures at stronger prosodic boundaries. However, phonetic measures at weaker prosodic boundaries show more substantial syntactically conditioned differences than ones at stronger prosodic boundaries.

The paper is structured as follows. Section 2 provides the background of the current study; Section 2.1 presents findings from previous studies on the correlation between speech rate and phonetic variables; Section 2.2 discusses elicitation and analysis methods of speech rate and describes related studies; Section 2.3 outlines the current experiment, with a special focus on the properties of the stimuli employed in our experiment. We also present our perspectives on prosodic structure and prosodic boundary strength in Section 2.3. Section 3 details experiment design and analyses methods. Section 4 presents the results, and Section 5 provides further discussion. Section 6 summarizes our main findings along with some contributions and future directions. It should be noted that we use the term “speech rate” in this paper to refer to “speaking rate”, which is a measure of rate that includes pauses, rather than “articulation rate”, which excludes pauses (cf. Kendall, 2013); thus, the terms “speech rate” and “speaking rate” are used interchangeably throughout the paper.

2. Background

2.1. Correlation between speech rate and phonetic variables

Many different sorts of phonetic measures are correlated with speech rate. First, previous studies have reported that articulation durations decrease at fast speech rate (e.g. Byrd & Tan, 1996; Engstrand, 1988; Gay, 1981; Pouplier et al., 2017). Although individual gestures may differ on the amount of shortening – for example, vocalic gestures are more likely to reduce than consonantal gestures (e.g. Engstrand, 1988; Gay, 1981), they are generally shortened at fast rates. Previous studies have also found evidence of an interaction between speech rate and coarticulation such that the overlap between successive gestures increases as speech rate is increased (e.g. Byrd & Tan, 1996; Engstrand, 1988; Gay, 1981; Krakow, 1993; Munhall & Löfqvist, 1992; Pouplier et al., 2017).

Second, speech rate has been found to correlate with articulatory movement amplitude and velocity. In general, movement amplitude is reduced at fast speech (e.g. Gay, 1981; Kent & Moll, 1972; Lindblom, 1964); however, the presence or the amount of reduction can differ by the manner or place of articulation of the gestures, the position of the gestures in the syllable (Byrd & Tan, 1996), or by individual participants (Ostry & Munhall, 1985). Regarding movement velocity, studies have found mixed results. For example, although both Gay (1981) and Kent and Moll (1972) observed reduction in the movement amplitude of the jaw at fast rates, movement velocity increased with rate in Gay (1981) but it remained constant or rather decreased for some participants in Kent and Moll (1972). On the other hand, Abbs (1973) found an increase in peak velocity at fast rates, while the amplitude remained unchanged. Some studies (e.g. Kuehn & Moll, 1976; Ostry & Munhall, 1985) found no common pattern across participants as the direction of the changes in velocity varied across individuals. These inconsistent results have led some researchers to suggest an alternative measure such as velocity profile (i.e. the shape of the velocity–time function) to investigate the correlation between speech rate and velocity more accurately (Adams et al., 1993).

An association between speech rate and suprasegmental features has also been identified. At slower speech, individuals tend to insert more pauses, and the durations of pauses or phrase-final elements are lengthened (e.g. Berkovits, 1991; Goldman-Eisler, 1968; Grosjean, 1979; Lane & Grosjean, 1973; Michelas & D’Imperio, 2012; Trouvain & Grice, 1999). Moreover, speech rate has been found to correlate with pitch level, local and global pitch range, and velocity of pitch movements. For example, Fougeron and Jun (1998) studied F0 contours of a narrative passage in French and found that their participants compress their local pitch range by adjusting the pitch level of the peak or valley at fast rates. The global pitch range, which was calculated with the highest and lowest F0 values in the narrative passage, was also compressed as speech rate is increased. However, they did not find a significant correlation between rate and the average velocity of pitch movements. Conversely, Caspers and van Heuven (1993) and Ladd et al. (1999) in their studies of Dutch and English respectively showed that the slope of the local F0 movement (i.e. velocity) becomes steeper at fast rates, while the excursion size remains unchanged. These studies all had in common that they found inter-participant differences on whether participants control F0 maxima and/or minima and how much they increase or decrease F0 according to variation in speech rate. Lastly, prosodic phrasing also changes with rate; specifically, the strength of a given prosodic boundary is lowered at fast rates (e.g. Caspers & van Heuven, 1991; Fougeron & Jun, 1998; Jun, 1993).

In sum, as speech rate is increased, gestures are shortened, co-articulated, and reduced in magnitude, and likewise, acoustic durations become shorter. F0 movements tend to be less extreme at fast rates, yet individual differences are observed in the control of F0 according to changes in rate. The causes of these correlations may be viewed in two different ways. On one hand, the correlations may be due to a cognitive mechanism that controls speaking rate, which leads to variation in articulatory timing and target parameters. Alternatively, it could be that no global control mechanism exists, and rather, differences in measures of rate emerge from local variation in the control of articulatory timing and targets. In other words, it is possible that there exists a global rate control parameter, or that no such parameter exists and that speakers change rate via localized articulatory adjustments. Although we believe it is important to explore the direction of causation between rate and phonetic variables, it is beyond the scope of the current study, and we remain agnostic on the mechanisms behind the correlations.

Although many previous studies have usefully demonstrated relations between speech rate and various phonetic measures, our understanding of the form of these relations is limited. In particular, do phonetic measures vary linearly or non-linearly as a function of speech rate? The present paper thus aims to obtain a more precise characterization of the functional relations between phonetic variables and speech rate. Specifically, we examine how articulatory and acoustic measures associated with phrase boundaries vary with rate.

Investigating functional relations is important for a number of reasons. First, it allows us to identify potential constraints on articulatory control. Previous studies have proposed that there may be compressibility and/or expandability constraints on segment durations (e.g. Cooper et al., 1985; Klatt, 1973, 1976). For example, Cooper et al. (1985) suggested the presence of an expandability constraint, based on the finding that the duration of a sentence-final word which bears focus does not increase as much as is predicted by the sum of the independent effects of phrase position and focus. On the other hand, Klatt (1973) found that vowels become incompressible beyond a certain amount of shortening. By inspecting functional relations between phonetic measures and speaking rate, we can examine the presence of expandability and/or compressibility constraints on articulation. Furthermore, our study examines not only segmental durations but also articulatory and F0 variables in the vicinity of phrase boundaries; this could help determine whether such constraints only apply for durational measures, or they can be extended to other aspects of articulation in general.

Second, through the investigations of functional relations, we can empirically test assumptions that are made in the theories of prosodic structure or local rate modulations. Specifically, given that the phonetic variables that we examine are associated with phrase boundaries, their functional relations with rate may provide phonetic evidence on categorical prosodic boundary shifts or instead suggest gradient phonetic variation as a function of prosodic boundary strength. The functional relations between rate and phonetic measures can also be interpreted in relation to the theories that predict local slowing at phrase boundaries, such as π-gesture theory (Byrd & Saltzman, 2003) or attentional modulation in the selection-coordination framework (Tilsen, 2018).

2.2. Elicitation and analysis of speech rate

To investigate functional relations between phonetic measures and speech rate, it is crucial to think carefully about how rate variation is elicited and what role speech rate should play in the analysis. Previous studies have generally used qualitative/categorical instructions to elicit a handful of rate categories – e.g. fast, normal, slow. This is sub-optimal for the goal of drawing inferences about effects of rate variation, because qualitative/categorical instructions are likely to elicit multimodal distributions of rate. Such distributions are also inappropriate for identifying non-linear relations because the continuum of rates is not sampled with sufficient granularity.

Furthermore, previous analyses have often used the experimentally determined rate manipulations as categorical predictors in statistical analyses. This is problematic for a number of reasons. For one, rate categories do not directly reflect quantitative variation in speech rate but rather discard such information. Moreover, rate categories do not reflect variation in the ways that participants interpret the terms “fast”, “normal”, “slow” etc. To some extent, this issue can be circumvented by including random speaker intercepts and slopes for each category in statistical models, but it fails to address the problem that each category may be associated with more or less variance. Ultimately, categorical analyses cannot provide information regarding how variables change with rate in a *continuous* sense. We believe it is much more reasonable to view rate as an intrinsically continuous phenomenon, as individual speakers can increase or decrease speech rate in fairly small and arbitrary increments.

In fact, there are several studies which have considered speech rate as a continuous variable; in particular, these studies used an empirical measure of rate such as duration of an utterance or syllables per second in statistical analyses. One such example is Byrd and Tan (1996). They examined how the articulation of consonant clusters across a word boundary varies as a function of speech rate, specifically examining the duration of each gesture, temporal latency and overlap between the two gestures, and movement amplitudes. They elicited a range of speech rates using a procedure whereby participants were instructed to increase their speech rate in blocks of sentences. Specifically, they read the first sentence in the block in their normal speech rate, and the subsequent sentences in the block were cued as “medium”, “faster”, and “fastest”. Instead of using these rate categories in their statistical analyses, Byrd and Tan (1996) used the duration of the region around the consonant cluster sequence as a measure of speech rate. They state that “the (these) rate instructions do not serve as variables in a categorical analysis, but rather, served only to engender a wide range of rate variation. Rate is considered a continuous variable in our analyses.” (1996: 268).

Cummins (1999) also used a similar strategy in eliciting and analyzing speech rate. He examined how lengthening associated with prosodic position (phrase-final lengthening) and with contrastive emphasis interact at various speech rates. Speech rate was elicited with the instructions “slow”, “comfortable”, “medium”, “fast”, and “very fast”, together with a graphic of an arrow pointing to the point on a five-point scale. The foot per second (foot/s) of the target region was used as a measure of speech rate in the analyses.

There is also a study which not only used an empirical measure of speech rate but also employed a continuous rate elicitation method. Munhall and Löfqvist (1992) investigated how the opening and closing gestures of the larynx in the /s/-/t/ cluster in the phrase *Kiss Ted* vary as a function of speech rate. Interestingly, rather than using a small number of rate categories to elicit various rates as in most previous studies, they instructed participants to begin speaking at a self-selected “slow” rate and increase rate in small increments until they reach their fastest rate. Thus, they did not pre-define the number of rates that participants should produce nor the rate increments. They further used the interval from the offset of the vowel in *Kiss* to the onset of the vowel in *Ted* as a measure of speech rate.

Since speech rate was treated as a continuous variable in these studies, it is possible to draw more detailed inferences regarding the functional relations between phonetic measures and speech rate. First, Byrd and Tan (1996) found that the temporal coarticulation between successive gestures in a consonant cluster increases linearly as speech rate is increased. However, it should be noted that they only fit a linear model between empirical speech rate and dependent measures, so we do not know whether a linear model explains the data as well as a non-linear model. In addition, as pointed out by the authors, the goodness of the linear fit (R^2) was not very high for most participants and stimuli; the highest R^2 that was found in the participant-pooled data was 0.32. They explained that the low R^2 values may be due to the low sampling rate (100 Hz) or factors other than speech rate that cause variations in dependent measures. Yet, it is possible that the low R^2 values are due to the presence of non-linear relations between the dependent measures and speech rate.

Second, Munhall and Löfqvist (1992) observed linear and non-linear relations between speech rate and laryngeal movement. They measured the interval between the two peaks of glottal opening – one in the /s/ in *Kiss* and the second in the /t/ in *Ted* in the phrase *Kiss Ted*. They found that the distance between the two peaks decreased as speech rate is increased, which eventually became zero at some fast rates as only one peak was observed for the /s+t/ cluster. This in itself is an example of a non-linear relation. Within the range of rates where only one peak was observed, the interval between the peak opening and the frication onset for /s/ increased linearly as speech rate is decreased.

In Cummins (1999), a non-linear relationship was observed between target syllable duration and speech rate in the data of one participant. For this participant, the target syllable duration decreased as speech rate increased, yet there was a region at fast rates where the compression was no longer observed (i.e. reached the maximum compression). The author explained that the non-linearity was observed for this participant as she produced much greater variety of speech rates than the other participants. The general finding of this study was that the lengthening effect induced by prosodic position (phrase-final lengthening) and contrastive emphasis combines in additive fashion in the medial range of speech rate, although the author proposed a possibility that the compression and stretching may be restricted beyond that medial range.

To more thoroughly address the question of how phonetic measures vary with speech rate, we use a recently developed visual motion-based elicitation method, in which a moving visual stimulus is used to iconically represent rate variation. Before producing a target sentence, participants see a red box (rate cue) that moves across the screen at various speeds. They are instructed to base the speed of their response upon the speed of the motion of the cue. Note that participants do not produce the utterance while the cue is moving, so it is not the case that they attempt to match their utterance duration to the period of time that the cue is visible. Ten different cue rates (i.e. the speed of the cue) are used, and they are varied randomly from trial to trial; this method discourages participants from interpreting rate in a categorical fashion.

In addition, we use an empirical measure of speech rate in the analyses; specifically, the duration of the target utterance is adopted as a measure of speech rate. We analyze the empirical speech rate rather than the experimental cue rate, because the latter does not accurately reflect how participants interpret the speed of the cue – i.e. the empirical rate is a better measure, since it reflects the rate of production as implemented by the speaker. It should also be noted that we use the global rate measure (total sentence duration) rather than the local measure (the durations of the target regions from which we extract our dependent variables) in the analyses; see Section 4.1 below for the discussion of global vs. local rate.

2.3. Speech rate and phonetic variables at phrase boundaries

In the current study, we conduct a sentence production experiment where we record acoustic and articulatory data (using electromagnetic articulography) from productions of sentences that contain non-restrictive (NRRCs) and restrictive relative clauses (RRCs). Examples of the two syntactic structures are presented in Table 1. The use of these syntactic contexts is important because they provide us multiple prosodic environments to examine rate effects. The two types of relative clauses are commonly accepted to be semantically and pragmatically distinct. An NRRC contributes information regarding the referent of the expression that it modifies (*Mr. Hodd* in Table 1), but the information is not essential to identifying that referent. NRRCs are often separated in orthography from a main clause by commas, dashes, or parentheses. Native speakers have the intuition that an NRRC can be produced as an “aside” or parenthetical, and silent pauses are felicitous

Table 1
Examples of non-restrictive and restrictive relative clauses and accompanying contexts used in the experiment. Phonetic measures before and after the relative clause (underlined) were examined for the analyses.

Context:	There is one Mr. Hodd. He knows Mr. Robb.
Non-restrictive relative clause (NRRC):	A Mr. Hodd, <u>who knows Mr. Robb</u>, often plays tennis.
Context:	There are two Mr. Hodds. Only one knows Mr. Robb.
Restrictive relative clause (RRC):	The Mr. Hodd <u>who knows Mr. Robb</u> often plays tennis.

before and after the relative clause. In contrast, the information provided by an RRC is essential to identifying a referent from a set of possible referents. For the example in Table 1, one can imagine that there are two different people who are named *Mr. Hodd*, but only one of them knows *Mr. Robb*. The RRC picks out that particular *Mr. Hodd*. It is less natural to pause before or after RRCs compared to NRRCs, and it is unusual to separate RRCs with commas or other punctuation in orthography.

These semantically distinct relative clauses are also considered to have different syntactic structures. There are a number of studies which have argued that the two relative clauses are syntactically distinct (see Arnold, 2007 for an overview). For example, some researchers like Fabb (1990) argued for a radical orphanage approach in explaining the syntax of NRRCs, claiming that an NRRC is not part of the syntactic representation of the sentence that contains the relative clause – in other words, it has no syntactic relation to its antecedent, while an RRC modifies its antecedent. An alternative approach holds that an NRRC is inside the syntactic representation though it is attached to somewhere high in the syntactic tree – for instance, in the root node – thus, making it syntactically distinct from an RRC (e.g. Emonds, 1979; McCawley, 1982). Although there does not exist a consensus on how the difference between the two structures should be represented syntactically, theories generally agree that the two types of relative clauses are structurally different.

Based on their semantic and syntactic differences, the two RCs have also been argued to have distinct prosodic structures. For example, Selkirk (2005) maintained that in NRRCs, the main clause subject (*A Mr. Hodd*) and the relative clause are separate intermediate phrases (ip), and together constitute an entire intonational phrase (IP). On the other hand, in RRCs, the main clause subject (*The Mr. Hodd*) and relative clause together comprise a single intermediate phrase (ip) and are part of an intonational phrase (IP) along with the main clause predicate. Thus, in NRRCs, there are two IPs in total – one for the subject and relative clause each of which is composed of an ip, and the other for the main predicate, whereas in RRCs, there is only one IP which is composed of two ips – one for the subject and relative clause and the other for the predicate. See (1) below. Another prominent analysis was developed by Nespor and Vogel (1986), who argued that an NRRC is obligatorily marked by an IP boundary before and after the relative clause.

(1)

NRRC:	**[[A Mr. Hodd, $_{ip}$]**	**[who knows Mr. Robb, $_{ip}$] $_{IP}$]**	**[[often plays tennis. $_{ip}$] $_{IP}$]**
RRC:	**[[The Mr. Hodd**	**who knows Mr. Robb $_{ip}$]**	**[often plays tennis. $_{ip}$] $_{IP}$]**

We examine phonetic measures associated with phrase boundaries that precede and follow the relative clause for the investigation of functional relations with speech rate. We refer to the boundary before the relative clause (i.e. between *Mr. Hodd* and *who*) as *pre-RC boundary* or *B1* and the boundary after the relative clause (i.e. between *Mr. Robb* and *often*) as *post-RC boundary* or *B2*. Specifically, we investigate articulatory, segmental, and F0 measures associated with the pre-RC and post-RC boundaries; these are the dependent variables in our analyses. The primary independent variable is an empirical measure of speech rate, i.e. here, the target sentence duration.

Bayesian generalized linear model fitting is conducted for the analyses. For a given variable at each boundary, we first compare a constant model (without any rate terms) and a linear model to determine whether that variable is responsive to rate variation. For those variables that varied substantially with rate, we subsequently compare the posterior predictions of linear and non-linear models (quadratic and cubic models) to draw inferences regarding functional relations between phonetic measures and speech rate. In all cases, the Widely Applicable Information Criterion (WAIC, Watanabe, 2013), which estimates out-of-sample deviance, is used to compare the models. Following previous studies which observed individual differences on controlling speech rate, participant-specific random intercepts and slopes are included in the regression models.

With the model comparison results, we first examine whether the functional relations between speech rate and phonetic measures differ by phrase boundaries – namely, pre-RC vs. post-RC boundary. Selkirk (2005) predicted that the two boundaries would differ in categorical boundary strength; in both types of RCs, a stronger prosodic boundary would occur at the post-RC boundary compared to the pre-RC boundary – i.e. NRRC: ip (pre-RC) vs. IP (post-RC), RRC: word boundary (pre-RC) vs. ip (post-RC). On the contrary, Nespor and Vogel (1986) maintained that in NRRCs, the pre-RC and post-RC boundaries are equal in categorical prosodic boundary strength (i.e. IP occurs at both boundaries). We thus investigate whether the functional relations between speech rate and phonetic measures differ by boundary location (pre-RC vs. post-RC boundary), which may or may not differ in prosodic boundary strength. Note that the current study discusses boundary strength in gradient terms, contrary to the previous studies which assumed categorical boundary strength; see below for our perspective on prosodic structure.

In addition, we analyze the effects of different syntactic structures on the relations between speech rate and phonetic variables. Given the semantic, syntactic, and prosodic differences between the two RC types, it is natural to expect that the phonetic measures, especially those at phrase boundaries, would significantly differ by RC type. Thus, in order to assess evidence for the effects of syntactic context on the functional relations with rate, we include syntactic context-specific rate coefficients in our regression models. For each variable at each boundary, we compare the posterior predictions of the best model of the two syntactic structures and examine the absolute normalized differences between them. Since our study examines phonetic variables that are realized at a wide range of rates, we can additionally find out where in the range of observed rates substantial syntactic differences arise. We can also compare which boundary (pre-RC vs. post-RC boundary) exhibits more or less syntactic differences.

Note that in this study, we refer to “boundaries” and “boundary strength” in a neutral sense: in our case, “boundaries” are merely the beginnings and ends of the phrase that corresponds to the relative clause, and we are agnostic regarding whether the “strengths” of these boundaries differ gradiently or categorically. Most theories of prosodic structure (e.g. Beckman & Pierrehumbert, 1986; Nespor & Vogel, 1986; Pierrehumbert, 1980; Selkirk, 1986; for an overview, see Shattuck-Hufnagel & Turk, 1996) hold that boundary strengths differ categorically in a way that reflects a hierarchical organization of phrase types. Although theories differ on the number and definition of phrasal categories and on their ability to nest recursively, the theories in general posit at least two types of phrasal units (e.g. ip and IP), in addition to the prosodic word.

Many empirical studies, however, have found evidence that calls into question the theories which propose a small set of phrase types and categorical boundary strengths. For example, Wightman et al. (1992) found evidence for four distinct prosodic boundaries based on the durations of phrase-final segments. They also suggested that if acoustic cues other than durations were additionally considered, more categories could be distinguished. Similarly, Swerts (1997) reported that their participants could distinguish six levels of discourse boundaries with pause duration. Krivokapić and Byrd (2012) conducted an articulatory experiment and found that speakers produce the same type of prosodic category – particularly, the IP – with different boundary strengths. They further showed that listeners are also sensitive to such fine-grained differences within the IP. In de Pijper and Sanderman (1994), untrained listeners judged the strength of boundaries on a 10-point scale, and the distribution of the scores did not form a small set of clusters, as predicted by the traditional theory of prosodic hierarchy. Taken together, these studies warrant caution in presupposing any particular analysis of prosodic structure.

There are indeed researchers who are opposed to a strictly categorical view of prosodic boundary strength. For instance, Wagner (2005, 2010) proposed that the strength of a given prosodic boundary is scaled “relative” to the boundaries produced earlier. Under this theory, prosodic boundaries may differ in quantitative phonetic effects, not necessarily by qualitative categories. In addition, Articulatory Phonology, specifically the π-gesture theory (Byrd & Saltzman, 2003), is agnostic regarding whether boundaries should be understood to differ categorically or gradiently. π-gestures are systems which are hypothesized to slow down the unfolding of articulatory constriction gestures in the vicinity of phrase boundaries; the more strongly active a π-gesture is, the more articulation is slowed. Byrd and Krivokapić (2021) stated that “within the π-gesture model, boundaries are represented gradiently rather than categorically or symbolically. So, boundaries encoding greater disjuncture (for example, in traditional terms, at an Intonational Phrase versus an Intermediate Phrase) have greater activation and thereby exert a greater slowing or “stretching” of the concurrent gestural activations”. See also Byrd (2006) and Krivokapić (2007) for the discussion on categorical vs. gradient prosodic boundary strength.

In keeping with the latter perspective on prosodic structure, specifically Byrd and Krivokapić (2021), we do not commit to a more conventional, categorical view of prosodic boundary or boundary strength in the current study. Throughout the paper, the term “boundary strength” will be used to indicate relative differences which may or may not be due to categorical differences in phrasal organization. Accordingly, “strong” vs. “weak” prosodic boundaries in this paper would refer to quantitative, phonetic differences; phonetic measures associated with stronger prosodic boundaries would be more extreme (e.g. longer durations, larger movements) than those associated with weaker prosodic boundaries.

In sum, the current study examines how articulatory and acoustic variables at phrase boundaries before and after the relative clause vary with speech rate; in particular, we examine whether there is substantial evidence for rate effects, and if so, whether those effects are best described as linear or non-linear. Following the recommendations in a recent special issue of this Journal – *Emerging Data Analysis in Phonetic Sciences*, eds. Roettger, Winter, and Baayen (2019), we use Bayesian statistical methods and adopt an exploratory approach – i.e. the current study aims to observe patterns and provide the basis for hypothesis generation. For each measure at each boundary, we fit constant, linear, and non-linear models (quadratic and cubic models), compare these models with an information-based criterion (i.e. WAIC), and analyze the model predictions that derive from posterior distributions of model parameters; see Section 3.3 for the details (e.g. random, fixed effects) of the statistical models. The functional relations between rate and phonetic variables are then examined with respect to boundary location and syntactic structure.

We do expect that most of the variables we examine will be responsive to variation in speech rate, based on the studies reviewed in Section 2.1. For the reasons we discussed above, in many cases, previous studies have not been well-suited for characterizing the functional relations between phonetic measures and rate. Thus, it is difficult to make specific predictions regarding which variables or phrase boundaries are more or less likely to exhibit linear or non-linear relations. Regarding syntactic effects, since the two types of RCs are argued to differ semantically, syntactically, and prosodically, we expect to observe substantial differences by RC type in most dependent variables at both boundaries.

3. Methods

3.1. Participants and task

Twelve native speakers of English (6M, 6F) with no speech or hearing disorders participated in the experiment. For six of these (3M, 3F), only acoustic data were collected (acoustic-only sessions); for the other six, articulatory data were collected in addition to acoustic data (articulatory sessions). The experimental task was identical in both cases. In articulatory sessions, participants were seated about 1.5 m from a computer monitor in a quiet room and recorded with a shotgun microphone at a distance of 1.5 m; in acoustic-only sessions, participants were seated in front of a computer monitor in a sound-attenuating booth and wore a condenser microphone

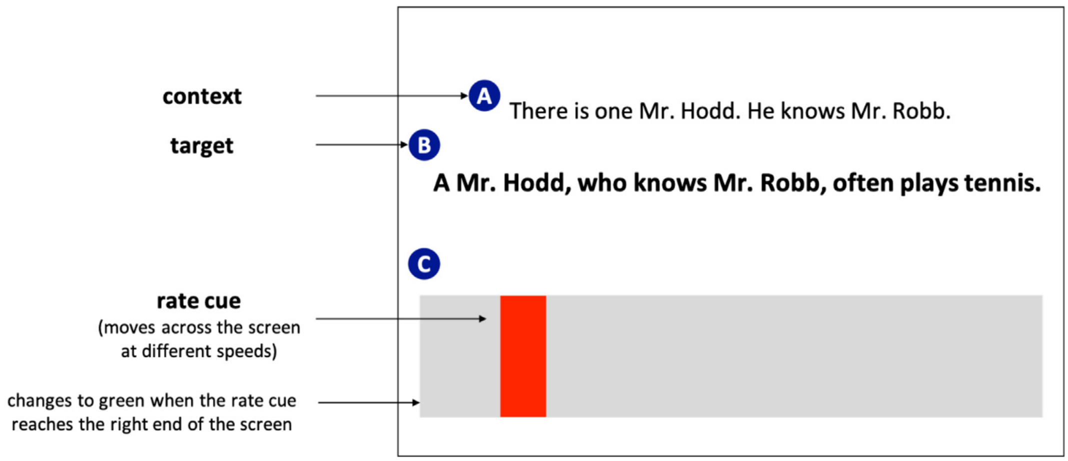

Fig. 1. Presentation of a single trial. Participants were instructed to read the context and target sentences silently when they appeared in sequence (A and B). The rate cue (red box) then showed up and moved across the screen at one of the 10 different cue rates (C). When the cue reaches at the end of the screen and the gray box changes to green (not shown in the figure), participants read the target sentence (B) in a way that reflects the speed of the cue.

(AKG C520 headset). There were six blocks of 40 trials in each experimental session. One type of RC was produced throughout a block, and the blocks alternated between the two types of RCs.

Participants were cued to the target sentence and rate as illustrated in Fig. 1. In each trial, participants first saw two sentences in sequence on the screen. The first sentence (Fig. 1(A)) provided context to the second sentence (Fig. 1(B)), which was the target sentence in the experiment. (see Table 1 for the RRC example.) The purpose of the context was to draw attention to the semantic/pragmatic differences between the two RCs in order to favor the correct syntactic interpretations of the target sentence. For example, the context "*There are two Mr. Hodds. Only one knows Mr. Robb*" picks out one *Mr. Hodd* in particular, encouraging the RRC interpretation of the target sentence. Conversely, the context "*There is one Mr. Hodd. He knows Mr. Robb*" favors the NRRC interpretation of the target, since the second sentence (*He knows Mr. Robb*) simply gives extra information about the referent (*Mr. Hodd*). In addition to the context sentence, definite/indefinite determiners in the beginning of the target sentence (*A, The*) reinforced the semantic difference between NRRCs and RRCs. The determiner "*the*" favors the RRC interpretation because it presupposes the existence of more than one person named *Mr. Hodd*. No participants reported any difficulty in interpreting the meanings of the two types of sentences.

Participants were instructed to read both the context and target sentences silently when they first appeared (Fig. 1(A) and (B)). After 1.5 seconds, the visual rate cue appeared, which is the red box in Fig. 1(C). This cue moved from left to right across the screen at one of 10 different speeds. These speeds were specified as periods of time and were equally spaced from 0.8 s to 4.1 s; the endpoints of this range were derived from the observed durations of the target sentence in pilot data, which were collected from two native speakers of English. We refer to these periods as *cue rates*. Note that the cue can be considered an "analog" stimulus because participants were not able to identify the 10 unique rates; rather, their subjective impression was that the cue rate varied on a continuous scale.

Participants were instructed that when the cue disappears and the gray box changes to green, they should produce the target sentence in a way that reflects the speed of the cue. Crucially, they were told to wait until the cue passes across the screen and the box becomes green before initiating their response. Therefore, they did not match their production to the period of time in which the cue was moving, but instead, they varied the speed of their production based upon their impression of how fast or slow the cue was moving. It is important to emphasize that the rate cue was used to indirectly and iconically elicit a wide variation in rate, rather than imposing a specific timeframe on production. The speed of the cue was varied randomly from trial to trial by selecting one from the set of 10 cue rates. (cf. A similar method was used in Tilsen and Hermes (2021) and Hermes et al. (2021), except that the cue rates were presented in ascending and descending orders in these studies.) Each cue rate appeared the same number of times in each experimental session. Although the cue rates were used to elicit various speech rates, we used empirical measures of rate, specifically, the utterance durations produced by participants, in the analyses (see Section 3.3 below).

The target words in the experiment were the names that follow the honorific "Mr*.*" All the names were monosyllabic with a CVC form. The onset consonants of these target words were /h/, /r/, or /l/, and the codas were /b/ or /d/. The vowel /a/ was used in all cases. Therefore, a total of six names were used throughout the experiment (*Hobb, Robb, Lobb, Hodd, Rodd, Lodd*). We restricted the set of coda consonants to /b/ and /d/ in order to facilitate articulographic analysis in the vicinity of phrase boundaries. At each boundary, the target word was randomly selected in a way that ensures each participant to produce 120 /b/-final and 120 /d/-final forms. Thus, it was not the case that each of the six target words appeared for the fixed number of times, but rather they were randomly selected based only on the coda consonants. Note that no trials had the same name at the pre-RC and post-RC boundaries. For the final word of the sentence, four different di-syllabic words (*tennis, hockey, soccer, baseball*) were used, and each word appeared the same number of times.

In order to prevent participants from putting emphatic focus on the names, which may induce prominence and affect natural phrasing of the sentence, they were explicitly instructed not to emphasize any of the words in the target sentence. In addition, participants performed sixteen practice trials (which included both NRRCs and RRCs) under experimenter supervision before beginning the experiment. If they put focus on the names, they were corrected by the experimenter.

3.2. Data collection and processing

In both articulatory and acoustic-only sessions, acoustic data were collected at 22,050 Hz. Acoustic segmentation was conducted using Kaldi (Povey et al., 2011). For each participant, six trials were manually labeled and used to train monophone HMMs. Then, a forced alignment was applied to all trials. The alignments of all trials were manually inspected and corrected when necessary. Acoustic durations of the coda and vowel portions of the target words were obtained from the segmented data.

F0 data were extracted using Praat. We first identified a participant-specific F0 range with the first 20 trials of each participant. Specifically, we collected all F0 values of the first 20 trials, removed outliers, and calculated the range of the remaining F0 values, which was also sanity checked considering the gender of the participant. With this participant-specific F0 range, F0 data were extracted in Praat with a timestep of 5 ms using auto-correlation method. F0 values near boundaries were then smoothed/interpolated with a cubic spline method. We defined the target region around phrase boundaries from the honorific "Mr." to two syllables after the boundary; for example, in "*A Mr. Hodd, who knows Mr. Robb, often plays tennis*", the target region for the pre-RC boundary is "*Mr. Hodd, who knows*", whereas it is "*Mr. Robb, often*" for the post-RC boundary. The linearly time-warped F0 contours of the target region from each participant in each syntactic structure are presented in Fig. 13 (pre-RC boundary) and Fig. 14 (post-RC boundary). At both boundaries, at least one peak was found at the pre-boundary region and post-boundary region, respectively. Yet, some participants had multiple peaks within each region; others had the highest F0 value at the beginning of the target region rather than the peak (see Fig. 13, Fig. 14). Thus, we recorded F0 maxima at each region, rather than the F0 value of the peak, and the difference between the two maximum values; see Fig. 2 for an example. Note that we do not analyze F0 valley (F0 minimum), which usually appeared adjacent to the phrase boundary. This is because in some participants or at certain rates, F0 valley was accompanied with creak, and moreover, it sometimes occurred at the coda of the target word, which was often devoiced. This made it difficult to reliably measure F0 minima across all participants and rates.

Articulatory data were collected with an NDI Wave Electromagnetic Articulograph (EMA) with a sampling rate of 400 Hz. Articulator sensors were located mid-sagitally on the upper lip (UL), lower lip (LL), gum below the lower incisors (JAW), tongue tip (TT, about 1 cm from tongue apex) and tongue body (TB, 4–5 cm posterior from TT). Reference sensors for head movement correction were located on the nasion and left and right mastoid processes.

Articulatory data were processed as follows. First, reference and articulator sensors were filtered at 5 and 10 Hz respectively, using 3rd order lowpass Butterworth filters. Then, head movement was corrected by transforming each frame of data such that the reference sensors were located at a fixed position. The horizontal and vertical coordinates of the articulator sensors were then resampled to 1000 Hz to allow for more precise identification of gestural landmark times. A lip aperture signal (LA) was defined by calculating the Euclidean distance between the UL and LL sensors. Fig. 3 shows the kinematic

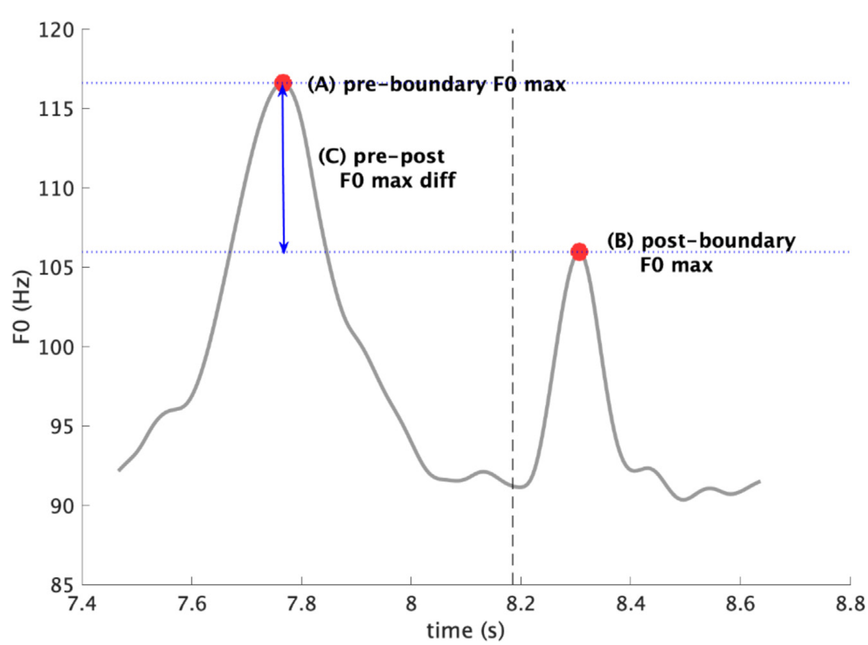

Fig. 2. F0 variables examined in the study. An example of a smoothed and interpolated F0 contour at the pre-RC boundary is shown (gray line). The vertical dashed line marks the end of the target name – i.e. the location of the phrase boundary. We measured (A) F0 maxima of the pre-boundary region, (B) F0 maxima of the post-boundary region, and (C) the difference between the two max values.

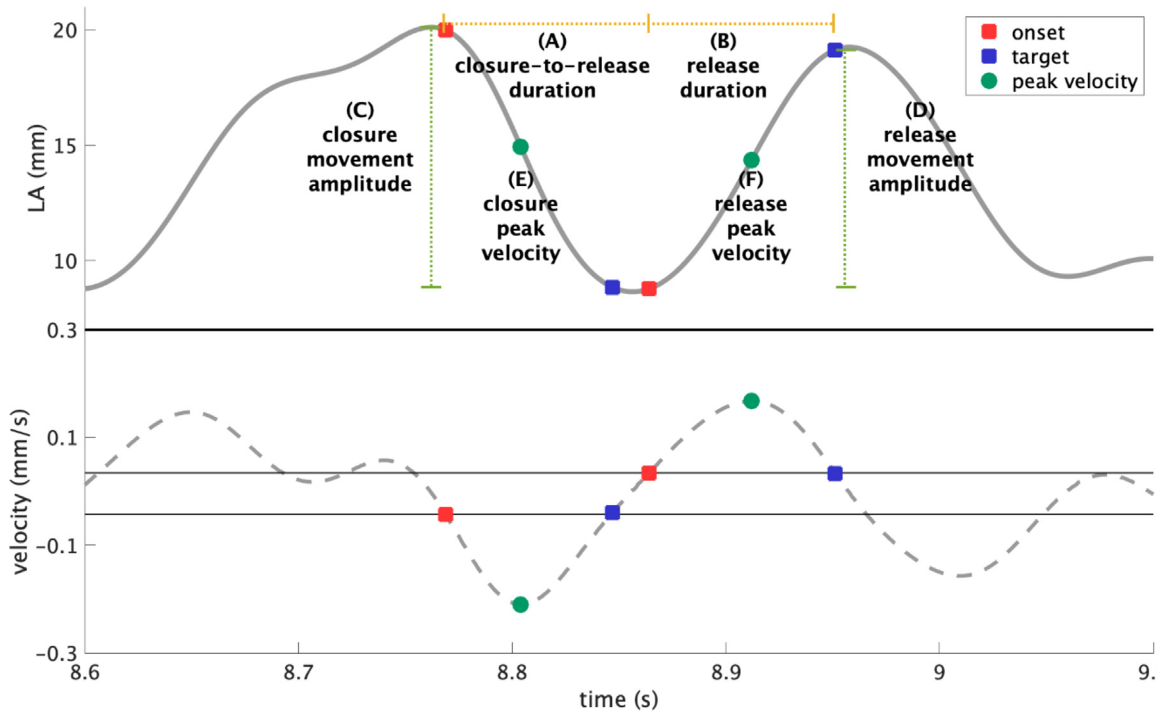

Fig. 3. Kinematic landmarks and dependent articulatory variables examined in the study. An example of an LA trajectory (top panel) and the accompanying velocity trajectory (bottom panel) for the target word *Robb* at the post-RC boundary is presented (gray lines). Squares and circles represent gestural onsets, targets, and peak velocities. For each boundary, closure-to-release durations (closure + plateau durations), release durations, and amplitudes and peak velocities of closure and release movements were measured ((A)–(F)).

landmarks along with the dependent measures for the articulatory analyses, using an example of the target word *Robb* at the post-RC boundary. Based on the acoustic segmentation, relevant velocity extrema were detected for consonantal closure and release gestures of the coda of the target words at each boundary (green dots in Fig. 3). Gestural onsets and targets were then identified in relation to velocity extrema. Specifically, the gestural onsets were the time points where the velocity signal first rises above 20% of the peak velocity, and the targets were the time points where the signal first falls below 20% of the maximum velocity, which are shown as red and blue squares respectively in Fig. 3. The target words in the experiment ended either in /b/ or /d/; thus, for the words that ended with /b/, the LA signal was analyzed, while for the words that ended with /d/, the vertical position of the TT sensor was analyzed. For articulatory analyses, we measured closure-to-release durations (closure + plateau durations), release durations, amplitudes of closure and release movements, and peak velocities at each boundary.

To characterize local speech rate slowing or pausing at prosodic boundaries, we used a hybrid articulatory/acoustic measure instead of pause durations. Pause durations are somewhat problematic because it can be hard to distinguish pauses from periods of low acoustic intensity associated with the often devoiced codas of the target words. Moreover, pause durations are undefined on trials in which no silent interval is detected, which makes them less suitable for use in regression analyses. To circumvent these issues, we adopted a hybrid articulatory/acoustic measure, the trans-boundary interval (TBI), which is well-defined on all trials. Specifically, the TBI is the period of time from the target achievement of the pre-boundary coda closure (measured articulatorily) to the acoustic onset of the post-boundary vowel, which closely corresponds to the onset of voicing. For example, consider the utterance "*A Mr. Hodd, who knows Mr. Robb, often plays tennis*". Here, the TBI of B1 is the time from the target achievement of the alveolar closure in *Hodd* to the onset of voicing for the vowel /u/ in *who*; likewise, the TBI of B2 is the time from the target achievement of the bilabial closure in *Robb* to the onset of voicing for the vowel /a/ in *often*. The TBI can thus be readily interpreted as a measure of slowing/pausing associated with a phrase boundary.

A total of 240 trials were collected for each of the 12 participants. For articulatory sessions (1440 trials), 127 trials (8.8%) were excluded from analyses due to speech errors, disfluencies, or problems in data collection, and 56 trials (3.9%) due to problems in detecting landmarks. For acoustic-only sessions, 50 out of 1440 trials (3.5%) were discarded for speech errors, disfluencies, and data collection problems. To identify outliers in a way that takes into account rate-related variation, mixed effects linear regressions with target sentence duration as a fixed effect and participant as a random intercept were conducted at each boundary for all variables. Datapoints whose standardized absolute residuals were larger than 2.326 (the 1/99 percentiles of a normal distribution) were excluded from subsequent analyses. Regarding speech rate, empirical speech rates elicited from all participants were z-score normalized, and the rates whose absolute residuals were larger than 2.576 (the 0.5/99.5 percentiles) were excluded from the analyses. This was to ensure that the speech rates we collected are considered as "natural speech". We reasoned that the rates should be comparable across participants to be considered as "natural". Although one can produce extremely fast or slow rates by deleting syllables or

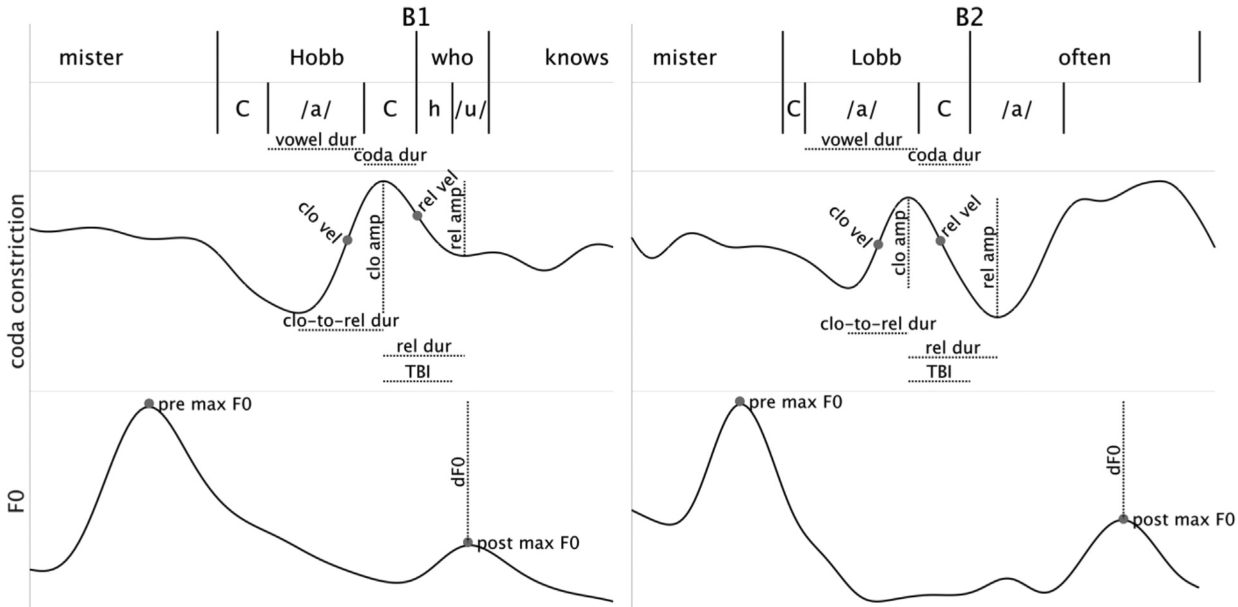

Fig. 4. Schematic representation of dependent variables in our analyses. The left panel shows variables at the pre-RC boundary (B1) and the right panel shows those at the post-RC boundary (B2). Top: segmental measures; middle: articulatory measures; bottom: F0 measures.

inserting extra pauses, if the rates that are produced this way are far from the range of rates from 12 participants, we should be skeptical on whether those rates can be considered as "natural". Thus, we excluded extreme rates that are outside the 99.5% range of the empirical speech rates, which resulted in the exclusion of 45 rate datapoints (out of 2647, 1.7%).

3.3. Data analysis

The rate measure in our analyses is the duration of the produced target sentence. Note that we used this global empirical rate measure rather than the local rate or the experimentally manipulated cue rate; see Sections 2.2 and 4.1 for motivation. Fig. 4 and Table 2 summarize the dependent variables in our analyses. 12 articulatory and acoustic variables were measured at each boundary (pre-RC and post-RC boundaries), resulting in a total of 24 variables.

We used Bayesian generalized linear model regression in our analyses. Following recommendations in McElreath (2020b), we used hierarchical models with regularizing priors, which we examined using prior predictive simulations. The regression procedure provides posterior distributions of model parameters that we analyze in several ways. Our motivation for conducting Bayesian analyses is that we do not intend to conduct confirmatory hypothesis testing regarding whether there exist certain forms of functional relations between speech rate and phonetic variables. Instead, we aim to characterize the strength of evidence for various types of relations without imposing decisions regarding which relations are present. The Bayesian approach is more appropriate given the exploratory nature of our study, and as mentioned above, adopting this

Table 2
Summary table of dependent variables. The following variables are measured at the pre-RC and post-RC boundaries respectively.

name	abbrev.	description
articulatory		
closure-to-release duration	clo-to-rel dur	the duration from the onset of the closure to the onset of the release of the target word coda (i.e. closure + plateau durations)
release duration	rel dur	the duration from the onset to the target of the release of the target word coda
closure movement amplitude	clo amp	the magnitude of the movement in the closure phase of the coda
release movement amplitude	rel amp	the magnitude of the movement in the release phase of the coda
closure peak velocity	clo vel	the peak velocity in the closure phase of the coda
release peak velocity	rel vel	the peak velocity in the release phase of the coda
trans-boundary interval	TBI	the duration from the target achievement of the coda closure (pre-boundary) to the onset of voicing (post-boundary)
segmental		
vowel duration	vowel dur	the duration of the target word vowel
coda duration	coda dur	the duration of the target word coda
F0		
pre-boundary F0 max	pre max F0	the F0 maxima of the pre-boundary target region
post-boundary F0 max	post max F0	the F0 maxima of the post-boundary target region
pre-post F0 max diff	dF0	the difference between the pre-boundary and post-boundary F0 maxima

approach is consistent with recommendations made in a recent special issue of this Journal (Nicenboim, Roettger, et al., 2018; Roettger et al., 2019; Vasishth et al., 2018).

A complete description of our regression procedures, along with R code, is provided in an Appendix: Bayesian generalized linear model regressions. Here we mention the most important details. All regressions were conducted in R (R Core Team, 2020) using the Hamiltonian Markov Chain Monte Carlo (MCMC) sampler in the RStan package (Stan Development Team, 2020). The independent variable of speech rate (RATE) and all dependent variables were z-score normalized for the regressions. Visual depictions of results in the manuscript, however, are presented in the original data units. Four different generalized linear models were fit to each variable: (1) *non-linear cubic model* (RATE3) that has intercepts and linear, quadratic, and cubic rate effects; (2) *non-linear quadratic model* (RATE2) that has intercepts, linear, and quadratic rate effects; (3) *linear model* (RATE1) that has intercepts and linear rate effects; and (4) *constant model* (RATE0) that has only intercepts. The intercepts and slope coefficients of all our models included random effects for each participant and relative clause combination and fixed effects of relative clause type. In other words, each term of the polynomial included an interaction between speech rate and clause type. In addition, except for the F0-related dependent variables, a fixed effect of the place of articulation for the target word coda (labial/coronal) was included in the intercepts.

To identify the variables that are responsive to rate changes and characterize their functional relations with rate, we compared the models using the Widely Applicable Information Criterion (WAIC, Watanabe, 2013) calculated with the *rethinking* package (McElreath, 2020a). We first compared the WAIC of the constant model (RATE0) and the linear model (RATE1) to determine which dependent measures exhibit substantial evidence of rate effects. We calculated the mean WAIC difference between the two models (i.e. WAIC of the constant model – WAIC of the linear model, henceforth ΔWAIC) and its standard error interval. In keeping with Bayesian philosophy, we do not impose an arbitrary threshold to decide whether a rate effect is or is not significant. Instead, we consider the strength of evidence for such effects by assessing the ΔWAIC of the models in light of the standard error of that difference. In general, we conclude that there is substantial evidence for one model over another when the ±2.0 standard error interval of ΔWAIC does not include 0. However, it is important to keep in mind that evidence for effects is always gradient and not directly amenable to categorical decisions about the presence/absence of effects. For those variables that did not show substantially better predictions in the linear model when compared to the constant model, we did not make further comparisons between the constant and non-linear models. This is because we reasoned that if there is no strong evidence for even a linear relation between a given variable and rate, there is not much to draw from the fact that the non-linear models provide better predictions than the constant model.

For those measures that were observed to be responsive to changes in rate, subsequent comparisons were made between the linear model (RATE1) and the non-linear models (RATE2, RATE3); this was to determine whether there is evidence that a given variable varies linearly or non-linearly with rate. Note that since the focus of this comparison is on linear vs. non-linear relations, we do not systematically examine which non-linear model – quadratic vs. cubic model – explains the data better; instead, we attempt to interpret the posterior predictions of the best non-linear model. Also for that reason, we limit our analyses only to the quadratic and cubic models rather than testing other non-linear functions or polynomials beyond 3rd order. As for the constant vs. linear model comparisons, we examined the ΔWAIC and its standard error interval to assess the strength of evidence for non-linear rate effects. These model comparison results were then analyzed with reference to boundary location (pre-RC vs. post-RC boundary).

To infer whether there is evidence for phonetic differences between relative clause types (NRRCs vs. RRCs), we assessed whether posterior predictions of the best model for a given variable substantially differ between the two syntactic structures. For this purpose, for each dependent measure at each boundary, we calculated the average of the absolute posterior differences in a 30-point grid defined over the range of normalized rates, after the differences have been normalized by their standard errors. This measure provides an interpretable characterization of whether the model predicts substantial differences between the relative clauses, in a way that incorporates all the rate terms of the model – i.e. the linear terms and non-linear terms when motivated.

4. Results

Our rate elicitation method resulted in a fairly wide range of rates that show a unimodal distribution. In the model comparisons, 15 out of 24 dependent variables were found to be responsive to rate variation; 10 of the 15 showed evidence for non-linear relations with rate. We found that the phonetic variables associated with the post-RC boundary (B2) were more likely to be responsive to rate changes compared to those at the pre-RC boundary (B1). Moreover, phonetic measures varied more extensively at B2, and this resulted in strong non-linear relations, specifically ceiling or floor attenuation effects at extreme rates. These results suggest that the phonetic variables at B2 are more susceptible to variation in speech rate than those at B1. Highly participant-specific patterns were observed in the posterior predictions of the F0 measures, less so for other variables. A relatively robust RC type effect was observed in the durational measures associated with the pre-RC boundary.

Section 4.1 presents distributions of empirical speech rate, which is the independent variable in our analyses. Section 4.2 shows the results of model comparisons for the purpose of identifying functional relations between speech rate and dependent variables. Section 4.3 presents the analysis of syntactic effects. Note that in most of the figures presented below, the rightward direction of the x-axis corresponds to slower rates (longer utterance durations).

4.1. Speech rate variation

Empirical speech rates observed in the experiment were highly correlated with cue rates and exhibited unimodal distributions. Fig. 5 shows the relation between empirical speech rates and cue rates, along with the distributions of empirical

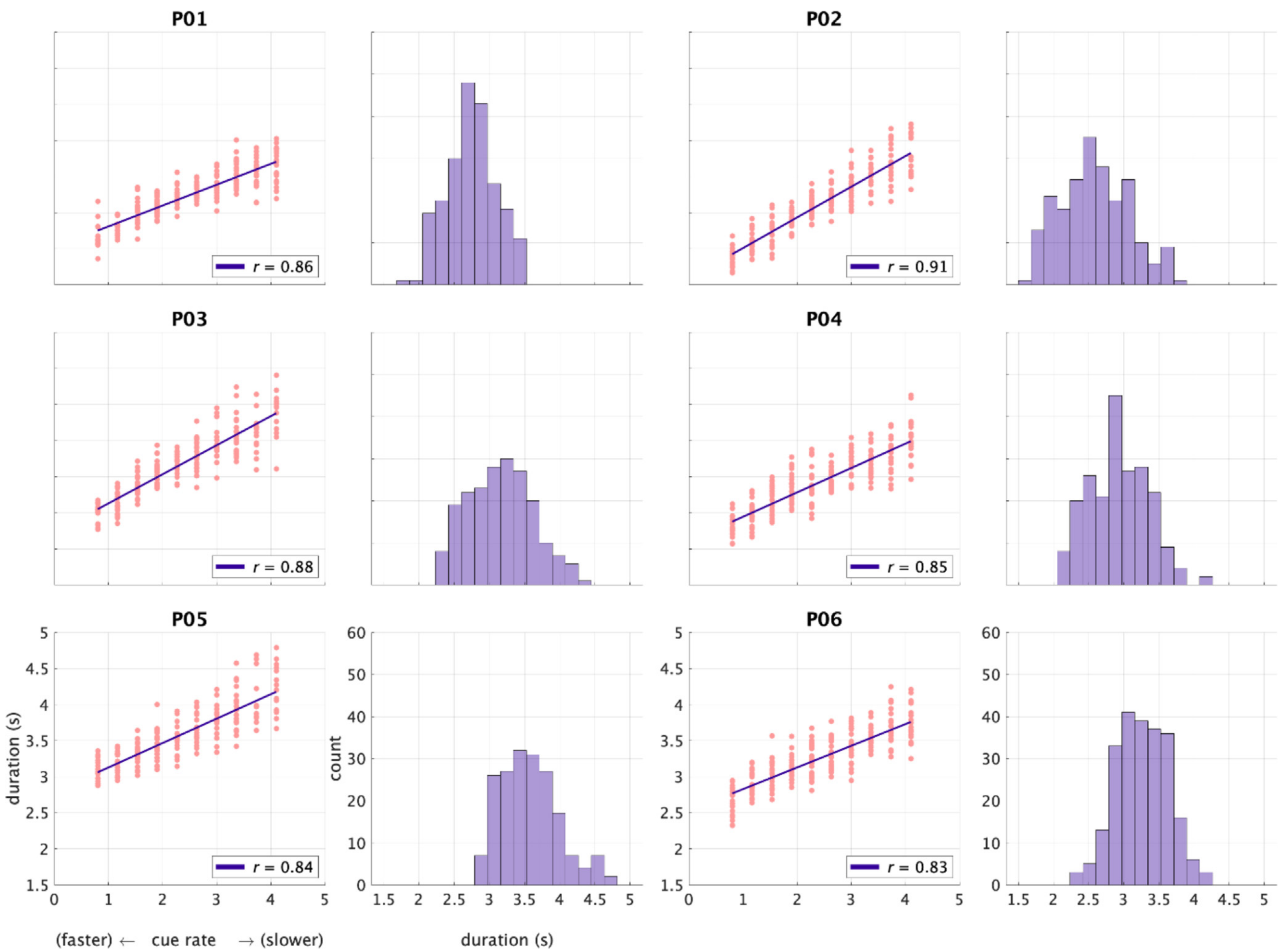

Fig. 5. The relation between empirical speech rates (target sentence durations) and cue rates (cue durations which varied in 10 steps from 0.8 s to 4.1 s) and the distribution of empirical speech rates in each participant. The figure shows six participants' data in the articulatory sessions. For each participant, the left panel shows how empirical rates vary as a function of cue rates. The dark blue line shows the result of a linear regression between the target sentence durations and cue rates, and the correlation coefficient (*r*) is presented on the bottom right. The right panel shows the histogram of the empirical speech rate for a given participant.

rates for each participant in the articulatory sessions. The average linear correlation coefficient between empirical rates and cue rates across all participants was 0.86. This confirms that participants did in fact vary speech rate in a way that reflects the speed of the rate cue. From the histograms in Fig. 5, it is evident that the empirical speech rates collected from each participant showed unimodal distributions. These distributions were more uniform (i.e. flatter, with thicker tails) than Gaussian distributions. The kurtosis (a measure of "tailedness") of the empirical rate distribution was below 3 (the kurtosis of normal distribution) for 10 of the participants, and the average kurtosis across all 12 participants was 2.72. Though the number of trials for each of the 10 cue rates was identical in each experimental session, the distribution of empirical rates had a clear modal value. Perhaps, this suggests that participants tend to be biased toward central rates. Although not shown in the figure, rate distributions from participants in the acoustic-only sessions were qualitatively similar to those participants shown in the figure.

Our rate elicitation method further resulted in a wide range of variation in speech rate. Across participants, the range of empirical rates we observed was 1.3 s/utterance to 4.79 s/utterance. When these durations are transformed to syllables per second (σ/s), the range that we elicited was 2.9–10.8 σ/s. For comparison, the 99.9 percentile range of rates in the Haskins Production Rate Comparison database (HPRC, Tiede et al., 2017) is 2.8–8.7 σ/s. Thus, in comparison to that corpus, the fastest rates that our dataset contains are faster, while the slowest rates are quite similar. However, it is important to emphasize that it is not necessarily possible to make direct comparisons of rates between utterances with different phrasal and syntactic properties, or between different tasks/elicitation methods. In particular, while the current experiment used long sentences that contain relative clauses and further probed the same two structures repeatedly, the sentences in HPRC database were much shorter and unique. Hence, it is not clear exactly how much rate variation we would expect, compared to the variation measured from the HPRC database. Although further research on how experimental stimuli and tasks/elicitation methods influence observed rates is warranted, the comparison with the HPRC corpus leads us to be confident that we have obtained a wide range of rate variation.

The rate measure that we used in our analyses is the duration of the target sentence, which is the most global measure we can use. Yet, speech rate is also modulated more locally, particularly at phrase boundaries, and it is important to confirm that the local rate modulations properly reflect the variations of the speed of the cue. We thus examined the correlations between the experimental cue rates and the durations of the target regions. As in extracting F0 (Section 3.2), we defined the target region for the pre-RC boundary as "*Mr. Hodd, who knows*", and the target region for the post-RC boundary as "*Mr. Robb, often*" in the sentence "*A Mr. Hodd, who knows Mr. Robb, often plays tennis*". The average correlation with the durations of the pre-RC region across participants was 0.83, and it was 0.85 for the post-RC region, both showing a very high degree of correlation (cf. average correlation with the total sentence duration: 0.86). These correlations show that participants also varied rate at target regions in a way that reflects the experimental cue manipulation. Since both total sentence and target region durations exhibited a high correlation with cue rates, we used the former as an independent variable in the analyses for simplicity. This also avoids the issue that more local measures of rate may be conflated with some of our dependent measures (such as segment durations).

4.2. Functional relations between speech rate and dependent variables

In the first set of model comparisons, we compared the WAIC of the constant model and the linear model for each dependent variable at each boundary. For a total of 15 out of 24 variables (62.5%), WAIC comparisons indicated that the linear model substantially outperformed the constant model. For those variables that were responsive to variation in speech rate, we subsequently compared the linear model and the best non-linear model – either quadratic or cubic model – to find out which model provides better predictions of phonetic measures as a function of rate. Out of the 15 variables, there were 10 variables (66.7%; cf. out of 24, 41.7%) for which one of the two non-linear models substantially outperformed the linear model. Table 3 provides a full list of variables, grouped by the best model.

Examining the model comparison results in relation to boundary location, we found that most of the articulatory variables at the pre-RC boundary (B1) did not exhibit strong rate effects; only the TBI was responsive to rate changes at this boundary. At the post-RC boundary (B2), however, several articulatory measures – movement duration, amplitude, and TBI – as well as acoustic variables did exhibit substantial variation with rate. Note that at the post-RC boundary, articulatory variables that are closer to the boundary (i.e. measures at the release phase of the coda and TBI) were more responsive to variation in speech rate. At both boundaries, movement velocities of the closure and release phases of the target word coda constriction did not vary with changes in rate.

For those variables that varied substantially with rate (second and third rows of Table 3), we found that the articulatory and segmental measures were more likely to exhibit non-linear relations at the post-RC boundary, whereas linear relations were found for those measures at the pre-RC boundary. There was only one exception: vowel duration of the target word at the pre-RC boundary showed evidence of a non-linear relation with rate. For F0 variables, the posterior predictions of the measures at the post-RC boundary showed linear relations, while the variables at the pre-RC boundary showed non-linear relations. For illustration of mean ΔWAIC and its standard error for the rate-responsive variables, see Fig. 6.

We examined the posterior predictions of a subset of the variables where we observed large non-linear rate effects – i.e. large WAIC differences between the linear and non-linear models. The coda duration at the post-RC boundary was first examined. Fig. 7 compares predictions of the quadratic model and the linear model in each syntactic structure. In both RCs, we observed a ceiling pattern in the predictions of the quadratic model: coda durations increase as the rate slows down, but at very slow rates, the durations attenuate as if they have reached a ceiling. This attenuation/ceiling effect was also observed for other variables such as the release amplitude at B2, vowel durations at B1 and B2, closure-to-release duration and release duration at B2.

Regarding the TBI at the post-RC boundary, we found a floor attenuation effect in both syntactic structures. As shown in Fig. 8, the post-RC TBI overall increased as speech rate was decreased; however, at fast rates, the amount of increase diminished. This suggests that the overlap between the gestures across a phrase boundary reached its maximum at fast

Table 3
Summary of model comparison results. Variables in the first row did not exhibit substantial rate effects. Variables in the second row exhibited substantial evidence of linear relations, and those in the third row exhibited evidence of non-linear relations.

	pre-RC boundary (B1)			post-RC boundary (B2)		
	articulatory	**segmental**	**F0**	**articulatory**	**segmental**	**F0**
constant	clo-to-rel dur rel dur clo/rel amp clo/rel vel			clo amp clo/rel vel		
linear	TBI	coda dur				pre max F0 post max F0 dF0
non-linear		vowel dur	pre max F0 post max F0 dF0	clo-to-rel dur rel dur rel amp TBI	vowel dur coda dur	

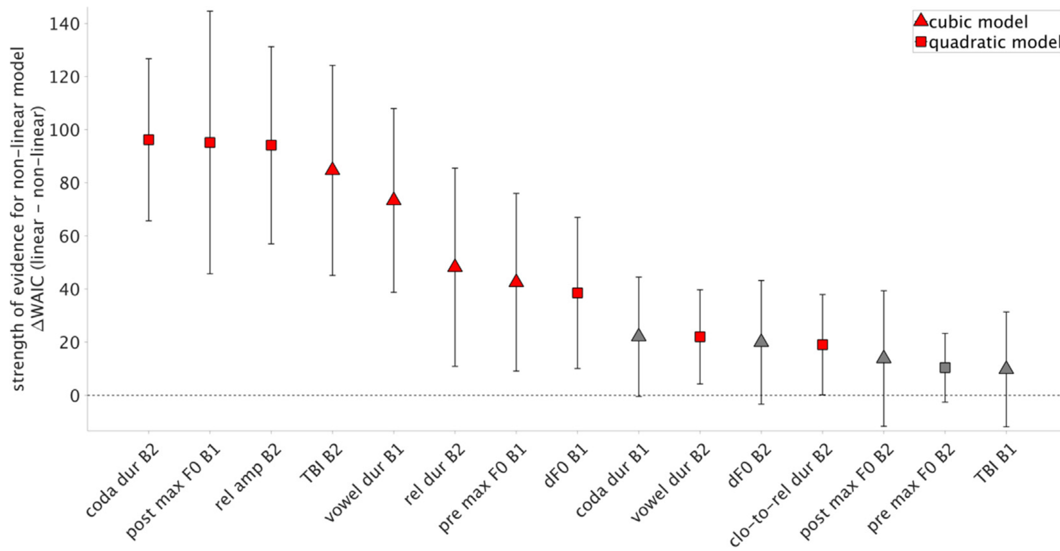

Fig. 6. The difference between the WAIC of the linear model and the best non-linear model and the standard error interval. Only variables that exhibited substantial evidence for linear models when compared with constant models are shown. For each measure, the marker shows the mean ΔWAIC, and the whiskers show the ±2.0 standard error intervals. For variables whose standard error intervals do not include 0, the markers are colored in red; the red triangle shows the variables with better predictions in the cubic model, whereas the red square shows the variables for the quadratic model. The variables whose interval included 0 are marked in grey; these are the cases where the linear model may provide better predictions than the non-linear models. The variables are listed in the order of ΔWAIC.

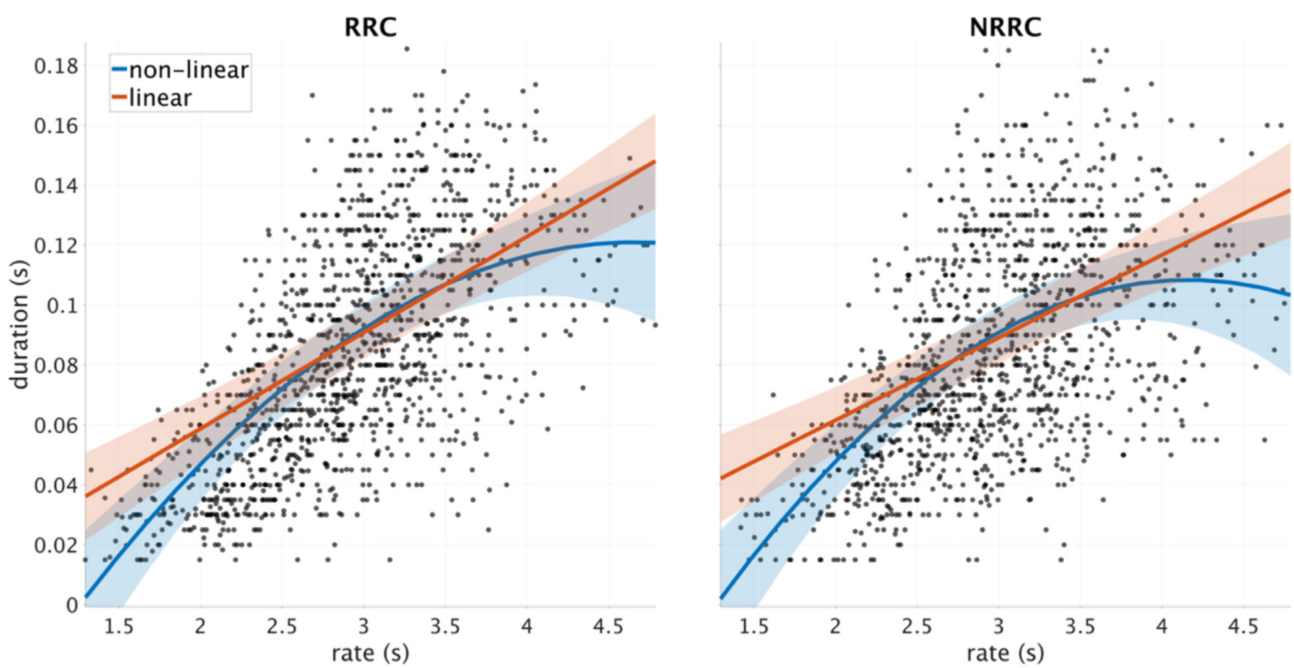

Fig. 7. Posterior predictions of the linear and non-linear models of coda duration at B2. The predictions are plotted separately for each syntactic structure; left panel: RRC, right panel: NRRC. In this and subsequent figures, lines show the mean posterior predictions, and the shaded area represents 5–95% confidence intervals. Red lines: linear model (RATE1), blue lines: best non-linear model – i.e. the quadratic model (RATE2).

rates (i.e. floor effect). Interestingly, both floor and ceiling attenuation effects were observed at extreme rates in NRRC.

We additionally examined the distributions of coda durations and TBI at each boundary and found that these measures at B2 expanded and compressed to a greater extent than the measures at B1. See Fig. 9 which shows Gaussian kernel density functions for coda duration (left panel) and TBI (right panel). The distributions of both measures at the post-RC boundary are broader, extending to larger values. For other variables which showed ceiling attenuation effects – the release duration and amplitude at B2, the measures at B2 also expanded further than B1. Yet, the distribution of the vowel durations, where non-linear relations were observed both at B1 and B2, was similar across the boundaries. Interestingly, the TBI at the post-RC boundary also decreased to a greater extent than it did at the pre-RC boundary; this is confirmed in the right panel of Fig. 9, where the peak in the Gaussian kernel density of B2 is located to the left of the peak of B1. This suggests that phonetic measures at B2 may not only expand but also in some cases compress to a larger extent compared to

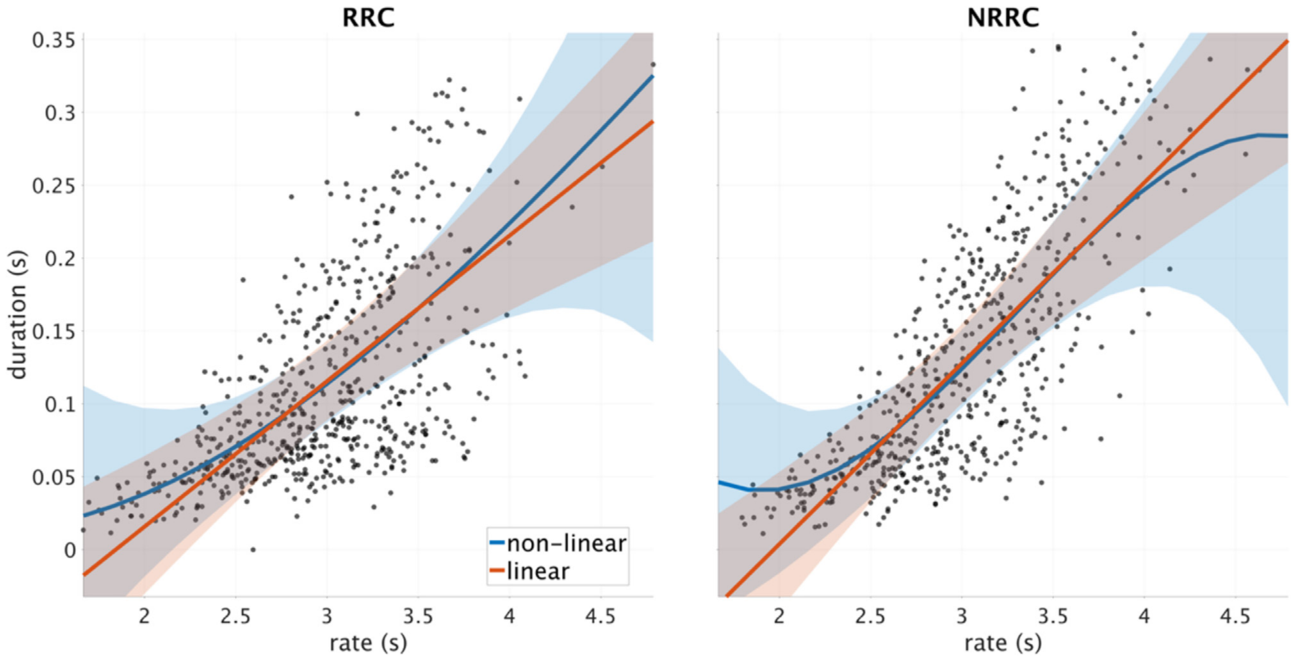

Fig. 8. Posterior predictions of the linear and non-linear models of the TBI at B2. Red lines: linear model (RATE1), blue lines: best non-linear model – i.e. the cubic model (RATE3).

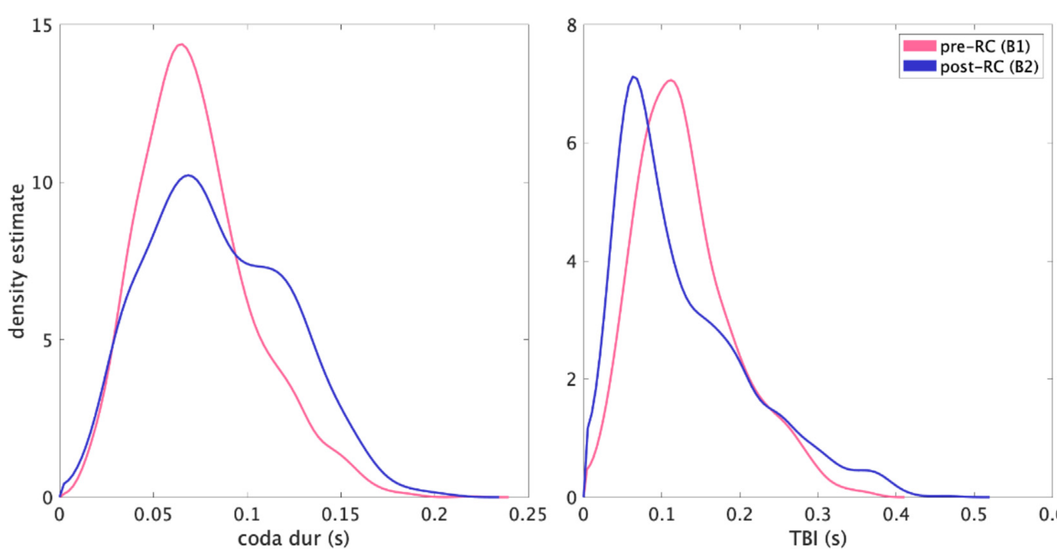

Fig. 9. Kernel density estimation of coda duration (left panel) and TBI (right panel) at each boundary. In both panels, the pink lines show the density estimates of the measures at the pre-RC boundary, whereas the blue lines show the measures at the post-RC boundary.

those at B1. This explains the observation that B2 measures are more likely to have non-linear relations with rate, specifically the attenuation effects at extreme rates.

Another variable with a substantial preference for the non-linear model was the post-boundary F0 maximum at B1. Unlike the above variables, where we could observe a clear distinction between linear and non-linear model predictions in fixed effects, the difference between the models was not as apparent. Inspection of posterior predictions of the random effects indicated a substantial amount of variation across participants. This is illustrated in Fig. 10. A variety of patterns were observed among participants: for example, some of them showed a floor pattern (e.g. P05 NRRC, 07, 08, 12), while others showed a ceiling pattern (e.g. P03, 06). Further, participants differed on which extremes of the range of rates they attenuated F0: ceiling effects were found at fast rates, whereas floor effects tended to show up at slow rates. These various patterns suggest that most of the differences between the linear and non-linear models for the F0 variables were manifested in the random effect parameters. Inter-participant differences were observed not just in the post-boundary F0 maxima (Fig. 10) but also in other F0 measures (i.e. pre-boundary F0 max, pre-post F0 max diff) at the pre-RC boundary. A detailed discussion on F0 variables, especially focusing on their participant-specificity, is presented in Section 5.3.

To sum up, among 24 variables we examined, 15 of them (62.5%) showed substantial evidence of rate effects. Acoustic

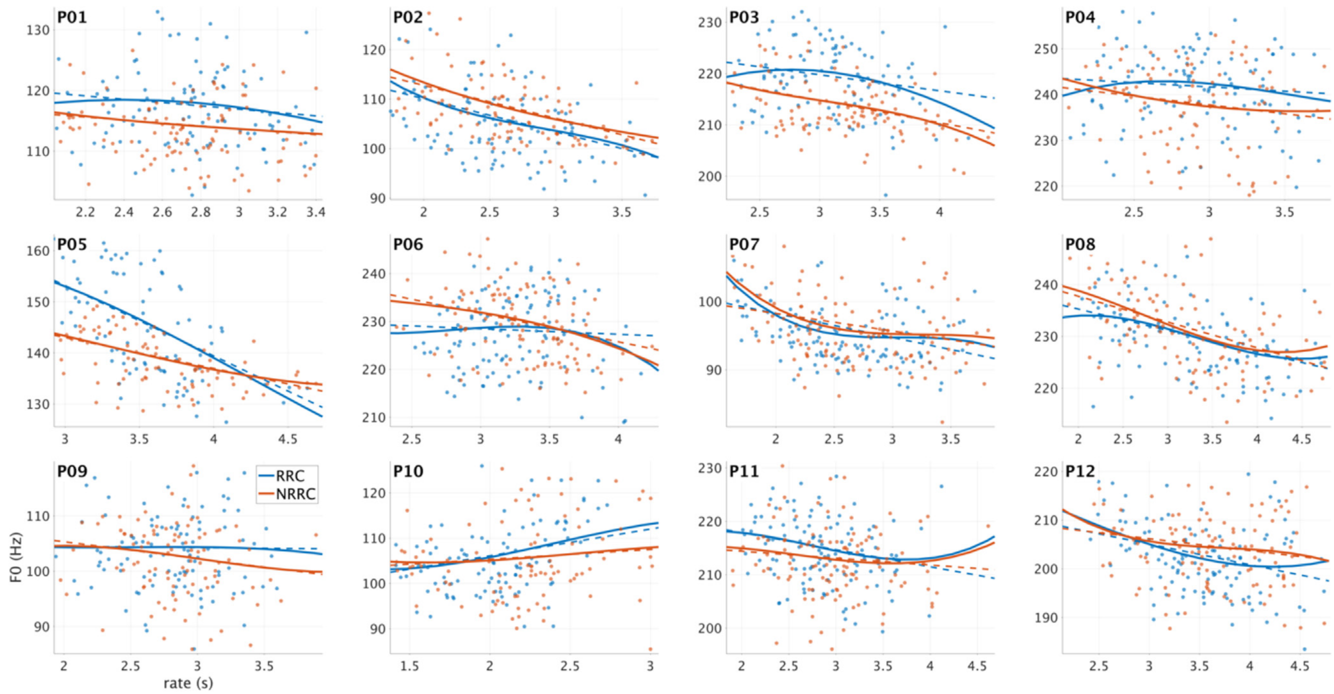

Fig. 10. Posterior predictions of the linear and non-linear models of the post-boundary F0 max at B1 in each participant. Dashed lines: linear model (RATE1), solid lines: best non-linear model. The red lines are the predictions of the NRRCs, and the blue lines are the predictions of the RRCs.

Table 4
The average of the absolute z-scored differences between the posterior predictions of NRRCs and RRCs for each variable at each boundary. The results of each boundary are presented in the order of the average z-scored difference.

pre-RC boundary (B1)		post-RC boundary (B2)	
TBI	1.27	clo-to-rel dur	0.53
coda dur	0.95	rel dur	0.45
rel dur	0.78	coda dur	0.35
clo-to-rel dur	0.51	TBI	0.32
rel amp	0.39	vel rel	0.22
dF0	0.36	vowel dur	0.21
rel vel	0.25	rel amp	0.14
vowel dur	0.23	dF0	0.14
clo amp	0.23	clo amp	0.11
pre max F0	0.15	pre max F0	0.07
post max F0	0.14	post max F0	0.06
vel clo	0.02	vel clo	0.01

measures were responsive to rate variation at the pre-RC boundary, whereas both articulatory and acoustic variables were responsive at the post-RC boundary. Among those variables that showed rate effects, articulatory and segmental variables at B2 and F0 measures at B1 tended to exhibit non-linear relations with rate, although the non-linear relations in the F0 measures were associated more strongly with participant-specific effects. We also observed scale attenuation effects in some variables at extreme rates.

4.3. Relative clause type differences

The durational measures associated with the pre-RC boundary showed substantial differences between clause types. To assess differences between syntactic structures in a way that accounts for both linear and non-linear components of the regression models, we analyzed the average z-score normalized difference between the posterior predictions of NRRCs and RRCs. The average normalized differences of each variable at each boundary are presented in Table 4. The variables that showed relatively large differences were TBI (1.27), coda duration (0.95), and release duration (0.78) at B1. Fig. 11 shows the posterior predictions and posterior prediction differences for the first two of these. Note that in both cases, the linear model was favored by the model comparison procedure (cf. Table 3).

Inspection of the posterior predictions showed that the linear rate slopes differed between clause types, such that there were larger rate effects on these variables in NRRCs than in RRCs. Furthermore, the difference between the two RCs emerged at the medial rate and increased as speech rate slowed down; the difference was absent at fast rates. For the release duration at B1, the constant model was favored by the WAIC criterion, and as in the coda and TBI, NRRCs showed longer durations than RRCs.

Interestingly, substantial differences between the two RCs were not observed in the variables associated with the post-RC boundary. Although closure-to-release duration at B2 showed a moderate difference (0.53), the average normalized differences were generally small at this boundary. Even for those variables that exhibited substantial syntactic effects at B1 – TBI, coda duration, and release duration, the normalized differences of these variables at B2 were relatively small. Thus, our data suggest that there may be a syntactically conditioned prosodic difference at the pre-RC boundary, but not at the post-RC boundary.

5. Discussion

In this study, we examined whether and how articulatory, segmental, and F0 variables associated with phrase bound-

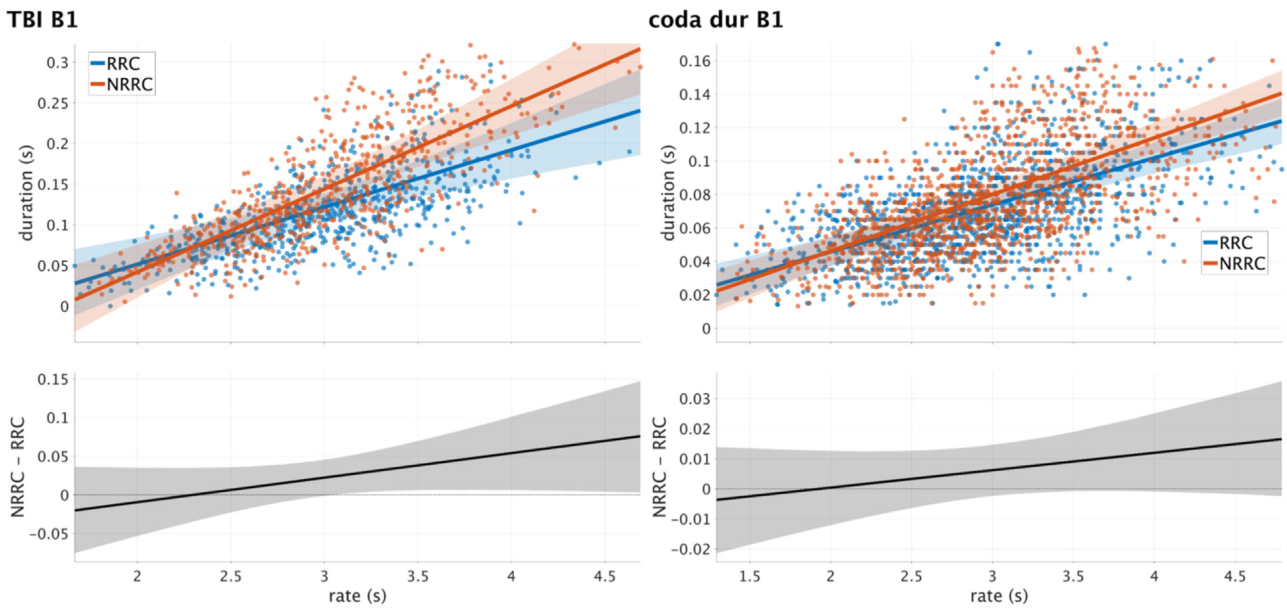

Fig. 11. Posterior predictions of the linear model of the TBI (left panel) and coda duration (right panel) at B1. In both variables, the top panel shows the mean model predictions (lines) of each syntactic structure along with the 5–95% confidence intervals (shaded area); red lines: NRRC, blue lines: RRC. The bottom panel shows the difference between the predictions and the confidence intervals.

aries vary with speech rate. Speech rate variation was elicited with a moving visual rate cue, which resulted in relatively broad, unimodal distributions of rates. Examination of the functional relations between phonetic measures and speech rate revealed some important differences between the pre-RC and post-RC boundaries. Put succinctly, the results show that while the post-RC boundary is more susceptible to variation in speech rate, the pre-RC boundary better reflects the distinction between relative clause types. We found phonetic evidence that suggests that a stronger prosodic boundary occurs at B2 than at B1: as shown in Fig. 9, the distributions of phonetic measures at B2 were broader, expanding to more extreme values, compared to those at B1. This altogether suggests that phonetic variables associated with a stronger prosodic boundary (the boundary aligned with the end of the RC) are more malleable to variation in speech rate than those associated with a weaker prosodic boundary (the boundary aligned with the start of the RC). In contrast, syntactic effects are robustly observed at the relatively weaker, pre-RC boundary.

Although our results are interpreted on the basis of gradient prosodic boundary strength, they can also be discussed in light of prosodic theories which assume categorical boundary strength. In particular, our results are more consistent with Selkirk (2005), who argued that the post-RC boundary is marked with a stronger prosodic category than the pre-RC boundary regardless of RC type (see (1) in Section 2.3 above). The results are less consistent with Nespor and Vogel (1986), who argued that the same level of prosodic boundary occurs before and after NRRCs. The results, however, do not entirely support Selkirk (2005), as substantial syntactic differences were found only at the pre-RC boundary. In contrast, Selkirk (2005) predicted that both pre-RC (NRRC: ip vs. RRC: word boundary) and post-RC (NRRC: IP vs. RRC: ip) boundaries would differ by relative clause type. Below we elaborate on these and other findings.

5.1. Speech rate variation and prosodic boundary differences

We found that the phonetic measures associated with the post-RC boundary are more responsive to rate changes than those at the pre-RC boundary. Specifically, of the 12 variables examined at each phrasal boundary, the linear model substantially outperformed the constant model for six variables at the pre-RC boundary, but for nine variables at the post-RC boundary. Further, we observed more evidence of non-linear relations at B2 than at B1; four variables exhibited non-linear relations with rate at B1, whereas six variables were found to exhibit such relations at B2. These findings are summarized in Fig. 12, where blue lines/dots indicate variables for which a linear relation was supported, and red lines/dots indicate variables for which a non-linear relation was supported. Interpreting the results with reference to prosodic boundary strength, our findings suggest that the phonetic measures associated with a stronger prosodic boundary (the boundary at the end of the RC) are more responsive to rate variation as well as more likely to exhibit non-linear relations with rate compared to those associated with a weaker prosodic boundary (the boundary at the start of the RC).

There are two subtly different interpretations of the asymmetry in rate effects between the pre-RC and post-RC boundaries. One is that stronger prosodic boundaries are more susceptible to global variation in speech rate, and therefore, phonetic variables at stronger prosodic boundaries exhibit changes with rate to a greater extent than those associated with weaker prosodic boundaries. Alternatively, we might infer that speakers manipulate articulatory timing more extensively at stronger prosodic boundaries than at weaker ones, in order to mark the strength of the boundary, and hence the asymmetry in effect derives from the local control of rate. This difference in interpretation amounts to whether the relations we observe between phonetic variables and rate are indirect

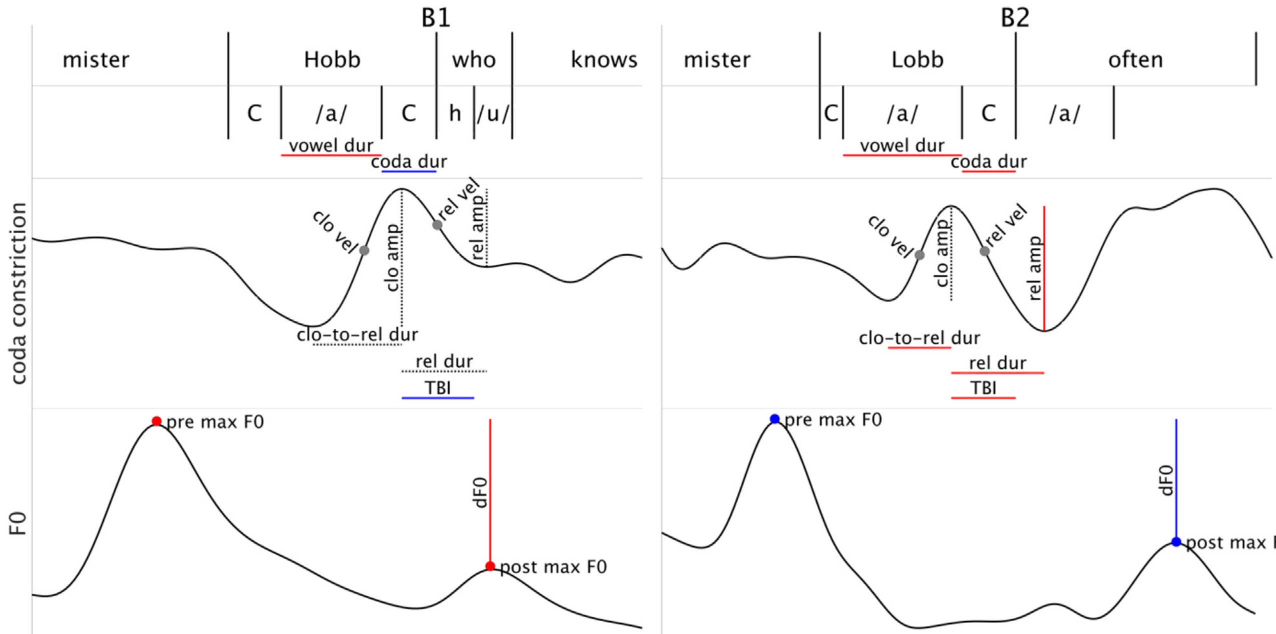

Fig. 12. Schematic representation of model comparison results. The variables that are marked in red represent cases where there was substantial evidence for non-linear relations, while the blue shows cases with evidence for linear relations. Gray dots or dashed lines mark the variables for which a constant model was supported.

consequences of differences in boundary strength or whether speakers actively use these variables to mark differences in boundary strength. Currently, we remain agnostic about these possibilities. Another interpretation of the asymmetry is that the pre-RC and post-RC boundaries have distinct communicative functions; that is, the pre-RC boundary marks the syntactic contrast, but the post-RC does not, and hence it is freer to vary with speech rate. This possibility is discussed further in Section 5.4 below.

One possible factor that is particularly relevant to the absence of rate effects in the articulatory variables at B1 involves differences in the segmental/gestural environments at the two boundaries. In the case of B2, the presence of a low vowel following the clause boundary allows for a greater range of articulator movement for the release of the coda constriction. This contrasts with B1, which is followed by labial protrusion and tongue body elevation gestures for /u/ – these gestures constrain the extent to which the lower lip and the tongue body can be lowered during the release of the pre-boundary coda, and hence may obscure the articulatory effects of rate variation. Another possible factor is that B1 occurs before B2, i.e. earlier in the utterance. This precedence confound cannot be addressed in the current design, but it may be possible to disentangle effects of boundary location and precedence with alternative sentence designs.

5.2. Non-linear relations between speech rate and phonetic variables

An important contribution of our analysis is that it revealed that some phonetic measures exhibit strong attenuation effects at extreme rates. We suspect that the attenuation effects arise because articulatory parameters reach intrinsic limits as rates become more extreme. In the case of durational measures, these limits correspond to how much an interval can expand or compress. The patterns we observed suggest that such constraints do exist, and we speculate that they derive from mechanisms of articulatory control or from listener-oriented factors such as perceptual recoverability.

Specifically, we observed a ceiling pattern in most of the temporal variables (i.e. coda duration, vowel duration, closure-to-release duration, and release duration) as well as release amplitude at B2. A ceiling pattern was also observed in the vowel duration at B1. In all these cases, the durations or amplitudes increase as speech rate is decreased, but at very slow rates, the measures attenuate, as if they have reached a ceiling. This finding is consistent with the presence of an expandability constraint on segment durations which was proposed by Cooper et al. (1985). It further shows that such constraints extend to kinematic variables such as movement amplitude, which is perhaps not surprising given the finite size of the vocal tract.

In addition, a floor-like attenuation effect was observed with increasing speech rate in the TBI at the post-RC boundary. The floor pattern of this variable suggests that there is a constraint at fast rates which imposes a maximum allowable overlap between the gestures across a boundary. This makes sense if a goal of perceptual recoverability plays a role in speech: too much overlap could compromise the ability of a listener to extract the content of the phrase-final word form. Note that we did not observe a floor-like pattern in the segmental durations. If a compressibility constraint proposed by Klatt (1973, 1976) was in effect, we might have observed a floor pattern in the coda or vowel durations as in the TBI; yet, this was not the case. Perhaps, the phrase-final environment of the segments we examined induces a lengthening that prevents those segment durations from reaching a floor at fast rates.

Besides discussing non-linearities in relation to articulatory and perceptual constraints, another way of interpreting non-linearities is that they arise from rate-dependent changes in prosodic organization. As discussed in Section 2.3, prosodic theories that assume categorical prosodic boundaries (e.g. Nespor & Vogel, 1986; Selkirk, 1986) mostly posit two phrasal categories in the prosodic hierarchy – for example, the categories ip and IP. Under these theories, it is reasonable to hypothesize that the prosodic phrasal organization of a sentence may change with speech rate, such that the boundary that is associated with a lower-level unit (e.g. ip) at fast rates may become associated with a higher-level unit (e.g. IP) at slow rates; see the post-RC boundary in example (2). A number of studies on speech rate indeed have shown that prosodic phrasing of a sentence changes with variation in speech rate as introduced in Section 2.1 (e.g. Caspers & van Heuven, 1991; Fougeron & Jun, 1998; Jun, 1993).

(2)

	B1	B2	
fast rates:	**[[A Mr. Hodd, $_{\text{ip}}$]**	**[who knows Mr. Robb, $_{\text{ip}}$]**	**[often plays tennis. $_{\text{ip}}$] $_{\text{IP}}$]**
slow rates:	**[[A Mr. Hodd, $_{\text{ip}}$]**	**[who knows Mr. Robb, $_{\text{ip}}$] $_{\text{IP}}$]**	**[[often plays tennis. $_{\text{ip}}$] $_{\text{IP}}$]**

The categorical changes in prosodic organization which are hypothesized to arise from changes in rate can give rise to mixture distributions of phonetic variables, which (depending on their parameters) can manifest as non-linear relations of the sort that we observed. However, this interpretation should be made with caution, as attenuation effects may interfere with the observation of mixtures. That is, the ceiling and floor effects we observed may arise from the presence of distinct prosodic categories but may also arise from a single prosodic category with externally derived upper and lower bounds on phonetic measures. In the absence of evidence to distinguish attenuation effects from categorical changes in prosodic organization, we draw no strong conclusions regarding those causes of the non-linearities.

Since the current study examines phonetic effects at phrase edges, our results are relevant to the π-gesture theory of Byrd and Saltzman (2003). According to this theory, the source of a local lengthening and slowing of gestures observed in the vicinity of phrase boundaries is the activation of systems (π-gestures) which slows the rate of an internal clock. Our results suggest that the π-gesture activation is influenced not only by boundary strength but also by global speech rate. Alternatively, the attentional modulation model of Tilsen (2018) holds that both global and local rate control are accomplished by adjusting the relative contributions of external sensory feedback and internal feedback, which govern the selection and suppression of articulatory gestures. In this approach, not only global speech rate but also other factors, such as the importance of a clause to a discourse, are allowed to influence the time course of gestural activation.

5.3. Participant-specific variation in the functional relations of F0

Regarding F0 variables, the relations we observed with speech rate were more participant-specific than they were for articulatory and segmental variables. The posterior predictions of the post-boundary F0 maxima at B1 in each participant (Fig. 10) showed that some participants had a floor attenuation, while others had a ceiling attenuation. Furthermore, participants differed with regard to the rates at which the attenuation of F0 maxima was observed. The variation across individuals was found in other F0 variables at the pre-RC boundary, where non-linear relations were observed as well. This suggests that it may not be appropriate to draw across-speaker inferences on the functional relations between speech rate and F0-related variables in the vicinity of phrase boundaries.

The origins of the participant-specific variation can be seen in aspects of their F0 trajectories which are not well-captured by our phonetic measures (F0 maxima and their differences). The figures below show linearly time-warped F0 contours of each participant in each syntactic context in the vicinity of the pre-RC (Fig. 13) and post-RC (Fig. 14) boundaries. They present the F0 contours at the target region which is defined from the honorific "Mr." to two syllables after the boundary (for example, in "*A Mr. Hodd, who knows Mr. Robb, often plays tennis*", the pre-RC target region is "*Mr. Hodd, who knows*", and the post-RC target region is "*Mr. Robb, often*"). Although some similarities exist, substantial variation is evident in the number, height, and timing of the peaks and valleys. For example, in the pre-RC boundary (Fig. 13), most participants showed a pattern where there is one peak before the boundary and another peak after the boundary. However, some participants showed multiple peaks at either pre- (P05, 07) or post-boundary region (P06, 08) or at both regions (P09, 11). The F0 at the peak was in general higher at the pre-boundary region, although some had similar or even higher F0 at the post-boundary region (P04, 05, 10). The timing of the boundary relative to the peaks/valleys was also different such that the boundary occurred at the F0 valley in P08 or P11, whereas the boundary followed the F0 valley in the rest of the participants.

Additional speaker-specificity was observed in the F0 contours at the post-RC boundary (Fig. 14). While some participants had one high peak at the pre-boundary region, others (P02, 05, 09, 10) had the highest F0 at the beginning of the target region. The number of peaks found within the target region also differed by participants. In addition, the timing of the boundary relative to the peaks/valleys varied such that the boundary occurred at the peak for some participants (P03, 04, 06, 09, 12), while it occurred near the valley for others (P01, 07, 10).

Differences in F0 contours between participants may arise for various reasons such as different prosodic structures, different nuclear pitch accents or boundary tones, individual differences in the extent to which speakers use F0 cues to mark boundaries, and differences in the extent to which participants "perform" the syntactic/semantic contrasts between RRC and NRRC sentences. Due to this variation, we do not believe it is appropriate to draw population-level inferences

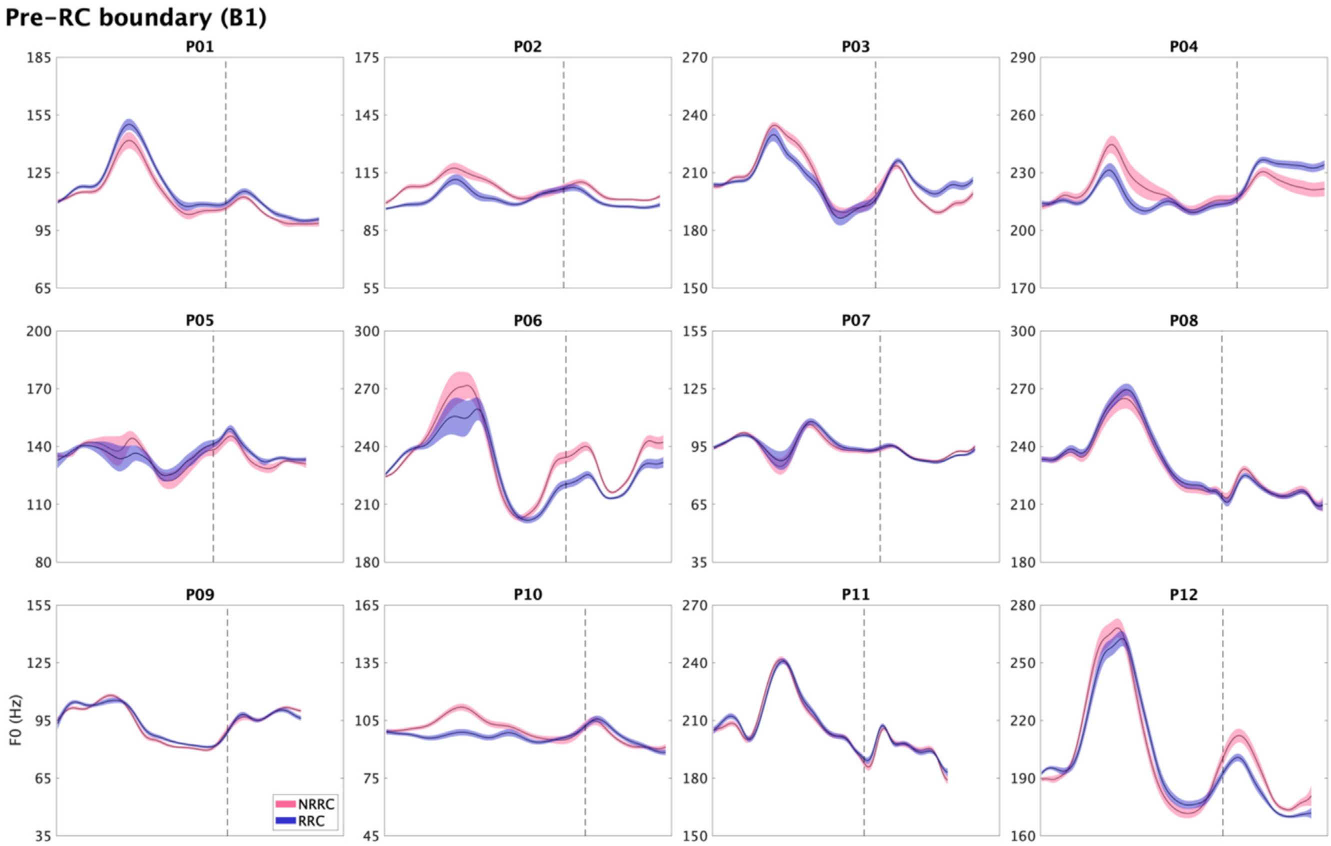

Fig. 13. Time-warped F0 contours of each participant at the pre-RC boundary. The figures show the target region at B1 (e.g. *Mr. Hodd, who knows*). All rates are combined. F0 contours are smoothed, and voiceless parts are interpolated with adjacent values; pink: NRRC, blue: RRC. The black lines show the mean F0 values, and the shades show 95% confidence intervals. The vertical dashed lines mark the ends of the target name (i.e. phrase boundary). The range of the y-axis is 120 Hz in all participants.

on the functional relations between rate and F0 variables. For datasets in which similar forms of participant-specific variation is present, it may be necessary to identify clusters (sub-populations) within which F0 patterns are more consistent; this would require a larger speaker pool than the one we examined. In addition, due to phrase-final devoicing of voiced segments and the occurrence of irregular/creaky phonation produced by some participants or at some rates, we were not able to obtain robust estimates of F0 minima near phrasal boundaries. Hence, we suggest that further investigation of the correlation between speech rate and F0 should be conducted with utterances containing a vowel or sonorant-final segment near boundaries. Moreover, such investigations must either take steps to reduce the occurrence of creaky voice or find a way to incorporate its occurrence into analyses.

5.4. Syntactically-*conditioned* prosodic differences

In the comparisons between syntactic constructions, differences between the two RCs were found in the durational measures – in particular, the TBI, coda duration, and release duration – at the pre-RC boundary. In addition, all of the aforementioned variables showed larger rate effects in NRRCs than in RRCs. A specific examination of the posterior predictions of the TBI and coda durations found that the difference between the two RCs emerged at the medial rate and increased as speech rate is decreased. Hence, the durational contrasts between the two RCs might be described as “neutralized” at fast rates. Under categorical views of prosodic organization, this would mean that at fast rates, the pre-RC boundary is associated with the same type of prosodic unit in both RCs. So, it could be inferred that the prosodic organization changes with rate, particularly in NRRCs. On the other hand, under gradient views of prosodic organization, the neutralization may arise simply from the fact that the relations between articulatory speech rate and articulatory control parameters at very fast rates are the same for both types of RCs.

An interesting dissociation in our results is that while more phonetic variables varied with rate and showed non-linear relations at B2 (Section 5.1), those associated with B1 exhibited stronger differences in posterior predictions by RC type. The results from the syntactic analysis are summarized in Fig. 15, which shows the magnitudes of the differences in posterior predictions between the two RCs: the differences are generally larger at B1. Recall that Selkirk (2005) maintained that NRRCs and RRCs differ in categorical prosodic boundary strength at both pre-RC and post-RC boundaries. At the pre-RC boundary, NRRC is marked with an ip, whereas RRC is marked with a word boundary; at the post-RC boundary, NRRC is associated with an IP, but RRC is associated with an ip. From the gradient view of prosodic structure, we would have observed a significant difference in the functional relations with rate by RC type at both phrase boundaries, if Selkirk (2005)’s proposal was correct; yet, the difference was not very substantial in the phonetic measures associated with the post-RC boundary.

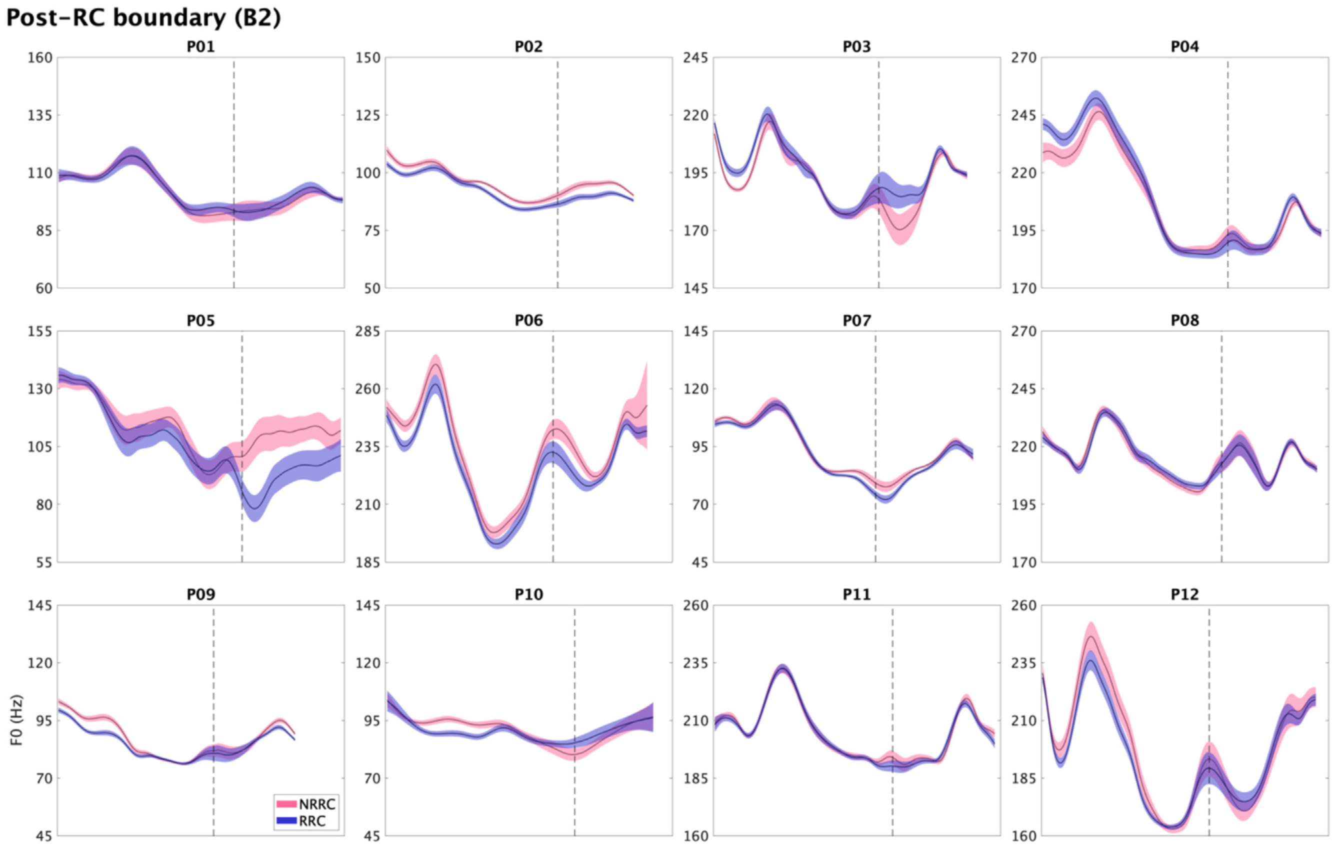

Fig. 14. Time-warped F0 contours of each participant at the post-RC boundary. The figures show the target region at B2 (e.g. *Mr. Robb, often*). The range of the y-axis is 100 Hz in all participants. The rest of the information is identical to Fig. 13.

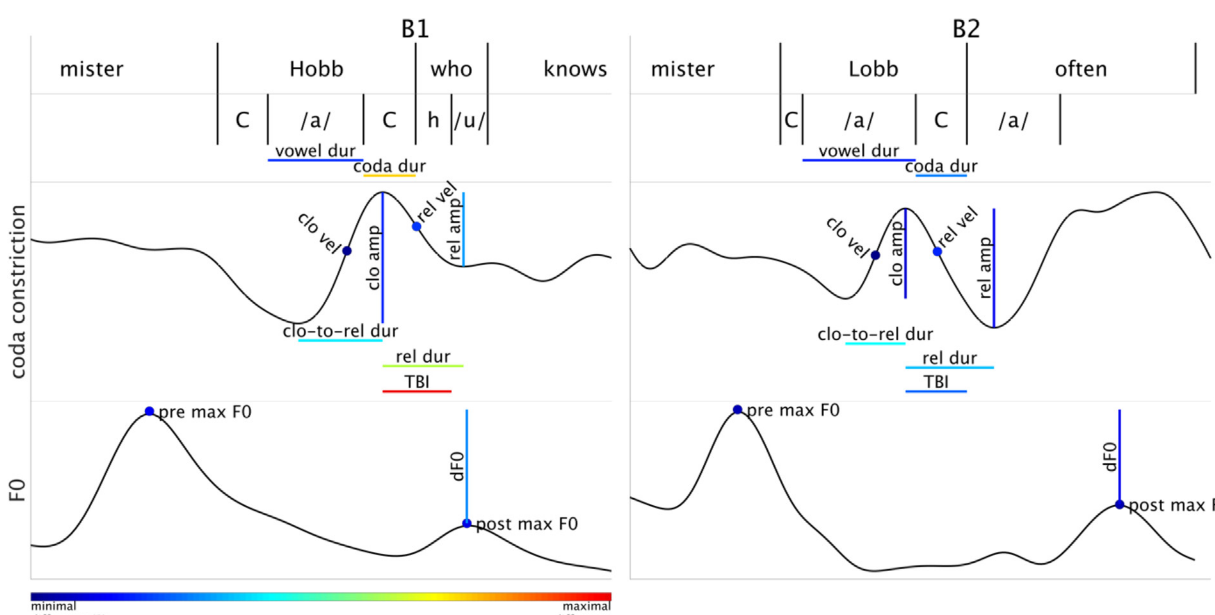

Fig. 15. Schematic representation of the magnitudes of relative clause type effects. The colors of the markers and lines correspond to effect magnitudes (blue is smaller, red is larger). The minimal difference in the range is 0, while the maximal difference is the largest difference found in our data (i.e. the difference of TBI at B1).

One possible interpretation of the weak syntactic effects at the post-RC boundary is that the two boundaries have distinct communicative functions. Perhaps, the pre-RC boundary cues listeners about the syntactic structure and disambiguates the semantic relations between the RC and antecedent in a way that the post-RC boundary does not. Hence, the pre-RC boundary is more likely to show evidence of a syntactic contrast. This also relates to the finding that rate variation is more apparent in the phonetic variables associated with the post-RC boundary (Section 5.1). Since B1 plays the important role of disambiguating syntactic/semantic possibilities, it must be kept less susceptible to rate changes; conversely, because B2 is less important for ambiguity resolution, it is freer to vary with rate. Caution is warranted in this interpretation as the results of the current study may depend on an exaggeration of the contrast between the stimuli of the experiment or on the repetition of structurally similar sentences. A related possibility is that at the post-RC boundary, the goal of marking a relatively strong prosodic break within each utterance outweighs the syntactically conditioned difference in prosodic boundary strength. Yet, this interpretation requires an assumption that participants may have difficulty producing two types of contrasts – B1 vs. B2 and NRRC vs. RRC – simultaneously. Further investigation with more participants and greater variation in experimental stimuli might be conducted to explore these possibilities.

6. Conclusion

This study examined how speech rate is correlated with articulatory and acoustic variables associated with phrase boundaries in sentences that contain non-restrictive and restrictive relative clauses. By conducting Bayesian regressions with a variety of models, we found that phonetic measures are more responsive to rate changes at the post-RC boundary than the pre-RC boundary. Phonetic variables at the post-RC boundary also tended to exhibit non-linear relations with rate, showing ceiling or floor attenuation effects at extreme rates. These results suggest that phonetic variables at stronger prosodic boundaries are more susceptible to variation in rate than those at weaker prosodic boundaries. In contrast, examinations of posterior predictions by RC type found that syntactic differences were more substantial in the variables associated with the relatively weaker, pre-RC boundary.

One important aspect of our study is that we treated speech rate as a continuous, empirically observed variable. While many previous studies have elicited speech rate with qualitative instructions and have treated rate as a categorical variable, we used a visual rate cue to elicit variation in rate and adopted an empirical measure to reflect its continuous nature. Through the novel experimental design, we elicited a wide range of rates, and this allowed us to examine functional relations between rate and various phonetic measures with more precision than in prior studies. Another important methodological aspect of our study is that we adopted a Bayesian statistical approach in conducting analyses. Since our goal was to explore the nature of functional relations between various phonetic measures and speech rate, Bayesian inference was more suitable than a null hypothesis significance testing approach.

The results of the study point to several future directions of investigation. First, the finding that phonetic variables are more susceptible to rate variation at stronger prosodic boundaries is important, because it is not fully consistent with the common idea that stronger prosodic boundaries are associated with longer durations, larger movements, and more extreme pitch excursions. In fact, our results indicate a more nuanced relation in which stronger prosodic boundaries change more with rate than weaker prosodic boundaries – this involves not only durational lengthening at slow rates but also durational compression at fast rates. This phenomenon warrants further investigation, ideally with more exact controls over segmental environments.

Second, although our study achieves a fairly detailed characterization of non-linear functional relations between rate and phonetic variables, we propose that there is room for further improvement. We considered quadratic and cubic models in the current analyses, yet fitting other non-linear models, for instance, exponential, logarithmic, or sigmoidal models, or smoothing splines may provide better characterizations of the functional relations between phonetic measures and rate. In order for this endeavor to be successful, greater statistical power may be necessary – i.e. more observations must be made within and/or across speakers.

Third, our analyses suggest that F0-related variables cannot be treated in the same way as articulatory and segmental variables. F0 patterns differed more substantially across speakers, and we expect this to be the case for any experiment in which complex syntactic environments are analyzed. To address this, a much larger speaker pool must be examined, and perhaps clustering analyses can be used to identify subpopulations of participants with consistent F0 patterns.

Finally, we suggest that it may be useful to consider different rate elicitation methods or adopt other empirical measures of rate in future analyses. In the current study, we used 10 different cue rates, and they were randomized from trial to trial. With different numbers of cue rates, one may obtain productions of more extreme rates or more fine-grained variations of rate. In addition, we may observe different distributions of rates if cue rates were not randomized but rather presented sequentially. We thus believe it would be valuable to employ different experiment designs to gain a better understanding of the nature of speech rate. Moreover, instead of using a durational measure of empirical rate, we might ask whether a measure such as sentences per second or syllables per second – which are inversely related to our rate measure – results in qualitatively similar findings. Furthermore, we used the most global rate measure in the current study; yet, there may be some value to investigating whether the functional relations between local rate and various phonetic measures differ from our results in any way. This would also inform us on the relations between global and local rate modulations.

The issue of how to best characterize and measure "speech rate" has not been systematically addressed in phonetic studies, despite the widespread usage of speech rate as a covariate. Our study constitutes a first step in this direction and in particular indicates that we cannot think of speech rate monolithically as a linear covariate of any given phonetic measure. Instead, relations between rate and phonetic variables depend on the nature of the variables examined, are contingent on the interaction between syntactic and prosodic structures, and can be sensitive to ceiling and floor effects.

Conflict of interest statement

The authors have no conflicts of interest to declare.

CRediT authorship contribution statement

Seung-Eun Kim: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Sam Tilsen:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision.

Acknowledgements

We would like to thank Draga Zec, Abby Cohn, and members of the Cornell Phonetics Lab for discussions of this research. We also thank Bruce McKee for technical support in conducting the experiments.

Appendix. Bayesian generalized linear model regressions

Bayesian regressions were conducted in R (R Core Team, 2020) using the Hamiltonian Markov Chain Monte Carlo (MCMC) sampler in the RStan package (Stan Development Team, 2020). The independent variable of speech rate (RATE) and all dependent variables were z-score normalized for the regressions. For each dependent variable, posterior parameter distributions were estimated for four generalized linear models: one with intercepts and linear, quadratic, and cubic rate effects (RATE3); one with intercepts, linear, and quadratic rate effects (RATE2); one with intercepts and linear rate effects (RATE1); and one with only intercepts (RATE0). The intercepts and slopes of all models included a random subject-by-relative clause effect and a fixed effect of syntactic structure. This means that all terms of polynomials included an interaction between speech rate and clause type. For articulatory and segmental variables, a fixed effect of the place of articulation of the target word coda (labial/coronal) was also included in the intercepts. The place effect was excluded from regressions of all F0-related variables because we have no a priori reason to believe that the coda place will have an influence on those variables. Following the guidelines in McElreath (2020b), we used index variables for subjects (SU), relative clause type (RC), and place of articulation of the target word coda (PL).

To obtain posterior samples, four MCMC chains were run with 6000 iterations per chain, 1000 of which were warmup iterations. To facilitate more efficient sampling of the posterior distribution, the random intercepts and slopes were non-centered. Comparisons of the RATE3, RATE2, RATE1, and RATE0 models were conducted using Widely Applicable Information Criterion (WAIC, Watanabe, 2013), calculated using the *rethinking* package (McElreath, 2020a). The WAIC is pointwise approximation of out-of-sample deviance calculated from the posterior. It guards against overfitting by incorporating a penalty term based on the sum of the variance of the log posterior probabilities of each observation (McElreath, 2020b, pp. 223–225). The stan code for the maximal model (RATE3) is shown below.

```
data{
  int<lower = 1> N; //Number of observations
  int<lower = 1> S; //Number of subjects
  int<lower = 1> C; //Number of relative clause types
  int<lower = 1> P; //Number of places
  vector[N] y;    //vector of observations
  int<lower = 1,upper = S> su[N];    //Vector of subject indices
  int<lower = 1,upper = C> rc[N];    //Vector of clause type indices
  int<lower = 1,upper = P> pl[N];    //Vector of place indices
  vector[N] rate3; //vector of rates^3
  vector[N] rate2; //vector of rates^2
  vector[N] rate1; //vector of rates
}
parameters{
  matrix[S,C] b3s_raw;
  matrix[S,C] b2s_raw;
  matrix[S,C] b1s_raw;
  matrix[S,C] b0s_raw;
  vector[C] b0;
  vector[C] b1;
  vector[C] b2;
  vector[C] b3;
  vector[P] c;
  real<lower = 0> b0s_sigma;
  real<lower = 0> b1s_sigma;
  real<lower = 0> b2s_sigma;
  real<lower = 0> b3s_sigma;
  real<lower = 0> sigma;
}
transformed parameters{
  matrix[S,C] b0s;
  matrix[S,C] b1s;
  matrix[S,C] b2s;
  matrix[S,C] b3s;
  b0s = b0s_sigma * b0s_raw;
  b1s = b1s_sigma * b1s_raw;
  b2s = b2s_sigma * b2s_raw;
  b3s = b3s_sigma * b3s_raw;
}
model{
  vector[N] MU;
  vector[N] B0;
  vector[N] B1;
  vector[N] B2;
  vector[N] B3;
  sigma ~ exponential(1);
  c ~ normal(0, 0.5);
  b3 ~ normal(0, 0.1);
  b2 ~ normal(0, 0.1);
  b1 ~ normal(0, 0.5);
  b0 ~ normal(0, 0.5);
  b3s_sigma ~ exponential(0.5);
  b2s_sigma ~ exponential(0.5);
  b1s_sigma ~ exponential(1);
  b0s_sigma ~ exponential(1);
  for (i in 1:S) {
    for (j in 1:C) {
      b3s_raw[i,j] ~ std_normal();
      b2s_raw[i,j] ~ std_normal();
      b1s_raw[i,j] ~ std_normal();
      b0s_raw[i,j] ~ std_normal();
    }
  }
  for (i in 1:N) {
    B3[i] = b3s[su[i],rc[i]] + b3[rc[i]];
    B2[i] = b2s[su[i],rc[i]] + b2[rc[i]];
    B1[i] = b1s[su[i],rc[i]] + b1[rc[i]];
    B0[i] = b0s[su[i],rc[i]] + b0[rc[i]] + c[pl[i]];
    MU[i] = B0[i] + B1[i] * rate1[i] + B2[i] * rate2[i] + B3[i] * rate3[i];
  }
  y ~ normal(MU , sigma);
```

(*continued on next page*)

```
}
generated quantities{
  vector[N] log_lik;
  vector[N] MU;
  vector[N] B0;
  vector[N] B1;
  vector[N] B2;
  vector[N] B3;
  for (i in 1:N) {
    B3[i] = b3s[su[i],rc[i]] + b3[rc[i]];
    B2[i] = b2s[su[i],rc[i]] + b2[rc[i]];
    B1[i] = b1s[su[i],rc[i]] + b1[rc[i]];
    B0[i] = b0s[su[i],rc[i]] + b0[rc[i]] + c[pl[i]];
    MU[i] = B0[i] + B1[i] * rate1[i] + B2[i] * rate2[i] + B3[i] * rate3[i];
    log_lik[i] = normal_lpdf(y[i] | MU[i] , sigma);
  }
}
```

References

Abbs, J. H. (1973). The influence of the gamma motor system on jaw movements during speech: A theoretical framework and some preliminary observations. *Journal of Speech and Hearing Research, 16*(2), 175–200. https://doi.org/10.1044/jshr.1602.175.

Adams, S. G., Weismer, G., & Kent, R. D. (1993). Speaking rate and speech movement velocity profiles. *Journal of Speech, Language, and Hearing Research, 36*(1), 41–54. https://doi.org/10.1044/jshr.3601.41.

Arnold, D. (2007). Non-restrictive relatives are not orphans. *Journal of Linguistics, 43*(2), 271–309.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3*, 255–309. https://doi.org/10.1017/S095267570000066X.

Berkovits, R. (1991). The effect of speaking rate on evidence for utterance-final lengthening. *Phonetica, 48*(1), 57–66. https://doi.org/10.1159/000261871.

Byrd, D. (2006). Relating prosody and dynamic events: Commentary on the papers by Cho, Navas, and smiljanić. In L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory phonology 8* (pp. 549–561). Mouton De Gruyter.

Byrd, D., & Krivokapić, J. (2021). Cracking prosody in articulatory phonology. *Annual Review of Linguistics, 7*, 31–53. https://doi.org/10.1146/annurev-linguistics-030920-050033.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics, 31*(2), 149–180.

Byrd, D., & Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics, 24*(2), 263–282. https://doi.org/10.1006/jpho.1996.0014.

Caspers, J., & van Heuven, V. J. (1991). Phonetic and linguistic aspects of pitch movements in fast speech in Dutch. *Proceedings of the 12th International Congress of Phonetic Sciences*, *5*, 174–177.

Caspers, J., & van Heuven, V. J. (1993). Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica, 50*(3), 161–171. https://doi.org/10.1159/000261936.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America, 77*(6), 2142–2156. https://doi.org/10.1121/1.392372.

Cummins, F. (1999). Some lengthening factors in English speech combine additively at most rates. *The Journal of the Acoustical Society of America, 105*(1), 476–480. https://doi.org/10.1121/1.424576.

de Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *The Journal of the Acoustical Society of America, 96*(4), 2037–2047. https://doi.org/10.1121/1.410145.

Emonds, J. (1979). Appositive relatives have no properties. *Linguistic Inquiry, 10*(2), 211–243.

Engstrand, O. (1988). Articulatory correlates of stress and speaking rate in Swedish VCV utterances. *The Journal of the Acoustical Society of America, 83*(5), 1863–1875. https://doi.org/10.1121/1.396522.

Fabb, N. (1990). The difference between English restrictive and nonrestrictive relative clauses. *Journal of Linguistics, 26*(1), 57–77.

Fougeron, C., & Jun, S.-A. (1998). Rate effects on French intonation: Prosodic organization and phonetic realization. *Journal of Phonetics, 26*(1), 45–69. https://doi.org/10.1006/jpho.1997.0062.

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica, 38*, 148–158. https://doi.org/10.1159/000260020.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York, NY: Academic Press.

Grosjean, F. (1979). A study of timing in a manual and a spoken language: American sign language and English. *Journal of Psycholinguistic Research, 8*(4), 379–405. https://doi.org/10.1007/BF01067141.

Hermes, A., Tilsen, S., & Ridouane, R. (2021). Cross-linguistic timing contrast in geminates: A rate-independent perspective. In *Proceedings of the 12th international seminar on speech production* (pp. 52–55).

Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody [PhD Thesis]*. The Ohio State University.

Kendall, T. (2013). *Speech rate, pause and sociolinguistic variation*. London, UK: Palgrave Macmillan. https://doi.org/10.1057/9781137291448.

Kent, R. D., & Moll, K. L. (1972). Cinefluographic analyses of selected lingual consonants. *Journal of Speech and Hearing Research, 15*, 453–473.

Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America, 54*(4), 1102–1104.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America, 59*(5), 1208–1221. https://doi.org/10.1121/1.380986.

Krakow, R. A. (1993). Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate. In M. K. Huffman & R. A. Krakow (Eds.), *Nasals, nasalization, and the velum* (pp. 87–116). Academic Press. https://doi.org/10.1016/B978-0-12-360380-7.50008-9.

Krivokapić, J. (2007). *The planning, production, and perception of prosodic structure [PhD Thesis]*. University of Southern California.

Krivokapić, J., & Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics, 40*(3), 430–442.

Kuehn, D. P., & Moll, K. L. (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics, 4*(4), 303–320. https://doi.org/10.1016/S0095-4470(19)31257-4.

Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (1999). Constant "segmental anchoring" of F0 movements under changes in speech rate. *The Journal of the Acoustical Society of America, 106*(3), 1543–1554. https://doi.org/10.1121/1.427151.

Lane, H., & Grosjean, F. (1973). Perception of reading rate by speakers and listeners. *Journal of Experimental Psychology, 97*(2), 141–147. https://doi.org/10.1037/h0033869.

Lindblom, B. E. F. (1964). Articulatory activity in vowels. *STL-QPSR, 5*(2), 1–5.

McCawley, J. D. (1982). Parentheticals and discontinuous constituent structure. *Linguistic Inquiry, 13*(1), 91–106.

McElreath, R. (2020a). *rethinking: Statistical Rethinking book package* (2.13) [Computer software].

McElreath, R. (2020b). *Statistical rethinking: A Bayesian course with examples in R and STAN*. CRC Press https://www.routledge.com/Statistical-Rethinking-A-Bayesian-Course-with-Examples-in-R-and-STAN/McElreath/p/book/9780367139919.

Michelas, A., & D'Imperio, M. (2012). When syntax meets prosody: Tonal and duration variability in French Accentual Phrases. *Journal of Phonetics, 40*(6), 816–829. https://doi.org/10.1016/j.wocn.2012.08.004.

Munhall, K., & Löfqvist, A. (1992). Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics, 20*(1), 111–126. https://doi.org/10.1016/S0095-4470(19)30242-6.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Nicenboim, B., Roettger, T. B., & Vasishth, S. (2018). Using meta-analysis for evidence synthesis: The case of incomplete neutralization in German. *Journal of Phonetics, 70*, 39–55. https://doi.org/10.1016/j.wocn.2018.06.001.

Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and confirmatory analyses in sentence processing: A case study of number interference in German. *Cognitive Science, 42*, 1075–1100. https://doi.org/10.1111/cogs.12589.

Ostry, D. J., & Munhall, K. G. (1985). Control of rate and duration of speech movements. *The Journal of the Acoustical Society of America, 77*(2), 640–648. https://doi.org/10.1121/1.391882.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation [PhD Thesis]*. Massachusetts Institute of Technology.

Pouplier, M., Marin, S., Hoole, P., & Kochetov, A. (2017). Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics, 64*, 108–126. https://doi.org/10.1016/j.wocn.2017.01.006.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., ... Vesely, K. (2011). The Kaldi speech recognition toolkit. *IEEE 2011 workshop on automatic speech recognition and understanding*.

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/.

Roettger, T. B., Winter, B., & Baayen, H. (2019). Emergent data analysis in phonetic sciences: Towards pluralism and reproducibility. *Journal of Phonetics, 73*, 1–7.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook, 3*, 371–405. https://doi.org/10.1017/S0952675700000695.

Selkirk, E. (2005). Comments on intonational phrasing in English. In S. Frota, M. Vigário, & M. J. Freitas (Eds.), *Prosodies: With special reference to Iberian languages* (pp. 11–58).

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193–247. https://doi.org/10.1007/BF01708572.

Stan Development Team. (2020). *RStan: The R interface to Stan* (2.21.2) [Computer software]. http://mc-stan.org/.

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

Tiede, M., Espy-Wilson, C. Y., Goldenberg, D., Mitra, V., Nam, H., & Sivaraman, G. (2017). Quantifying kinematic aspects of reduction in a contrasting rate production task. *The Journal of the Acoustical Society of America, 141*(5), 3580. https://doi.org/10.1121/1.4987629.

Tilsen, S. (2018). Three mechanisms for modeling articulation: Selection, coordination, and intention. *Cornell Working Papers in Phonetics and Phonology 2018*.

Tilsen, S., & Hermes, A. (2021). Nonlinear effects of speech rate on articulatory timing in singletons and geminates. In *Proceedings of the 12th international seminar on speech production* (pp. 56–59).

Trouvain, J., & Grice, M. (1999). The effect of tempo on prosodic structure. *Proceedings of the 14th international congress of phonetic sciences*, *2*, 1067–1070.

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics, 71*, 147–161. https://doi.org/10.1016/j.wocn.2018.07.008.

Wagner, M. (2005). *Prosody and recursion [PhD Thesis]*. Massachusetts Institute of Technology.

Wagner, M. (2010). Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory, 28*(1), 183–237. https://doi.org/10.1007/s11049-009-9086-0.

Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research, 14*, 867–897.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America, 91*(3), 1707–1717.