

## Research Article

# Pause postures: The relationship between articulation and cognitive processes during pauses



Jelena Krivokapić<sup>a,b,\*</sup>, Will Styler<sup>c</sup>, Benjamin Parrell<sup>d</sup>

<sup>a</sup> University of Michigan, Department of Linguistics, 421 Lorch Hall, 611 Tappan Street, Ann Arbor, MI 48109-1220, United States

<sup>b</sup> Haskins Laboratories, 300 George St 9th Fl, New Haven, CT 06511-6624, United States

<sup>c</sup> University of California, San Diego Department of Linguistics, 9500 Gilman Drive #0108, La Jolla, CA 92093-0108, United States

<sup>d</sup> University of Wisconsin-Madison, Department of Communication Sciences and Disorders, Goodnight Hall, 1975 Willow Drive, Madison, WI 53706, United States

## ARTICLE INFO

## Article history:

Received 21 January 2019

Received in revised form 10 December 2019

Accepted 17 December 2019

Available online 21 February 2020

## Keywords:

Articulatory settings

Pause postures

Pauses

Speech planning

Prosodic boundaries

Speech production

## ABSTRACT

Studies examining articulatory characteristics of pauses have identified language-specific postures of the vocal tract in inter-utterance pauses and different articulatory patterns in grammatical and non-grammatical pauses. Pause postures—specific articulatory movements that occur during pauses at strong prosodic boundaries—have been identified for Greek and German. However, the cognitive function of these articulations has not been examined so far. We start addressing this question by investigating the effect of (1) utterance type and (2) planning on pause posture occurrence and properties in American English. We first examine whether pause postures exist in American English. In an electromagnetic articulometry study, seven participants produced sentences varying in linguistic structure (stress, boundary, sentence type). To determine the presence of pause postures, as well as to lay the groundwork for their future automatic annotation and detection, a Support Vector Machine Classifier was built to identify pause postures. Results show that pause postures exist for all speakers in this study but that the frequency of occurrence is speaker dependent. Across participants, we find that there is a stable relationship between the pause posture and other events (boundary tones and vowels) at prosodic boundaries, parallel to previous work in Greek. We find that the occurrence of pause postures is not systematically related to utterance type. Lastly, pause postures increase in frequency and duration as utterance length increases, suggesting that pause postures are at least partially related to speech planning processes.

© 2019 Elsevier Ltd. All rights reserved.

## 1. Introduction

A long line of research has examined acoustic pauses during connected speech, which can be grouped into grammatical and non-grammatical pauses. Grammatical pauses are a part of prosodic boundaries, which are planned events that indicate linguistic structure. Non-grammatical pauses, on the other hand, are not planned events and are, broadly, the result of speech planning processes (e.g., pauses that are related to the time a speaker needs to plan an upcoming word or utterance, or filled pauses, such as *uh* and *uhm*). Recent work has started examining articulations during pauses (e.g., Gick, Wilson, Koch, & Cook, 2004; Katsika, Krivokapić, Mooshammer, Tiede, & Goldstein, 2014; Ramanarayanan,

Goldstein, Byrd, & Narayanan, 2013; Rasskazova, Mooshammer, & Fuchs, 2018) and has identified various articulatory patterns during pauses. These seem to have language-specific characteristics, but also exhibit large variability within languages in terms of speaker and context, and crucially depend on the type of pause (grammatical vs. non-grammatical).

A major unanswered question in this research is what the cognitive function of these patterns is. The current study addresses this question by examining the production of pause postures at prosodic boundaries. Pause postures are specific movements of articulators during acoustic pauses (described in detail in sections 1.2 and 1.3). Four questions are addressed in this study:

- (1) Do pause postures exist in American English? We examine whether specific movement patterns, as have been identified for Greek during pauses, termed pause postures, also exist

\* Corresponding author.

E-mail address: [jelenak@umich.edu](mailto:jelenak@umich.edu) (J. Krivokapić).

in American English. Foreshadowing the results for this question, we indeed find evidence of pause postures in American English.

- (2) We examine how pause postures are timed relative to other gestures at the boundary. Evidence of systematic timing patterns with other gestures at the boundary would provide further evidence of pause postures as cognitive units.
- (3) The main focus of our study examines what the cognitive processes underlying these pause postures are. Specifically, we examine if their occurrence and articulatory properties are related to utterance type and to speech planning processes.
- (4) Finally, we address a methodological question, namely, can pause postures be detected automatically using machine learning models? This provides both a method for future annotation, as well as secondary validation of the presence and detectability of this phenomenon.

### 1.1. Cognitive processes related to pauses during speech

Studies of grammatical pauses have established a number of factors determining the likelihood of occurrence and length of a pause (see overview in Fletcher, 2010; Fuchs, Petrone, Krivokapić, & Hoole, 2013). For example, faster speech typically leads to shorter and fewer pauses (e.g., Fletcher, 1987; Goldman Eisler, 1968; Lane & Grosjean, 1973). In read speech, but not in spontaneous speech, pauses occur only at syntactic boundaries (Goldman Eisler, 1968).

Discourse content has also been shown to affect pausing in speech. Analyses based on theoretical approaches to discourse, regardless of the specific theory, consistently show that hierarchically higher discourse boundaries are associated with longer pauses (Den Ouden, Noordman, & Terken, 2009; Hirschberg & Nakatani, 1996; Tyler, 2013; Yang, Xu, & Yang, 2014). Other studies have examined how a change in topic affects pause duration. A robust finding in these studies is that topic change has an effect on pause duration, such that topic shift between utterances leads to longer pauses than topic continuation (Bannert, Botinis, Gawronska, Katsika, & Sandblom, 2003; Smith, 2004; Swerts & Geluykens, 1994; Yang et al., 2014), though there is some evidence that this could be dependent on speaking style (Gustafson-Čapková & Megyesi, 2002).

The occurrence and duration of pauses in general is further influenced by a number of other structural factors. The more complex the linguistic (syntactic or prosodic) structure preceding or following a boundary, the likelier a pause is to occur and to be longer in duration (e.g., (Cooper & Paccia-Cooper, 1980; Ferreira, 1991, 1993; Grosjean, Grosjean, & Lane, 1979; Sanderman & Collier, 1995; Watson & Gibson, 2004)). Similarly, the longer the preceding or following utterance (in terms of feet, syllables, or phonological words), the likelier a pause is to occur and to be longer in duration (e.g., Ferreira, 1991; Fuchs et al., 2013; Kentner, 2007; Krivokapić 2007a, 2007b; Sternberg, Monsell, Knoll, & Wright, 1978; Wheeldon & Lahiri, 1997; Zvonik & Cummins, 2003), though the strength of the effect of each of these factors is not well understood (see Krivokapić, 2007a; Yang et al., 2014). While the relationship between prosodic boundary strength and pause duration has been examined only in a few studies, there is evidence that the stronger the boundary, the more likely a pause is to occur and to increase in length (Choi, 2003; Ferreira, 1993;

Gollrad, 2013; Horne, Strangert, & Heldner, 1995; Petrone et al., 2017; Strangert, 1991; Zellner, 1994).<sup>1</sup>

Pauses at prosodic boundaries (the type we are examining here) are grammatical pauses but they have been argued to have multiple functions (see for an overview Ferreira, 2007). Specifically, pauses are one of the phonetic markers of prosodic boundaries (the structural function) but processing of preceding and following utterances is also known to take place during prosodic boundary pauses. The effect of the material preceding the pause has been argued to be related either to the time for a listener to process information, or to the time for the speaker to deactivate information processed in the preceding phrase, though of course it could be related to both (e.g., Krivokapić, 2007a; Watson & Gibson, 2004), while the effect of the upcoming material is related to the planning of the upcoming utterance (e.g., Gee & Grosjean, 1983; Krivokapić 2007a, 2012; Watson & Gibson, 2004). The idea behind this is that more upcoming structural units (syntactic, phonological, prosodic) will lead to longer pauses, because cognitive load increases with the number of units to be processed and the longer pauses allow for more time for the speaker to plan the upcoming utterance (e.g., Butcher, 1981; Cooper & Paccia-Cooper, 1980; Ferreira 1991, 1993; Fuchs et al., 2013; Goldman Eisler, 1968; Grosjean et al., 1979; Krivokapić, 2007a, 2012; Levell, 1989; Watson & Gibson, 2004).

The present study builds on the findings reviewed in this section to examine if pause posture occurrence and properties are related to discourse structure and speech planning as has been shown for pause acoustics.

### 1.2. Articulatory behavior during pauses

Articulation during pauses has only been investigated recently, as technological advances have allowed scientists to do so. It has been long postulated that during pauses, the vocal tract assumes a default position, i.e., an “articulatory setting” (Honikman, 1964; Jenner, 2001; Laver, 1978), and the first observations of articulatory settings in kinematic data come from Öhman (1967) and Perkell (1969). Gick et al. (2004) were the first to systematically investigate these settings with the goal of understanding the phonological status of articulatory settings. For read speech in French and English, with five speakers from each language, they examined seven articulatory parameters (pharynx width, velopharyngeal port width, tongue body to palate distance, tongue tip to alveolar ridge distance, jaw aperture, upper lip protrusion, and lower lip protrusion) during pauses, at a point in time after articulators had stopped moving for the preceding utterance and before they started moving for the upcoming utterance. They found that English and French speakers differ in four of these parameters: upper and lower lip protrusion, pharynx width, tongue tip-to-alveolar ridge distance, tongue body-to-palate distance. These differences indicate that articulatory settings are at least partially language specific. They further argued, based on the

<sup>1</sup> Note that some of these studies (e.g., Horne et al., 1995; Sternberg et al., 1978) investigate syntactic structure, and others prosodic structure; given the evidence showing the dependence of pause duration on prosodic rather than syntactic structure (Grosjean et al., 1979; Ferreira, 1993), we assume here that the syntactic boundaries under investigation were in fact prosodic boundaries as well.

spatial stability of five of the vocal tract parameters, that the differences between languages in articulation during pauses were caused by targeted, language-specific articulatory settings, and as such might be part of the phonological and phonetic inventory of the language. Further evidence for this argument comes from [Wilson and Gick \(2014\)](#) who in a study of eight English-French bilinguals found that bilinguals who are perceived as native-like in both languages, but not those who aren't, used distinct articulatory settings for the different languages.

In read speech, all pauses are typically planned, in the sense that they are structurally determined and encode prosodic structure. However, in spontaneous speech differences may exist between different types of pauses. Thus we can distinguish between planned vs. unplanned pauses, where unplanned, or non-grammatical pauses are pauses specifically introduced to allow speakers more time to plan the upcoming chunk of speech, whether to find an appropriate word or for purposes of structural encoding. It should be clarified that we are discussing here two types of planning: In one case we are talking about planning in the sense of linguistic (in this case prosodic) structure encoding, and in the other, we mean planning in the sense of planning an upcoming chunk of speech. Note also that we assume that in both types of pauses, planning of the upcoming utterance might take place, the difference being that in the case of unplanned pauses, they do not mark a structural unit, instead occurring because the speaker needs additional planning time. [Ramanarayanan, Bresch, Byrd, Goldstein, and Narayanan \(2009\)](#) examined pause articulation during spontaneous speech for seven speakers, capturing both grammatical pauses (defined there as pauses occurring at major syntactic constituents) and non-grammatical pauses (all other pauses). [Ramanarayanan et al. \(2009\)](#) found that grammatical pauses, but not non-grammatical ones, showed a significant decrease in speed of articulator movement during the pause as compared to the pre-pause period. The period following both grammatical and non-grammatical pauses showed an increase in speed of articulators. There was also higher variation in articulator speed during and after the pause for non-grammatical pauses in comparison to grammatical pauses. More variability is generally assumed to mean less targeted, less structurally controlled movements, indicating, as [Ramanarayanan et al.](#) discuss, that the grammatical, but not the non-grammatical pauses are planned, targeted articulations. As [Ramanarayanan et al. \(2009\)](#) argue, these results suggest that different types of articulation during pauses can reflect different cognitive processes (planned grammatical breaks encoding linguistic structure vs. active cognitive planning processes).

[Ramanarayanan et al. \(2013\)](#) examined, for five speakers, vocal tract postures during acoustic pauses before or after speech ("absolute rest positions"; these were pauses at the beginning and end of a data acquisition interval), pauses directly prior to speech onset (speech-ready pauses), and grammatical silent or filled acoustic pauses during both spontaneous and read speech (inter-speech pauses). They found that the vocal tract postures differed during absolute rest position (with the articulators indicating a more closed vocal tract) compared to during inter-speech pauses and pauses prior to speech onset. They further iden-

tified differences in postures between read and spontaneous speech (with a higher jaw and lower tongue in spontaneous compared to read speech). Finally, they found a trend such that absolute rest positions showed higher variability than pauses directly prior to speech onset, which in turn showed more variability than pauses during read speech. Based on this, [Ramanarayanan et al. \(2013\)](#) suggest that inter-speech pauses in read speech are planned in the sense of structurally controlled, targeted positions, while the absolute rest pauses are likely to be the least planned, i.e. the least linguistically controlled.

Taken together, these studies provide articulatory evidence that pausing during speech can arise through multiple cognitive processes, and that these processes differentially affect the control of vocal tract articulation during the production of the pause. Articulatory configurations during pauses can be the result of targeted movements controlled by linguistic representations, but also may reflect other cognitive processes such as for example non-grammatical pauses arising through speech planning in the sense of planning an upcoming utterance ([Ramanarayanan et al., 2009, 2013](#)) or a variation in cognitive load (such as more demanding speech planning in spontaneous than in read speech, [Ramanarayanan et al., 2013](#)).<sup>2</sup>

The studies discussed so far examined specific positions of articulators or movements of the whole vocal tract. [Katsika et al. \(2014\)](#) and [Katsika \(2012\)](#) examined movements of individual articulators for eight speakers of Greek and identified pause postures which occurred during pauses in read speech at strong prosodic boundaries between sentences. These pause postures, which were visible on both the lip aperture and tongue dorsum trajectories, were spatially stable across repetitions and can be described as a movement away from a straight interpolation between a pre-boundary vowel and a post-boundary preparatory position for the upcoming post-boundary gesture (see [Fig. 3](#)), thus introducing an additional movement between the gestures of the consonants and vowels. [Katsika et al. \(2014\)](#) suggested that the pause posture could be the default articulatory setting for Greek. They developed an account of articulatory events at prosodic boundaries within the  $\pi$ -gesture model ([Byrd & Saltzman, 2003](#)), showing how pause postures show specific timing patterns with temporal, constriction, and tonal gestures (this model is described in the next section). [Katsika et al. \(2014\)](#) thus provide a way how articulatory settings could arise in relation to other linguistic events. [Katsika et al. \(2014\)](#) further suggested that the identified properties of pause postures (their temporal relationship to other linguistic

<sup>2</sup> Another set of recent studies has examined the effect of the upcoming utterance on articulation during pauses. Articulatory movements of course start before the end of the acoustic pause, but there is large variability in how early articulation starts across studies (cf. for example the results in [Mooshammer et al., 2012](#), [Rasskazova et al., 2018](#), [Drake & Corley, 2015](#), and [Tilsen et al., 2016](#)), large inter-speaker variability ([Schaeffler, Scobbie, & Mennen 2008](#), [Rasskazova et al. 2018](#), [Tilsen et al., 2016](#)), and differences depending on how much articulatory preparation is possible in a task ([Tilsen et al., 2016](#), see also [Kawamoto, Liu, Mura, & Sanchez, 2008](#) for indirect evidence for this). While the current study will not address the issue of how early articulation of an upcoming utterance starts, possible co-articulatory effects of the upcoming utterance need to be taken into account in studies of articulatory settings (our study controls for these by having the post-boundary utterance start with the same word in all stimuli).

units and spatial stability) indicate that they may be targeted, controlled movements (i.e., cognitive units).<sup>3</sup>

Rasskazova et al. (2018) also examined articulatory movements during grammatical pauses for eight speakers of German in read speech and found “rest” trajectories and “transitions”. Transitions refer to tongue movements that proceed from the pre-pause utterance to the post-pause utterance in a “smooth” movement, while “rest” trajectories consist of either articulators not moving after completion of the pre-boundary gesture, or of the tongue moving to the palate and staying there (given that this is an additional movement, it might be a pause-posture). Rasskazova et al. (2018) found that the frequency of occurrence of these two types of articulations during pauses differs between speakers, with some speakers predominantly having “transitions” and other speakers predominantly producing “rest” trajectories (see also Schaeffler, Scobbie, & Mennen, 2008 for speaker specific articulatory behavior during pauses), and that speakers who predominantly produced “rest” trajectories also had longer pauses, but not slower speech rate, which seemed fairly constant across speakers.

The above studies showed large variability in the articulation of acoustic pauses. What is not clear from these studies is what cognitive function this articulatory behavior reflects, which is an essential question if we are to understand pauses. It is evident that articulatory behavior during pauses is not just language-specific and speaker-specific, but it is also pause-type specific, and variation is shown to occur even during the same type of boundary. On the assumption that movements in the vocal tract are not random but reflect either cognitive processes (possibly each with their own articulatory target) or physiological needs, these differences indicate that absent a physiological explanation, variation in cognitive processes could underlie these different types of articulations. Thus, the primary goal of this study is to begin to examine which cognitive processes underlie articulations during pauses. We will focus on pause postures at prosodic boundaries, with the understanding that they might be the default articulatory settings of the vocal tract.

### 1.3. Theoretical account of pause postures and prosodic boundaries

Before we can examine the cognitive processes associated with pause postures, we first need to establish their existence in American English. The only account of properties of pause postures is given in Katsika et al. (2014), within the framework of Articulatory Phonology. This model, which accounts for the interdependence of tonal and temporal properties of boundaries, stress, and pause posture, will be introduced in some detail in this section as the predictions of the model will be used to identify the timing of pause postures with other gestures. We will only discuss properties relevant for this study; many aspects of it, and the motivation behind it will not be dis-

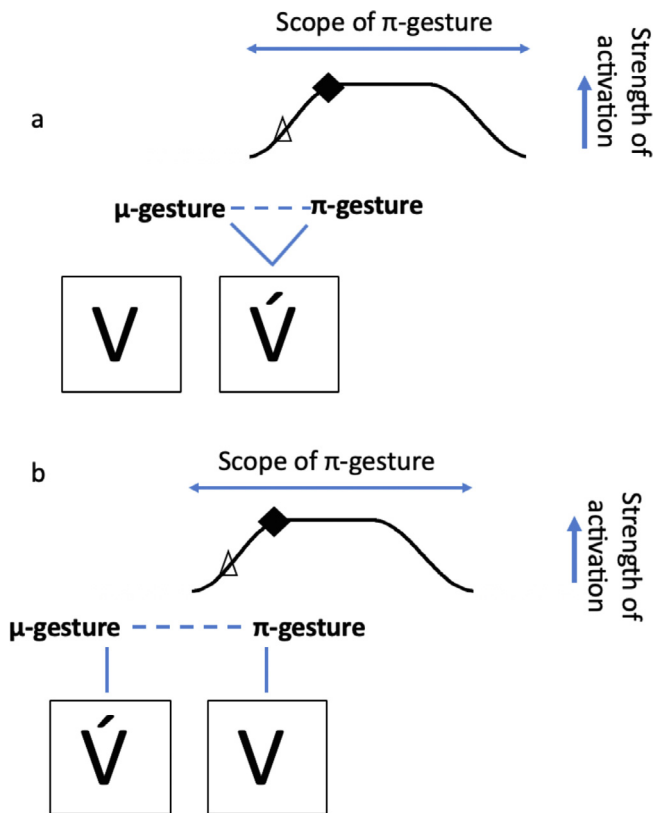
cussed (for a more detailed review see Krivokapić, 2014, 2020).

Within Articulatory Phonology, boundaries are understood to arise through the interplay of prosodic gestures ( $\pi$ -gesture,  $\mu$ -gesture, and tone gestures) with constriction gestures. Prosodic gestures model prosodic properties, while constriction gestures model segments. The  $\pi$ -gesture (Byrd & Saltzman, 2003) extends over an interval and during that interval slows the clock that controls a speaker’s speech rate. The scope of the  $\pi$ -gesture is at this point an empirical question, as is the question whether a boundary has one or two  $\pi$ -gestures. There are two possibilities: it could be one gesture spanning the period starting somewhere towards the end of the prosodic phrase and ending somewhere at the beginning of the following prosodic phrase, or one  $\pi$ -gesture at the end of the phrase and one at the beginning of the following phrase (see Byrd & Saltzman, 2003; Katsika, 2016 for discussion). The effect of the  $\pi$ -gesture is that co-active gestures become slower, spatially larger and temporally longer (accounting, among other things, for the well-known lengthening at prosodic boundaries), and less overlapped. The strength of the effect of the  $\pi$ -gesture is determined by its activation level, with stronger activation levels leading to stronger boundary effects (such as more lengthening for a more strongly activated  $\pi$ -gesture than for a less strongly activated  $\pi$ -gesture). Another prosodic gesture, the  $\mu$ -gesture (Saltzman, Nam, Krivokapić, & Goldstein, 2008), models temporal effects of lexical stress, also lengthening gestures co-active with it (both the  $\pi$ -gesture and the  $\mu$ -gesture lengthen gestures co-active with them, the difference between these two types of temporal gestures is mainly in their implementation within the computational model of Articulatory Phonology). Finally, tone gestures have been proposed to model lexical tone (Gao, 2008). Tone gestures have as their goal linguistically relevant F0 targets (such as H and L tones). They have also been suggested to account for pitch accents (e.g., Mücke, Nam, Hermes, & Goldstein, 2012) and boundary tones (Katsika et al., 2014). Based on their analysis of the effect of prosodic boundaries and prominence in Greek, Katsika et al. (2014) suggest the following account of the temporal relationships of prosodic gestures at the boundary (the schematic representation of the model is shown in Fig. 1). As one aim of this study is to establish the existence and grammatical status of pause postures, we present aspects of this model that make predictions about the temporal patterns in our data:

- (1)  $\pi$ -gestures are coordinated with the phrase-final vowel at the boundary and, weakly, with the  $\mu$ -gesture of the stressed syllable. Depending on where the lexical stress of the phrase-final word is, the  $\mu$ -gesture co-occurs either with the final vowel (if stress is on the last syllable of a word, Fig. 1a) or with an earlier vowel in the phrase final word (if stress occurs earlier in the word, Fig. 1b). If stress (and the  $\mu$ -gesture) occurs earlier in the word, then the coordination of the  $\pi$ -gesture with the  $\mu$ -gesture will lead to a slight shifting of the  $\pi$ -gesture towards the  $\mu$ -gesture. This means that the  $\pi$ -gesture starts earlier when stress is on the first syllable than when it is on the second (compare Fig. 1b to Fig. 1a).
- (2) The boundary tone is triggered when the  $\pi$ -gesture reaches a certain level of activation. While “certain level of activation” is not a clearly defined point, the implication of this is that there will

<sup>3</sup> Relatedly, within the framework of Articulatory Phonology it has been suggested that default articulatory settings may emerge from each articulator returning to its default position via the neutral attractor of each articulator. The neutral attractor takes the articulator from its constriction position to its neutral position when the articulator is not under active control of a gesture, thus ensuring that an articulator does not remain in the constriction position once a gesture is not active any more (Saltzman & Munhall 1989, see also the discussion in Ramanarayanan et al., 2013).





**Fig. 1.** Simplified schematic representation of boundary events as given in Katsika et al., 2014 for words with stress on the second syllable (a) and for words with stress on the first syllable (b); The model is simplified to only represent aspects of it relevant for the issue to be examined here. “V” indicates stressed vowels. The lines indicate coordination between gestures, and the dashed lines indicate a weaker coordination. The triangle marks the onset of the boundary tone and the rhomboid the onset of the pause posture. Figure adapted from Katsika et al., 2014.

not be a boundary tone without lengthening (note that by virtue of having a boundary tone, this boundary will be an Intonation Phrase (IP) boundary in the model of Beckman & Pierrehumbert, 1986). Given the shift of the  $\pi$ -gesture towards the stressed syllable (as described in 1), this threshold is reached earlier, and thus the boundary tone occurs earlier, when stress is on the first than when it is on the second syllable of a bisyllabic word.

- (3) Pause postures are triggered by an even stronger activation of the  $\pi$ -gesture than boundary tones are. Again, the term “stronger” is not well defined, but evidence for a stronger activation of the  $\pi$ -gesture is in general more lengthening, and for this particular assumption, the implication is that there will be no cases of PPs occurring in utterances without boundary tones. This prediction is consistent with the fact that only strong IPs have pauses. A further implication is that the temporal relationship between the boundary tone and the pause posture onset is stable, independent of the position of the stressed syllable: both boundary tone and pause posture will be triggered by specific levels of activation of the  $\pi$ -gesture, and these will occur earlier when the stressed syllable is earlier in the word than when it is later (as described in point 2). However, the relationship between the boundary tone and the pause posture will be stable (schematically shown in Fig. 1).

These patterns are evidenced in temporal relations between the boundary tone, the final vowel, and the pause posture, and thus lead to specific predictions for the temporal

relationships between certain articulatory landmarks of these gestures. We will describe these in the methods section.

Finally, it is generally assumed that controlled, targeted movements show relatively little variability (for discussion of this point and a more nuanced view on variability see Gick et al., 2004; Riley & Turvey, 2002; Whalen, Chen, Tiede, & Nam, 2018). Thus, in order to examine the cognitive status of pause postures, we will also examine variability of the pause posture (see also Katsika et al., 2014 for such an analysis).

#### 1.4. Motivation for the current study

One of the main questions of this study is to examine when pause postures occur. As pause postures seem to be tied to strong prosodic boundaries, they also might be tied to discourse structure, given the close link between prosodic and discourse boundaries. Prosodic and discourse boundaries are related in the sense that while discourse boundaries serve to mark larger discourse units (such as change of topic) at or above the level of a sentence and prosodic boundaries mark smaller units, namely prosodic phrases (which often but not always correspond to syntactic phrases), both are marked phonetically in a similar way, with the difference that discourse boundaries are stronger, for example having longer pauses than typical sentence level prosodic boundaries (Lehiste, 1975; Swerts & Geluykens, 1994; Beckman, Hirschberg, & Shattuck-Hufnagel, 2005). Given the observed variability in articulations at different types of pauses, the first question we address is whether articulatory settings are tied to specific discourse-pragmatic uses. Specifically, we examine the role of discourse in the occurrence of pause postures. However, the existing studies examining discourse used longer stretches of spontaneous speech, which, while ideal for the analysis of discourse and pragmatics, is not feasible for a study as controlled as the current one needs to be in order to investigate the articulation during pauses. We therefore constructed sentences varying in syntactic structure, meaning, and punctuation, with the goal of eliciting a variety of discourse-pragmatic interpretations, in the expectation that some of the sentences will elicit more pause postures than others. At this point, we did not formulate a hypothesis more specific than this; as there are no indications in the literature as to the cognitive functions of pause postures/articulatory settings (other than the brief points in Ramanarayanan et al. (2009, 2013) that various cognitive processes could underlie articulatory settings, and that they could be interacting with other processes of speech production), more well-founded hypotheses cannot be made. If our expectation is met, future studies will examine more specific discourse-related questions.

We further examine how pause postures may relate to cognitive function. As mentioned above, acoustic pauses serve multiple functions. Ferreira (2007, see also 1988, 1993) suggests, as a strong hypothesis, that the acoustic pause at prosodic boundaries can be divided into two fundamentally different parts: she suggests that the first part is the implementation of the prosodic boundary, while the second is related to planning. While we expect that both parts of the pause are implementations of the prosodic boundary, and that planning proceeds throughout the utterance, including through both parts of the pause, it is an empirical question if planning takes

place predominantly in some parts of the boundary or evenly throughout. One interesting possibility is that these two functions of the boundary are indicated by different articulatory behavior. Although the current study was not initially designed with this question in mind, the stimuli used allow us to additionally examine the role of speech planning in the occurrence of pause postures and thus shed some light on the question whether the boundary is divided into different cognitive parts which are reflected in articulation. Specifically, as discussed above, it is known that longer upcoming phrases take longer time to plan. We examine whether there is a relationship between the amount of planning needed (as indicated here by the number of syllables in the upcoming prosodic phrase) and pause posture occurrence and duration. We test the hypothesis that pause postures are more frequent and longer before longer upcoming phrases, allowing speakers more time to plan an upcoming utterance.

Before examining these two questions, we first need to examine whether there is a pause posture (PP) in American English, specifically asking (1) whether we see evidence of a consistent articulatory pattern during pauses (similar to the pause posture in Katsika et al., 2014) and (2) whether this articulatory pattern shows a consistent temporal relation to other linguistic events at the boundary, which would be additional evidence of the status of the PP as a cognitive, controlled, unit. We use a subset of the measures Katsika used in developing her model of gestural coordination at prosodic boundaries (Katsika et al., 2014) as a diagnostic for addressing the second question.<sup>4</sup> As an additional measure, we further examine spatial variability (also following Katsika et al., 2014; Gick et al., 2004; Ramanarayanan et al., 2009, 2013). To begin the investigation of pause postures in American English, we decided to focus on pause postures of the lip aperture (LA). While previous research has examined various articulators, and each articulator could be used to address this question, we focused on LA as it is relatively straightforward to label, which is useful for a study of a fairly new phenomenon. Based on what we know from existing research, there are two theoretical possibilities of how pause postures could occur: One is that they occur after the last active gesture for each articulator (e.g., on the tongue body after the last vowel in an utterance, on the tongue tip after the last coronal consonant of the utterance), thus occurring at different times for different articulators. The other possibility is that the pause posture occurs across the whole vocal tract simultaneously. According to Katsika et al. (2014) the latter should be the case (since, as discussed in Section 1.3., pause postures are triggered at a certain level of activation of the  $\pi$ -gesture). This is an empirical question, but we will not address it in our study, where we will focus on one articulator only, as it will suffice to address our main questions.

As the question of pause postures is new, and as established methods for determining a pause posture do not exist, there is a potential problem in identifying the presence and extent of pause postures. That is, it may be difficult to distinguish the targeted movements associated with the pause posture from the background of noise and interpolative motion. To

address this issue, in our analysis we will first start with labeling by a human annotator, but we will also use machine learning to identify pause postures. Although the human annotations alone could provide us with the data required to study the distribution and potential triggers of pause postures, and any supervised machine learning task will necessarily be guided by (but not beholden to) the human judgements which provide the training data, for this analysis of new and not-well-understood phenomena, we must ensure that the phenomena under discussion are measurable and reproducible through a mathematically predictable and transparent means, rather than based solely on a human judgement or heuristic. Finally, we hope that this work will provide a useful and generalizable approach to gestural detection, and ultimately result in a model which can identify pause postures in novel data using the same criteria as previously employed.

Shaw and Kawahara (2018) present one possible approach to solving this issue as a component of their suite of tools for identifying phonological targets in phonetic data. They use discrete cosine transform (DCT) to model continuous articulatory data as a series of four coefficients, then classify tokens according to their likelihood of being targeted movement using a straightforward Bayesian classifier. This technique combines the flexibility of DCT-based curve modeling with the decision-making abilities of a Naïve Bayes classifier, and provides a more nuanced manner of determining the status than a simple binary human decision. We present a similar but more generally applicable method, using Support Vector Machines (SVMs) for evaluating the presence or absence of pause postures, again using curvature analysis and machine learning, and describe how these models can be used to categorically describe annotated data.

To summarize, the goals of the study are to examine (1) whether pause postures occur in American English, (2) whether they can be considered cognitive units, (3) whether they are related to discourse structure and speech planning, and (4) whether they can be detected automatically using curvature analysis and machine learning.

## 2. Methods

We present an electromagnetic articulometer (EMA) study examining the existence and kinematic properties of pause postures in American English.

### 2.1. Participants

Eight participants (four male and four female) with no reported history of speech or hearing disorders participated in the current study. Data from one participant were not processed as he had difficulties with the set-up and with reading the sentences. The participants were students at the University of Southern California and were naïve as to the purpose of the experiment.

### 2.2. Stimuli

Fourteen sentences were designed to elicit a range of boundaries with varying syntactic structures and pragmatic uses (Table 1). To begin the investigation of pause postures

<sup>4</sup> Not all the analyses from Katsika (2012, Katsika et al., 2014) were conducted, as this went beyond the scope of the paper. We conducted those that would be most indicative of the timing patterns of pause postures to other gestures at the boundary.

Table 1

Stimuli for the target word “Mlma”. The same sentences were recorded with the target words “miMA” and “biBU”. Stress on the target word is indicated by capital letters. Participants read aloud target sentences (T) which were sometimes preceded by a context sentence (C), which participants read silently. The number of syllables represents the number of syllables preceding the boundary and the number of syllables of the first prosodic phrase following the boundary. The investigated boundary is marked by “#” (the pound sign was not in the stimuli presented to the participants).

	Stimuli	Number of syllables before the boundary	Number of syllables in the first prosodic phrase after the boundary	Boundary
1.	C: You two know everything! T: I don't know about Mlma. # Mini does know though.	7	5	IP boundary
2.	C: What should we talk about? T: There's a lovely story I know about Mlma. # Mini doesn't like it though.	12	7	IP boundary
3.	C: I will ask you about Mlma later. T: I don't know about Mlma. # Mini doesn't tell me these things.	7	8	IP boundary
4.	T: I don't know about Mlma— # Minni does—but I know about birds.	7	3	IP boundary
5.	T: I know about Mlma, # Mini, and the rest of the gang.	6	2	IP boundary
6.	T: Here is what I know about Mlma: # Mini doesn't like her.	9	6	IP boundary
7.	T: <i>I don't know about Mlma # mini-dolls going on sale.</i>	7	7	Word boundary
8.	C: Does Mina know about Mlma? T: Does she know about Mlma? # Mini discovered Mlma!	7	7	IP boundary
9.	C: You two know everything! T: I don't <b>know</b> about Mlma... # Mini does know though	7	5	IP boundary
10.	C: So you think you will get to know about Mlma? T: I <b>hope</b> we get to know about Mlma... # Mini seems to like her.	10	6	IP boundary
11.	C: So Bob certainly knows about Mlma? T: Bob <b>may</b> know about Mlma... # Mini doesn't think so though.	7	7	IP boundary
12.	C: So Bob certainly knows about Mlma? T: Bob <b>could</b> know about Mlma... # Mini doesn't think so though.	7	7	IP boundary
13.	T: # Mlma mini-dolls are going on sale.	0	10	Phrase-initial IP boundary
14.	T: I know all about Mlma. #	7	0	Phrase-final IP boundary

in American English, we decided to focus on the lip aperture (LA). We therefore chose target words that contain bilabial consonants so that we can control LA trajectories. There were three target words: *Mlma*, *miMA*, *biBU* (capitalization indicates lexical stress), pronounced as [ˈmimə, mɪˈmɑ, bɪˈbu]. The words *Mlma* and *miMA* varied lexical stress so that the temporal relationship of the pause posture with the utterance final word could be examined (as the predictions of the model developed in Katsika et al. (2014) can be tested using different stress patterns of phrase-final words), while *biBU* was included to vary vowel context (since the reason to include this target word was only to examine the effect of the vowel context, not to examine the effect of stress per se, lexical stress was not manipulated on this target word). While these target words required participants to learn new words, they allowed for phonetically controlled boundaries, which was necessary for the purposes of this study.

The stimuli for the target word *Mlma* are shown in Table 1, and the sentences with the target word *miMA* and *biBU* were identical except for the target word. Of the 14 utterances, one had the target word utterance-initially (sentence 13) and one utterance-finally (sentence 14), in order to examine pause postures utterance-initially and utterance-finally, i.e., without preceding and upcoming articulatory targets respectively. In the other twelve sentences, the boundary (#) occurs in the string “know about [target word] # mini”. In eleven of these twelve sentences, the boundaries were targeted to be IP boundaries (sentences 1–6, 8–12), while in one, the boundary was targeted to be a word boundary (sentence 7; this sentence was collected for another experiment but was included here since it provides additional data for the machine learning model). The twelve sentences containing the target word utterance medially varied in their syntactic structures and in their punctuation, with the goal of creating prosodic, syntactic, and pragmatic variety. They consisted of: three declarative sen-

tences where the target word was followed by a period (sentences 1, 2, 3), one sentence where the target word preceded a parenthetical sentence (sentence 4), one sentence where the target word was part of a list (sentence 5), one where the target word was followed by a colon to elicit a boundary (sentence 6), one rhetorical question (sentence 8), four sentences where the target word was followed by ... indicating an ellipsis and uncertainty (sentences 9–12), and one sentence with the target word placed utterance medially and expected not to have a prosodic boundary after the target word (sentence 7). In four of the sentences the target word was deaccented (sentences 9–12). This was done for the purposes of another experiment but contributed to the overall goal of introducing variety. The nuclear pitch accent in these sentences was indicated to the participants by the preceding bolded word. All other sentences had the nuclear pitch accent on the target word. Eight sentences (sentences 1 through 4 and 9 through 12) had a targeted L-H% sequence at the boundary (i.e., L-H% was the expected phrase accent-boundary tone sequence on the target word). This tonal sequence would allow the examination of the relationship between the boundary tone and pause postures. To elicit the targeted prosodic contours and meanings, a number of sentences were preceded by context sentences which the participants saw but did not read aloud. A note: In sentence 9, the post-boundary word was given as “Minni” instead of “Mini”. This was a typo in our stimuli, but participants produced it identically, as “Mini”.

The sentences were presented on a screen using the MARTA experiment control program (Tiede, Haskins Laboratories), and participants were instructed to read the sentences as if reading a story to someone. Eight repetitions were recorded of each utterance, for a total of 336 utterances (3 target words × 14 sentences × 8 repetitions) per participant, except for one participant, F2, where 7 repetitions were recorded, for a

total of 294 utterances. Sentences were pseudo-randomized in blocks of 14 sentences. In case of error or disfluency, participants were asked to read the sentence again, and the original reading was discarded.

### 2.3. Experiment procedure and data acquisition

Prior to the experiment, the participants familiarized themselves with the sentences by first listening to a recording of a native speaker of American English reading the sentences and then reading them aloud. The goal of the recording was to (a) familiarize the participants with the productions of the novel words and (b) indicate the targeted prosodic structure in case there are multiple ways to produce the intended sentences.

Kinematic data were recorded using an electromagnetic articulometer (EMA; WAVE, Northern Digital), at a sampling rate of 400 Hz. Sensors were placed midsagittally on the tongue tip (1 cm posterior to the apex) and the tongue body, on the upper and lower lip, and on the lower incisors (to track jaw movement). To correct for head movement, reference sensors were placed on the upper incisors and on the left and right mastoid processes. Acoustic data were acquired simultaneously using a Sennheiser shotgun microphone at a sampling rate of 16 kHz. After data collection, the articulatory data were smoothed with a 9th-order Butterworth low pass filter with a cut-off frequency of 15 Hz, corrected for head movement and rotated to the occlusal plane. Velocity signals were calculated as the first difference of the filtered position data.

### 2.4. Data labeling

A research assistant naïve to the purposes of the experiment but trained in ToBI (the Tone and Break Indices labeling system, Beckman & Ayers Elam, 1997) examined all sentences to ascertain the target words were produced accurately (with the correct stress), that the produced boundaries were IP boundaries, and that the sentences did not have disfluencies. Tokens where this was not the case were excluded (see Table 3). To examine the relationship of the boundary tone to other gestures (according to Katsika et al., 2014), a second research assistant, also naïve to the purposes of the experiment and trained in ToBI, labeled manually the onset of the boundary tone in the *Mlma* and *miMA* sentences in Praat (Boersma & Weenink, 2017). This was only done in cases where there was an L-H% tone sequence, as in these cases the boundary tones could be reliably identified. Excluded were the control no-boundary condition (sentence 7) and the utterances with the target word at the beginning of the utterance and therefore no boundary on the target word (sentence 13), and utterances with the target word at the end of the utterance but with a L-L% boundary tone (sentence 14). The onset of the H% boundary tone was taken to be the F0 minimum immediately preceding the increase in the F0 contour for the H%. In Praat, the drawing method in pitch setting was set on “speckles” to allow for more precision in measurement. A large number of cases could not be labeled due to creaky voice. The speakers had the following number of tokens that could be labeled for the boundary tone (with the percentage of the total number given in parentheses): F1 had 58 tokens (33%), F2

had 53 (32%), F3 had 80 (44%), F4 had 22 (12%), M2 had 38 (22%), M3 had 122 (67%), and M4 had 47 (24%). Some utterances were excluded because the pre-boundary or post-boundary consonant could not be labeled. The result was a total of 2190 tokens that were submitted to human data labeling and to a Support Vector Machine classifier.

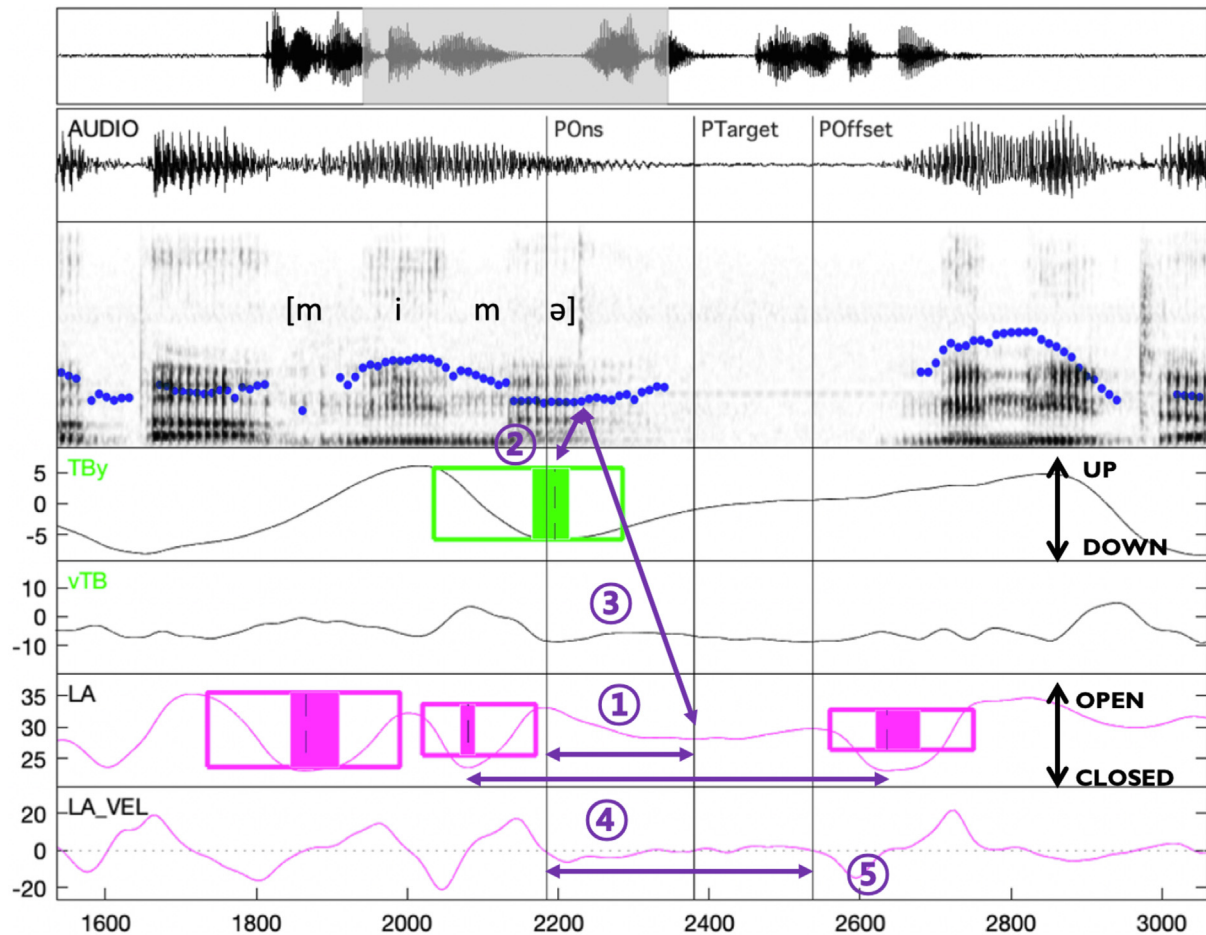
The two bilabial consonants of the pre-boundary target word, the bilabial consonant of the post-boundary word, and the phrase-final vowel were semi-automatically labeled using mview (Haskins Laboratories, under development). See Fig. 2 for an example of data labeling.

Vowels were labeled in order to examine the effect of stress on gestural timing, since Katsika et al. (2014) suggest that the  $\pi$ -gesture is coordinated with the last vowel and the stressed syllable. The consonants were labeled in order to delineate the boundary and to examine spatial stability of the PP in comparison to the most similar constriction, which is the first consonant of the target word. Vowels were labeled on the tongue body (TB) vertical displacement trajectory and consonants were labeled on the lip aperture (LA) signal, defined as the Euclidean distance between the upper and lower lip sensors. The landmarks were identified using velocity of the LA and tangential velocity for the tongue body. The landmarks were gesture onset (marked as the left end of the boxes in Fig. 2), maximum constriction (marked as a dashed line in the shaded boxes in Fig. 2), gesture offset (marked as the end of the boxes in Fig. 2). The velocity threshold was 20%.

Pause postures (PP) were identified on the basis of lip aperture using mview (Haskins Laboratories, under development) as movements that were not clear interpolations between the pre-boundary and post-boundary consonant constrictions, in the sense that they had a movement which clearly deviated from the expected interpolation trajectory (see Fig. 3; see also Schaeffler et al., 2008; Katsika, 2012; Katsika et al., 2014 for a description of pause posture as movements away from an interpolation). On this trajectory, three data points were identified: the onset of the pause posture was identified as the velocity zero crossing preceding a change in direction of movement towards the pause posture. The offset was identified as the zero-crossing preceding the maximum lip opening before the following bilabial consonant. The target of the PP was defined as the maximum constriction of the lip aperture, i.e., the point where the lip aperture was the smallest. While some cases were clearly movements away from an interpolation, as in Fig. 3, other cases seemed less clear to be a targeted movement. As a preliminary analysis, every movement away from a straight interpolation line (towards closing the lips), 1 mm or larger, between the pre-boundary and the post-boundary consonant, was labeled a PP. These manual identifications were subsequently used to train a machine-learning model to independently confirm the presence of pause postures on the basis of curvature and to evaluate the feasibility of future automatic annotation. We used the results of this machine learning model to determine which utterances contained pause postures, but for all pause postures, the pause posture landmarks were determined as described.

From these landmarks the following six variables (shown in Fig. 2) were derived. The first four will be used to examine the relationship of the pause postures to other linguistic events.





**Fig. 2.** Constriction labeling for the sentence “I don’t know about Milma, Mini does, but I know about birds”, showing the labeling for the pre-boundary vowel on the TB trajectory, and on the LA trajectory the two bilabial consonants before the boundary, the pause posture, and the bilabial consonant after the boundary. For the bilabial consonants and the vowel, the boxes indicate gesture onset (left end of the box), gesture offset (right end of the box), and the dashed line indicates maximum constriction. The three vertical lines show the pause posture onset, target (maximum constriction) and offset. The boundary tone was labeled in Praat and F0 is overlaid here only for purposes of illustrating the temporal intervals. The temporal intervals examined in this study are: 1 = Pause posture formation duration, 2 = Boundary-tone to V-target interval, 3 = Boundary-tone to PP-target interval, 4 = Boundary duration, and 5 = Pause posture duration. LA: lip aperture trajectory and velocity, TB: tongue body trajectory and velocity.

- (1) *Pause posture formation duration (interval 1 in Fig. 2): time from maximum constriction of the pause posture to the onset of the pause posture*

Katsika et al. (2014) found that the duration of the PP formation movement was longer in words with stress on the second syllable than in words with stress on the first syllable. The explanation for this was given as follows: If the  $\pi$ -gesture (which models the boundary) is coordinated with the  $\mu$ -gesture (which models stress), it will shift towards the stressed syllable and therefore it will start earlier in words with stress on the first in comparison to words with stress on the second syllable. A consequence of this would be, as was argued in Katsika et al. (2014) and Katsika (2016), that the  $\pi$ -gesture ends earlier, and therefore there is less lengthening of the pause posture formation duration, in words with stress on the first syllable.<sup>5</sup>

- (2) *Boundary-tone to V-target interval (interval 2 in Fig. 2): time from the onset of the boundary tone to the maximum constriction of the pre-boundary vowel*

Based on the analysis of Katsika et al. (2014), it is predicted that the boundary-tone to V-target interval will be longer if stress is on the second than when it is on the first syllable of the target

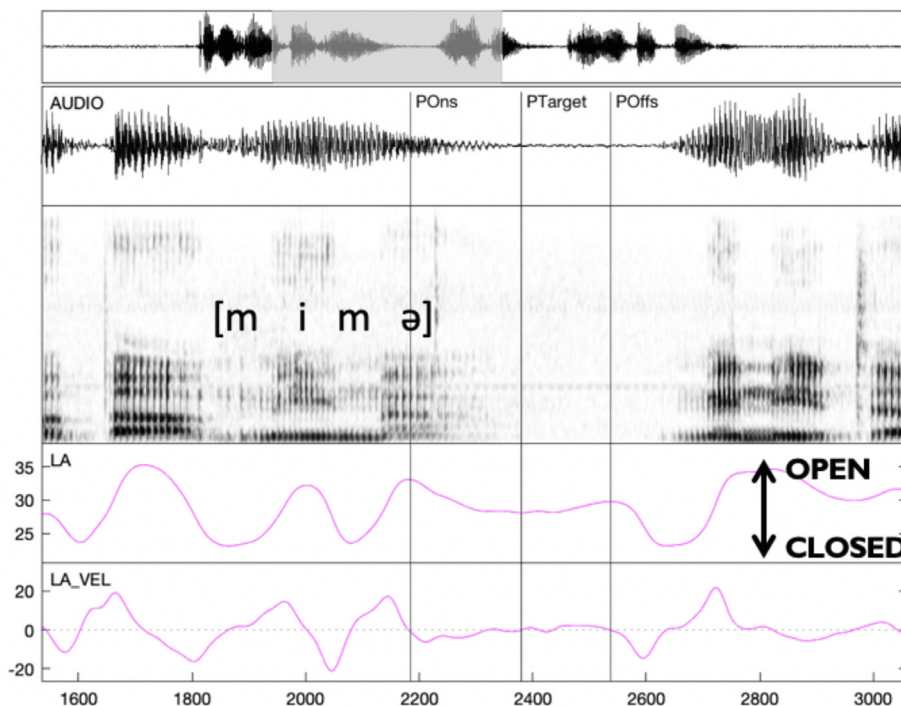
word. This is based on Katsika’s suggestion that the boundary tone is triggered when the  $\pi$ -gesture reaches a strong level of activation. Since the  $\pi$ -gesture shifts to the stressed syllable, the relevant activation level will be reached earlier within the target word when stress is on the first syllable, and therefore the boundary tone onset, while always occurring during the final vowel, will occur earlier when the stress is on the first syllable (Katsika et al., 2014). Consequently, the interval between the vowel target and the boundary tone will be longer when stress is on the second syllable.

- (3) *Boundary-tone to PP-target interval (interval 3 in Fig. 2): time from maximum constriction of the pause posture to the onset of the boundary tone*

The third assumption of Katsika et al. (2014) is that both the boundary tone and the PP are triggered when the  $\pi$ -gesture reaches a certain level of activation, and therefore the temporal relation between these two events is stable. Thus, stress is not expected to have an effect on the boundary-tone to PP-target interval.

- (4) We also compare the *spatial variability* of the maximum lip aperture constriction of the pause posture (pause posture target in Fig. 2) and the maximum constriction of the first consonant of the pre-boundary target word (dashed line in first pre-boundary consonant in Fig. 2). If the PP is a cognitive unit

<sup>5</sup> Note that under this assumption, there are two  $\pi$ -gestures at the boundary (one phrase-finally and one phrase-initially).



**Fig. 3.** Pause posture labeling for the sentence “I don’t know about Milma, Mini does, but I know about birds”. The identified landmarks are pause posture onset, maximum constriction (target), and offset. LA: lip aperture trajectory and velocity.

(i.e., is a targeted movement), it should show similar variability as other targeted movements. Therefore, the prediction is that the consonant and the pause posture maximum constrictions will show no difference in variability.

The following two measures will be used in our examination of the effect of planning in [Section 3.3.2](#) and will be explained in more detail there.

- (5) *Boundary duration*: time from maximum constriction of the post-boundary consonant to maximum constriction of the pre-boundary consonant (interval 4 in [Fig. 2](#)). This interval is a good approximation of boundary strength as it contains both pause duration and the parts of the surrounding phrases that contain the most final and initial lengthening. While we primarily examine boundary duration in relation to planning, we also test a prediction of Katsika’s model. The prediction is that PPs are more likely to occur at stronger prosodic boundaries, since PPs are triggered by a high level of activation of the  $\pi$ -gesture (and a high level of activation of the  $\pi$ -gesture leads to strong prosodic boundaries).
- (6) *Pause posture duration* (interval 5 in [Fig. 2](#)): time from the offset to the onset of the pause posture.

## 2.5. Support vector machine model

Following annotation, the lip aperture was exported for all trials, yielding discrete measures of lip aperture at 400 Hz. Combining these data with the human annotations allows us to extract the portion of the trajectory corresponding to the prosodic boundary, defined here as the period from gesture offset in the pre-boundary consonant to gesture onset in the post-boundary consonant. The annotator pause posture judgements were extracted as well. The annotator coded these as binary decisions for all speakers. This resulted in 2190 trajec-

tories across the seven speakers, of which 30.5% were labeled by the human annotator as containing pause postures.

As previously discussed, we will formally evaluate the presence or absence of pause postures by modeling the differences in trajectory captured by these human judgements using mathematically-driven statistical approaches in the R statistics computing environment ([R Core Team, 2013](#)). Note that we did not apply this method to the two sentences that had pauses examined at utterance beginning (sentence 13) and utterance end (sentence 14), since their kinematic trajectories were quite different from those of utterance medial pause postures.

We have approached the problem of identifying gestures (in this case, pause postures) in noisy data as a classification problem, well suited to machine-learning-based approaches (see [Shaw & Kawahara, 2018](#) for a similar approach). To this end, we extracted mathematically-derived features from each individual trajectory, then fed those features, along with the annotator’s judgements, into a Support Vector Machine (SVM), a specialized statistical model designed to optimally separate data into pre-defined groups. A successful algorithm will be able to, on the basis of the extracted features alone, identify pause postures, and reliably distinguish them from interpolative trajectories without pause postures, with high agreement with the human annotations.

First, the extracted pause trajectories were time-normalized by interpolating the trajectory to 1000 points, then rescaled to a 0–1 range using the minimum and maximum of each curve. This normalization allows the direct comparison of pauses with different durations, and from different words and speakers (which may have larger or smaller raw values for lip aperture). In addition, as one salient characteristic of pause postures is the deviation from the maximally efficient trajectory, a line of

direct interpolation (between the offset and onset of the surrounding gestures) was calculated. This interpolated trajectory was then subtracted from the actual trajectory, yielding a new curve effectively measuring (normalized) deviation from interpolation over time.

These two curves, representing the normalized trajectory and deviation-from-interpolation, were then used for feature extraction. Many possible features were tested, ranging from descriptive statistics on the curves, to discrete cosine transform (as used in Shaw & Kawahara, 2018), to direct measures of curve complexity. Ultimately, the best results (in the sense of largest agreement with human annotations) for SVM-based classification (described below) came from Functional Principal Component Analysis, a variant of conventional PCA which takes as input the normalized trajectories for all 2190 pauses, with rows for each trajectory and columns for each timepoint. From this  $2190 \times 1000$  matrix, fPCA determines the dominant, orthogonal patterns of temporal variation in the data. fPCA was conducted separately for the scaled trajectory and deviation-from-interpolation curves, as the scaled curves will show temporal patterns of change in position over time, and the difference-from-interpolation curves will show temporal patterns of movement away from the optimally efficient tongue path. Thus, each individual trajectory was assigned scores for each of the six most dominant components on each of the two curves (with the first six PCs accounting for 99.2% of the variability in raw curves, and 99.6% of the variability in the difference curve analysis), yielding twelve features which capture the majority of patterns of curvature in the data. The temporal patterns captured by these twelve components are displayed in Figs. 4 and 5, below, showing the patterns derived from a single  $\pm$  standard deviation shift away from a PC score of zero.

These twelve PC scores were then treated as features and combined with the binary annotator judgements of the presence or absence of a pause posture, creating a final dataset for classification. A random subset of 20% of the curves were then held aside as testing data, and the remainder was used as training data for the SVM classifier.

An SVM is, at its core, a statistical analysis designed to look at instances (here, trajectories) in a many-dimensional space defined by the measured features (here, our 12 curve components), and determine the optimal hyperplane, or many-dimensional line, which separates the data into predefined classes (here, 'pause posture' or 'none'). For the SVM described here, we used the `svm()` function in the `e1071` package for R (Dimitriadou, Hornik, Leisch, Meyer, & Weingessel, 2008), a radial kernel and 10-fold cross-validation for model training. In addition, as pause postures seem to appear in only  $\frac{1}{3}$  of the data, a class-weighted SVM was used, modifying penalties (here, 80/20 in favor of pause posture data) to help overcome the effects of class imbalance during the training process.

The resulting SVM was then evaluated in terms of its overall accuracy (relative to the human-annotations) at identifying a given curve as either having a pause posture ("1") or lacking one ("0") in the randomly-chosen test data subset which contained 375 total curves. If the algorithm is not able to replicate the human judgements, whether due to poor feature selection, or due to the pause posture pattern itself being of insufficient

consistency or robustness, we would expect the classifier to perform no better than chance. In the testing set, 71.4% of the curves contain no pause postures, thus, even a model completely ignoring the data (and simply always guessing 'no') could achieve 71.4% accuracy. To correct for this, we evaluated the models using Cohen's Kappa (Cohen, 1960) measure of agreement, where agreement is calculated with respect to expected agreement due to chance.

Using this metric, we tested a variety of features and groupings, ultimately selecting the 12 component model described above, using six PCs from scaled curves, six from deviation-from-interpolation<sup>6</sup>. In the test set, this model achieved 95.4% accuracy relative to human annotation, with a kappa of 0.89. Full results are presented below in Table 2.

The final model, when classifying the entire dataset, disagreed with the human annotator on a total of 51 trajectories (approximately 2.3% of trajectories). Although there were some cases where the SVM judgement appears incorrect, the majority of disagreements with the human annotator were either 'borderline' cases (where the trajectory could plausibly be labeled as a pause posture or not, often due to low gestural magnitude) or, among false negatives, cases where the pause posture gesture did not encompass the entire pause, with a leading or trailing area of interpolation. But the majority of SVM judgements replicate the human judgements, on the basis of these mathematical curvature characteristics alone.

Given the accuracy of the model at capturing the human-observed patterns, the high kappa (indicating that the SVM is a 'reliable annotator', c.f. McHugh, 2012, among others), and the fact that it produces judgements solely based on curvature with no preconceptions or subconscious bias, we have elected to use the SVM classifications, rather than the human judgements for our subsequent analysis of the frequency and distribution of pause postures. As the two sentences that had pauses at utterance beginning (sentence 13) and utterance end (sentence 14) were not analyzed with the SVM model, for these we use the results of the human annotator, and these serve only as preliminary results on pause posture occurrence, so as to establish whether these positions merit further investigations.

Of the 2190 utterances, 16 were excluded because the pauses were too short for the SVM model used to determine the presence of pause posture, resulting in 2174 utterances in the final analysis. The last row in Table 3 shows the total number of utterances per participant that were included in the analysis.

### 3. Results

#### 3.1. Frequency of pause postures

Of the 2174 utterances, 668 (31%) had a pause posture. There were two additional PPs identified by the model, for a total of 670, but these two pause postures occurred during the vowel production in sentences without a prosodic boundary (sentence number 7) and were likely model errors and were

<sup>6</sup> Post-hoc analysis of feature importance using the RandomForest algorithm indicate that components 2 and 3 from the scaled data PCA and component 1 from deviation-from-interpolation provide the most useful information for the classification of pause postures in this dataset, and an SVM model using these three component features alone can achieve a kappa of 0.83.

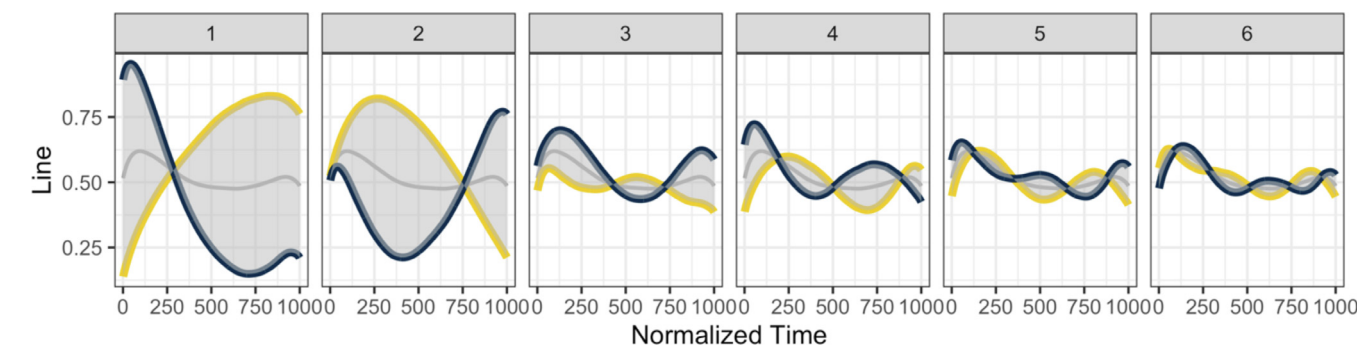


Fig. 4. High (yellow) and low (blue) variation components for scaled trajectories around the grey mean Lip Aperture trajectory. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

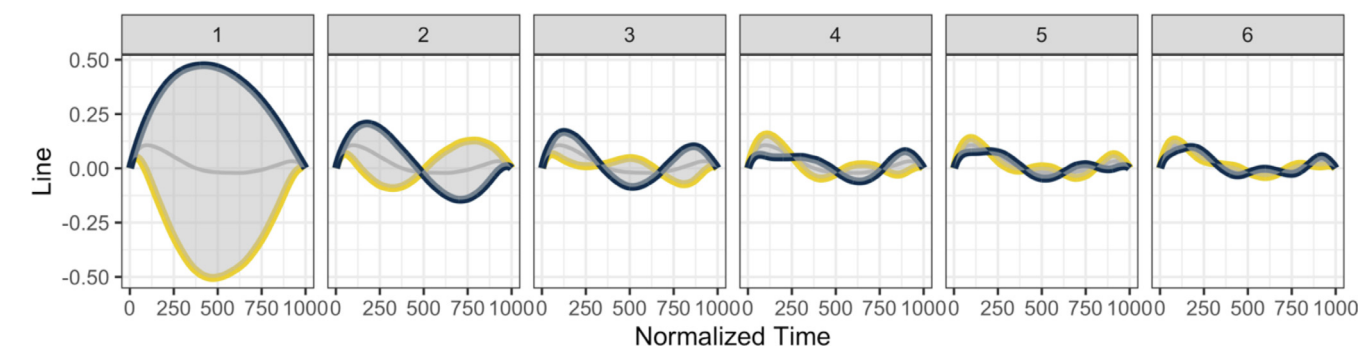


Fig. 5. High (yellow) and low (blue) variation components for deviation-from-interpolation curves around the grey mean Lip Aperture trajectory. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**  
Confusion matrix and accuracy for human annotator vs. 12-component SVM.

	Annotator 'No'	Annotator 'Yes'
SVM 'No'	258	7
SVM 'Yes'	10	100
SVM Accuracy		95.4%
Cohen's Kappa		0.89

therefore excluded from any further analyses (additional visual inspection of the tokens further confirmed that these were likely model errors). Divided by target word, out of 695 utterances with the target word *biBU*, 293 had a pause posture (42%). Utterances with the target word *miMA* had a pause posture in 175 out of 740 cases (24%), and utterances with the target word *Mlma* had a pause posture in 200 out of 739 cases (27%). Of the total number of pause postures, 44% were with the target word *biBU*, 30% had the target word *Mlma*, and 26% had the target word *miMA*. The distribution of pause postures by speaker is given in Table 4.

**Table 3**  
Data excluded due to disfluencies, prosodic errors, and unlabelability, as well as data used in the SVM model.

Participant	Disfluent/prosodic errors	Could not be labeled	Number of utterances included in the SVM model	Number of utterances included in the analysis according to the SVM model
F1	8	5	323	318
F2	2	6	286	286
F3	1	43	292	288
F4	3	4	329	329
M2	7	20	309	306
M3	2	16	318	314
M4	2	1	333	333

Of the 2174 utterances, 148 were control utterances (sentence 7) which were targeted to be produced without boundary and therefore not expected to have a pause posture. If the control sentence is excluded, the percentage of pause postures is slightly higher (33% of the 2026 utterances).

3.2. The cognitive status of pause postures

In the following set of analyses, we examine the cognitive status of PPs. We examine the temporal relationships of PPs to other linguistic events at the boundary (section 3.2.1). Related to this question, we also examine the spatial stability of the PP target (section 3.2.2).

3.2.1. The timing of the pause postures with linguistic events at the boundary

The goal of the following analyses (Section 3.2.1) is to establish whether there is a stable pattern between articulatory landmarks that are indicative of specific timing patterns



**Table 4**

Number of tokens with pause postures: All tokens with pause postures, tokens that could be used in the analysis of the temporal patterns of the pause posture, tokens that had a pause posture and a labelable boundary tone.

Speaker	Total number of pause postures and percentage (all target words included)	Total number of miMA/ Mlma tokens with pause postures for the analysis of temporal patterns (miMA/ Mlma/total)	Total number of miMA/ Mlma tokens with pause postures for which the boundary tone could be labeled
F1	117 (37%)	22/34/56	27 (48%)
F2	56 (20%)	5/18/23	2 (9%)
F3	36 (13%)	6/4/10	5 (50%)
F4	17 (5%)	2/0/2 (not included)	0 (0%)
M2	154 (51%)	41/33/74	21 (28%)
M3	144 (46%)	47/35/82	52 (63%)
M4	144 (43%)	22/43/65	41 (63%)

between gestures at prosodic boundaries. We do this by testing the predictions from the model from Katsika et al. (2014). Specifically, we test the effect of stress on 1) the pause posture formation duration (interval 1 in Fig. 2), on the boundary-tone to V-target interval (interval 2 in Fig. 2), and 3) boundary-tone to PP-target interval (interval 3 in Fig. 2). Utterances with the target word *biBU* were not examined in this set of analyses as this target word was only included to examine the effect of segmental context on pause posture occurrence, and therefore there is not a target word contrasting in stress to *biBU*. Furthermore, as mentioned earlier, the sentences where the pause posture occurred at the beginning or at the end of the utterance were also excluded. Data from speaker F4 could not be used in this set of analyses as this speaker only had two pause postures for the examined set of data (*miMA* and *Mlma* data). The complete number of tokens for this set of analyses is shown in Table 4.

ANOVAs were conducted (using JMP statistical software) for each of the conducted comparisons for each speaker separately, given the well-known differences in how prosodic structure is realized by different speakers (see e.g., Byrd & Saltzman, 1998; Kim, 2019; Krivokapić & Byrd, 2012; Parrell, Lee, & Byrd, 2013). Significance is assessed here as  $p < 0.05$ .

**Table 5**

The effect of stress on duration of the pause posture formation. ANOVA results and means and standard deviations of pause posture formation when stress is on the first syllable (stress 1) and when it is on the second syllable (stress 2). Means and standard deviations are in milliseconds.

F1	F2	F3	M2	M3	M4
stress 1 152 (35)	stress 1 157 (38)	stress 1 160 (44)	stress 1 306 (108)	stress 1 260 (72)	stress 1 216 (80)
stress 2 208 (50)	stress 2 260 (48)	stress 2 215 (23)	stress 2 376 (121)	stress 2 282 (62)	stress 2 325 (102)
$F(1, 55) = 24.3181, p < 0.0001$	$F(1, 21) = 25.0498, p < 0.0001$	$F(1, 9) = 6.9354, p = 0.03$	$F(1, 72) = 6.9618, p = 0.0101$	n.s.	$F(1, 63) = 23.8961, p < 0.0001$

**Table 6**

The effect of stress on the duration of the boundary-tone to V-target interval. ANOVA results and means and standard deviations of the boundary-tone to V-target interval when stress is on the first syllable (stress 1) and when it is on the second syllable (stress 2). Means and standard deviations are in milliseconds.

F1	F2	F3	M2	M3	M4
stress 1 -53 (32)	NA	NA	NA	stress 1 -22 (34)	stress 1 -9 (27)
stress 2 -37 (29)				stress 2 58 (67)	stress 2 49 (52)
n.s.				$F(1, 51) = 26.6012, p < 0.0001$	$F(1, 40) = 22.6526, p < 0.0001$

3.2.1.1. *Stress and duration of pause posture formation movement.* The prediction for this interval was that there is less lengthening of the pause posture formation duration in words with stress on the first syllable compared to words with stress on the second syllable. The effect of stress was significant for all speakers except M3. The effect in all cases was such that the intervals are longer when stress is on the second (*miMA*) than when it is on the first syllable (*Mlma*), supporting the hypothesis. Table 5 shows the results.

3.2.1.2. *The effect of stress on the duration of the boundary-tone to V-target interval.* We tested the prediction that the interval between the vowel target and the boundary tone will be longer when stress is on the second syllable. The analysis of the duration of the boundary-tone to V-target interval had a number of tokens missing, as in many cases the boundary tone could not be labeled (see Table 4 for the number of included tokens). The analyses were only conducted for participants where there were more than four tokens with f0 data for each target word (F1, M3, M4). The results show that stress has an effect on two speakers, such that the duration of the boundary-tone to V-target interval is longer when stress is on the second syllable, providing support for the prediction. Results are shown in Table 6.

3.2.1.3. *The effect of stress on the boundary-tone to pause-target interval.* We tested the prediction that stress does not have an effect on the boundary-tone to pause-target interval. We tested the effect for 3 speakers (F1, M3, M4), as, like above, the other participants had too few tokens with f0 data to be analyzed. For these speakers, there was no significant effect, as predicted. Results are shown in Table 7.

### 3.2.2. Variability of pause postures

If the PP is a controlled, targeted movement in English, it is expected that it will have a spatially stable target. To examine this question, we compare the spatial variability of the maximum lip aperture constriction of the pause posture and of the first consonant of the pre-boundary target word (C1). If the

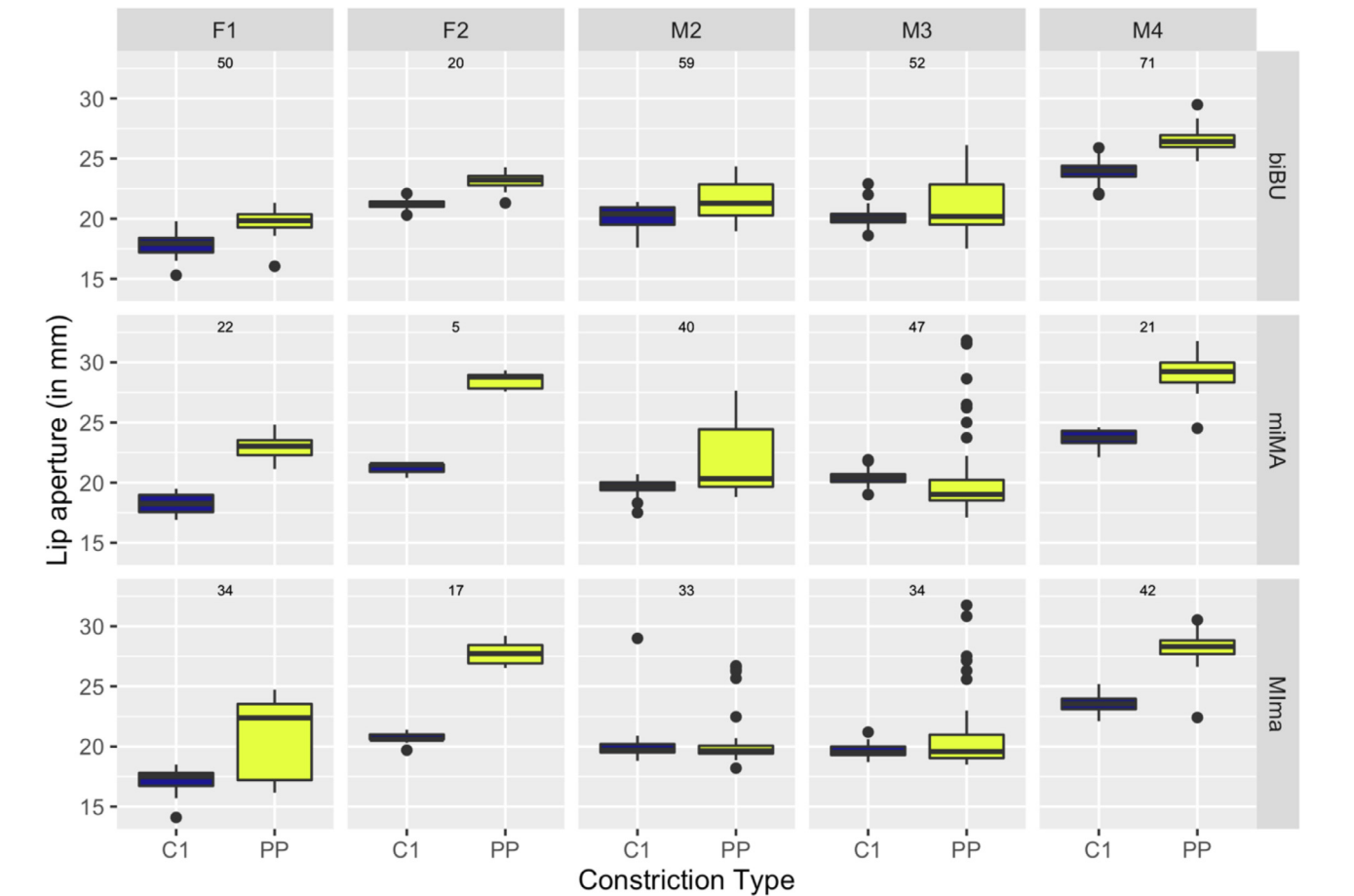
**Table 7**  
The effect of stress on the boundary-tone to pause-target interval. Means and standard deviations of the boundary-tone to pause-target interval when stress is on the first syllable (stress 1) and when it is on the second syllable (stress 2). Means and standard deviations are in milliseconds.

F1	F2	F3	M2	M3	M4
stress 1 223 (45)	NA	NA	NA	stress 1 298 (72)	stress 1 251 (98)
stress 2 214 (40)				stress 2 267 (91)	stress 2 276 (124)
n.s.				n.s.	n.s.

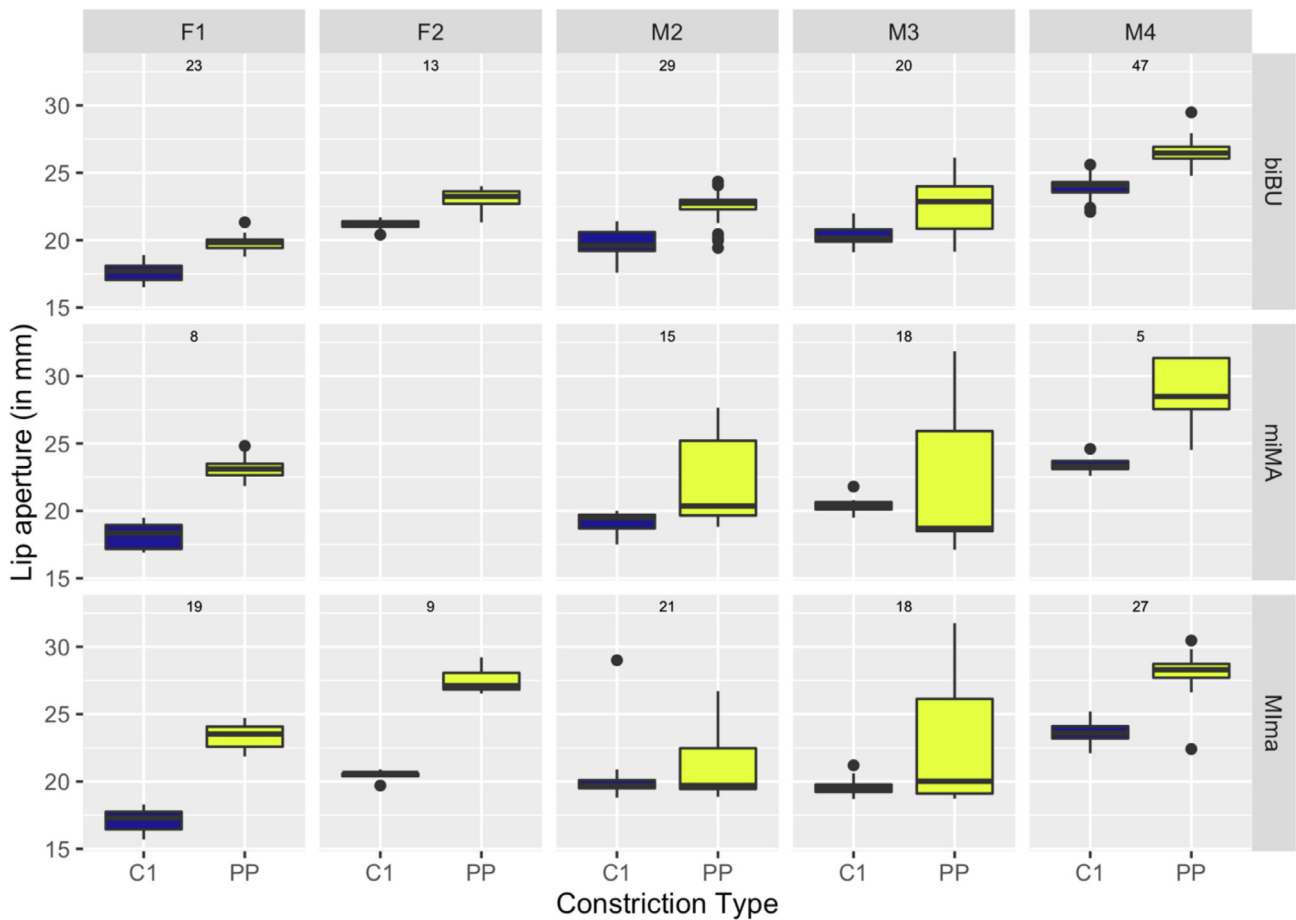
PP has a target equivalent to that of other cognitive units, we predicted that there should be no difference between pause postures and consonant constrictions in their variability. Additionally, following a reviewer’s helpful suggestion, we included in this analysis the effect of length of the pause posture itself, as we might expect less ability for the articulators to reach a given articulatory target during short pause postures (i.e., we might expect undershoot in short pause postures), and therefore more variability in shorter pause postures than in longer ones. This length was calculated by taking the by-speaker z-score of pause posture duration, z-scoring to account for across speaker differences in pause posture duration. Sentences with the PP occurring before the target word or the PP occurring at the end of the utterance were excluded from this analysis, as these were kinematically different.

We included the *biBU* sentences, since the manipulation of stress is not a prerequisite for this analysis. Data from speakers F3 and F4 were excluded as these speakers exhibited too few pause postures to conduct the analysis described below.

Figs. 6–8 compare the articulatory variability of pause postures and C1, in all pause postures (Fig. 6), in shorter-than-average duration ( $z < 0$ ) pause postures (Fig. 7), and longer-than-average duration ( $z \geq 0$ ) pause postures (Fig. 8). As seen in Fig. 6, when all pause postures are considered together, PPs have more variability than the consonant constriction, and this same increased variability holds when looking at shorter-than-average pause postures in Fig. 7. However, in Fig. 8, we see that the articulatory variability of longer-duration pause postures is, with the exception of M2’s *miMA*,



**Fig. 6.** Spatial variability for pause postures (PP, yellow) and the first consonant of the target word (C1, blue). The numbers in each panel indicate the number of consonants and pause posture tokens in the analysis (in all cases there is an equal number of C1 and PP tokens). Larger values for LA indicate more open lips. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



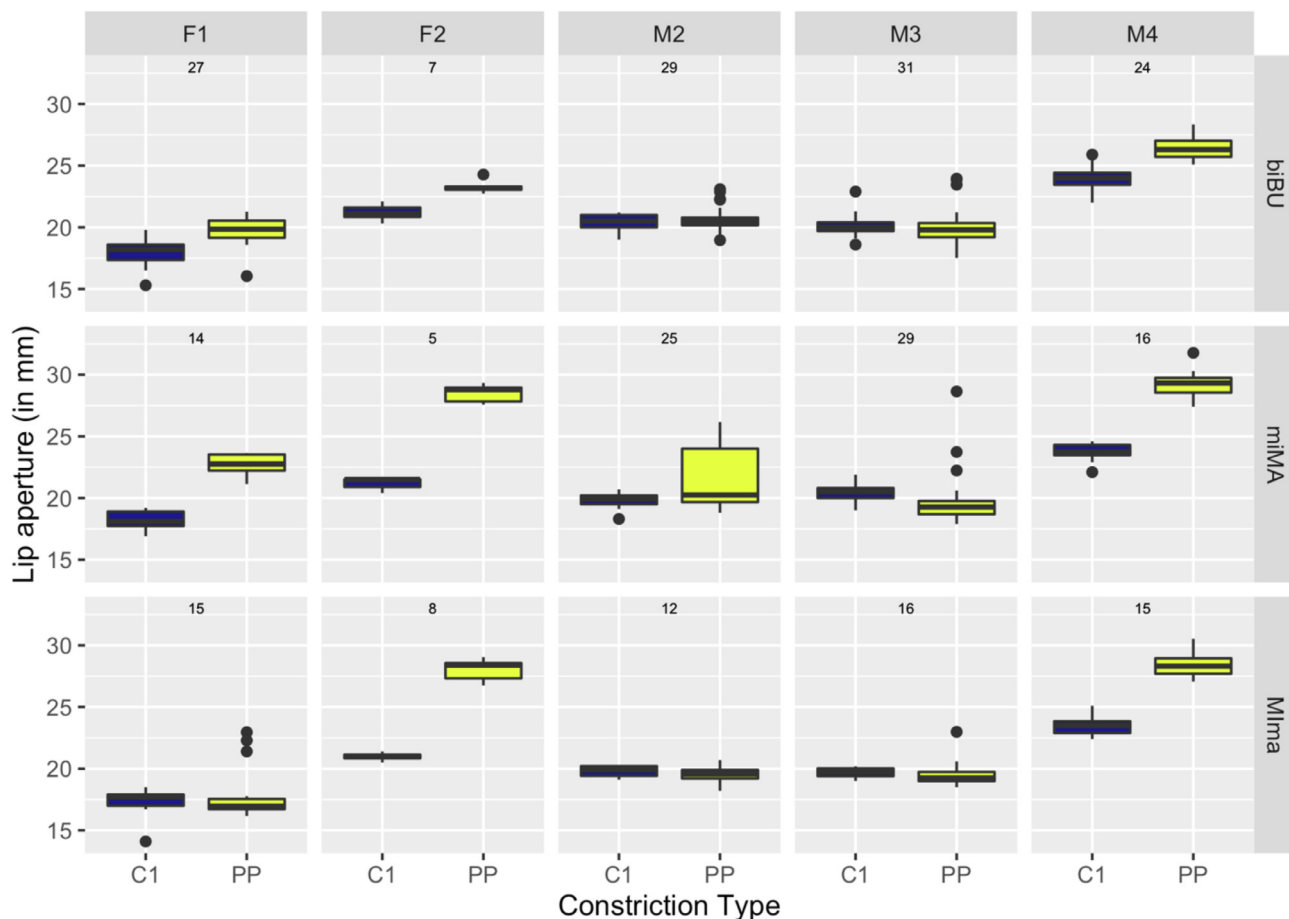
**Fig. 7.** Spatial variability for pause postures (PP, yellow) and the first consonant of the target word (C1, blue) for pause postures with a duration less than the by-speaker mean ( $z < 0$ ). The numbers in each panel indicate the number of consonants and pause posture tokens in the analysis (in all cases there is an equal number of C1 and PP tokens). Larger values for LA indicate more open lips. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

approximately equivalent to those of C1, pointing to the strong possibility of targeted movement in these longer PPs.

To test the significance of differences in variance, both between C1 and PP, as well as between target words, accounting for differences between PPs of different durations, we implemented quantile regression using the 'quantreg' package in R (Koenker, 2018). After centering the LA data around the mean by speaker for both PP and C1, to remove across-speaker differences in mean measurements, the resulting measurements were pooled, and a quantile regression was performed, estimating the 25th and 75th percentiles of lip aperture with respect to constriction type, target word, and z-scored pause posture duration (see full model output in Appendix I, Tables 1a and b. Note that due to the large number of both aggregated and by-speaker models in this paper, we are presenting the full model output tables for all quantile regression and GLM tests in the Appendix. In the text, we will discuss only those elements of the models which bear directly on the arguments being made). The regression modeled LA by constriction type (C1 or PP), interacted with target word (*biBU*, *Mlma*, or *miMA*). In this analysis on centered data, larger variance is indicated by a lowering of the 25th percentile and a raising of the 75th percentile (a widening of the overall variance for the item). In the 25th percentile, we found an overall increase in variability ( $p < 0.0001$ ) of PP in comparison to C1, and an increase in variability for *Mlma* ( $p = 0.0023$ , indicat-

ing that the effect was driven by *Mlma*, although there was no significant two-way interaction between pause posture and target word in the lower quantile estimate. We also see a three-way interaction between PP, *Mlma*, and z-scored pause posture duration, an effect for which we do not have an explanation. In the 75th percentile, there was a main effect of pause posture ( $p = 0.0107$ ) showing an overall decrease in variability compared to C1, and there was an increase in variability associated with pause postures specifically in the words *miMA* and *Mlma* ( $p < 0.0001$ ) relative to *biBU*. The net result is that although *biBU* tokens show a decrease in the 75th percentile in a pause posture context, articulations associated with the other two target words are more variable during PPs at the 75th percentile (but there was no significant interaction between pause posture and target word in the higher quantile estimate). We also see a sharp decrease in variability associated with longer pause postures ( $p < 0.0001$ ) for the 75th percentile, indicating that the asymmetry in variance associated with duration changes seen in the figures above represents a substantial effect. These results support our hypothesis that PPs are spatially as stable as other controlled targets, when accounting for the increased variability associated with short pause postures, with the PPs following *biBU* being less variable than those following *Mlma* and *miMA*.

To summarize the results so far: We found that pause postures are present in 33% of the data, and that the PP has



**Fig. 8.** Spatial variability for pause postures (PP, yellow) and the first consonant of the target word (C1, blue) for pause postures with a duration greater than the by-speaker mean ( $z \geq 0$ ). The numbers in each panel indicate the number of consonants and pause posture tokens in the analysis (in all cases there is an equal number of C1 and PP tokens). Larger values for LA indicate more open lips. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

specific temporal relations with other gestures at the boundary and spatial variability comparable to that of consonant targets for longer pause postures. Overall the results indicate the existence of a pause posture in American English.

### 3.3. Factors affecting the occurrence and duration of pause postures

The next set of analyses examine when pause postures occur and whether they can be tied to planning processes. We address the following questions: (1) what is the effect of sentence, testing whether specific constructions are more likely to lead to pause postures (Section 3.3.1), and (2) is there a relationship between speech planning and pause postures (Section 3.3.2).

For these analyses, all target words were included, as the stress difference on the target words was not a prerequisite for addressing these questions.

#### 3.3.1. The effect of sentence type on pause posture occurrence

A generalized linear model (GLM) was fitted to the data testing the effect of sentence on pause posture occurrence using the *glm* package in R (R Core Team, 2013). Note that for this and the following across-speaker analyses, mixed effects models did not converge (as we did not have enough data). All sentences except the control (no boundary) sentence were tested (see Appendix I, 2. Sentence Effect for full

model output both pooled ('Overall') and by speaker). There is an overall effect of sentence in the pooled data ( $p < 0.05$ ), as well as in five of the speaker-specific models ( $p < 0.05$ , for all speakers except F2 and F3). However, examination of the individual results (shown in Fig. 9) indicates substantial between-speaker variability in which sentences were more likely to have pause postures. There is no overall pattern of sentence effect that can be identified, so it can be concluded that sentence type does not have a single, homogenous effect on pause posture occurrence across speakers. It is worth pointing out, however, that the two sentences where the PP was targeted to occur at the beginning and at the end of the utterance (sentences 13 and 14 in the stimuli), do not behave differently from other sentences, indicating that PPs occur utterance initially and utterance finally as well as sentence medially.

#### 3.3.2. The effect of planning on pause posture occurrence

To further examine the conditions under which pause postures occur, we examine whether planning has an effect on pause posture occurrence and duration. As discussed in the introduction, acoustic pauses are known to be related to planning, with longer pauses indicating increased planning for the upcoming utterance. Therefore, in this section we examine the effect of the upcoming utterance length on pause posture



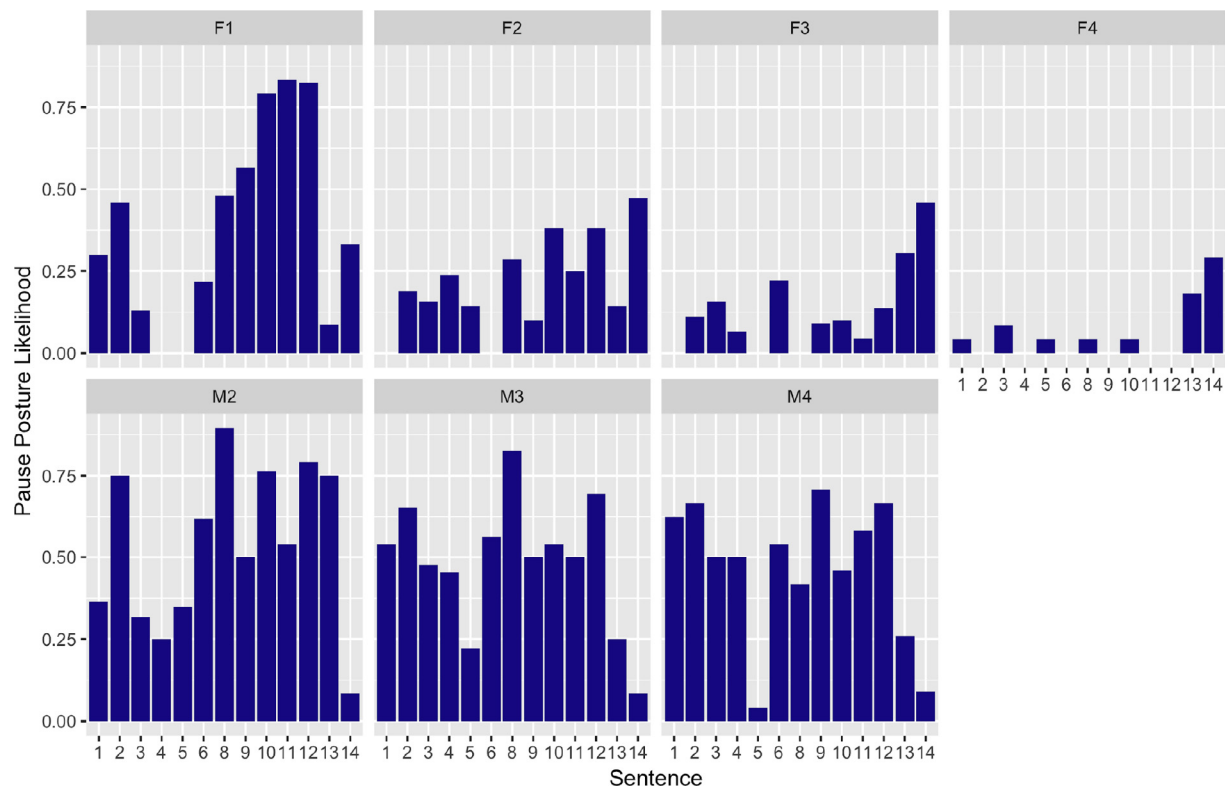


Fig. 9. The effect of sentence on the occurrence of pause posture for individual speakers.

occurrence and duration, testing the hypothesis that PPs might provide additional planning time for the upcoming utterance.

Length is defined in terms of number of syllables of the first post-boundary prosodic phrase. The number of syllables is a crude measure in this study, as the sentences were not designed to test this question and they vary considerably in syntactic structure. However, given the robust effect syllable numbers have on pause duration (e.g., Fuchs et al., 2013; Krivokapić 2007a, 2012; Zvonik & Cummins, 2003), we expect them to still have an effect on planning in the present study.

For this set of analyses, we examined all target words and all sentences except the control no-boundary, target word phrase initial, and target word phrase final sentences (sentences 7, 13, and 14 respectively). Sentences 7 and 13 differ from the others in their structure (by not having a boundary or by not having a preceding utterance) and planning might be affected by this. Sentence 14 does not have any upcoming scripted material, so the PP in that case cannot be affected by the upcoming scripted utterance.

We start by examining the relationship between boundary duration (interval 4 in Fig. 2) and pause posture occurrence, to test whether they are independent (Section 3.3.2.1). As will be shown below, they are not, and therefore further testing (in Sections 3.3.2.2. and 3.3.2.3) will need to incorporate boundary duration as a relevant variable. Section 3.3.2.2 examines the effect of the upcoming phrase length on pause posture occurrence, and Section 3.3.2.3 on pause posture duration. Finally, we test the effect of repetition on the occurrence and frequency of PPs (Section 3.3.2.4). As we explain later, we do not have predictions for this question, the contribution of this examination is exploratory in nature.

**3.3.2.1. The effect of boundary duration on pause posture occurrence.** We test the effect of boundary duration on pause posture occurrence to examine how closely related they are. A GLM was fitted to the data testing effect of increased boundary duration on pause posture likelihood (see Appendix I, 3. Boundary Duration for full model output). Longer boundary duration increases the likelihood of the occurrence of the pause posture (Figs. 10 and 11). These results were significant pooled across participants ( $p < 0.0001$ ) and for every participant individually. This result shows that boundary duration and pause posture are closely related, and that when examining the effect of planning, we need to ensure that whatever effect planning has on PPs is not simply due to boundary dura-

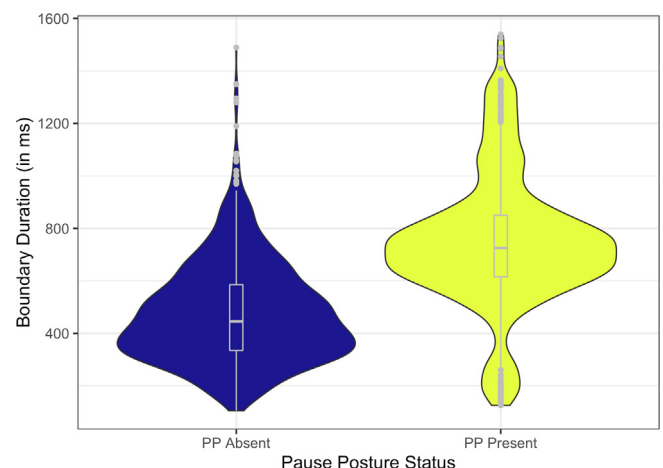


Fig. 10. The effect of boundary duration on the occurrence of pause posture, all speakers combined.

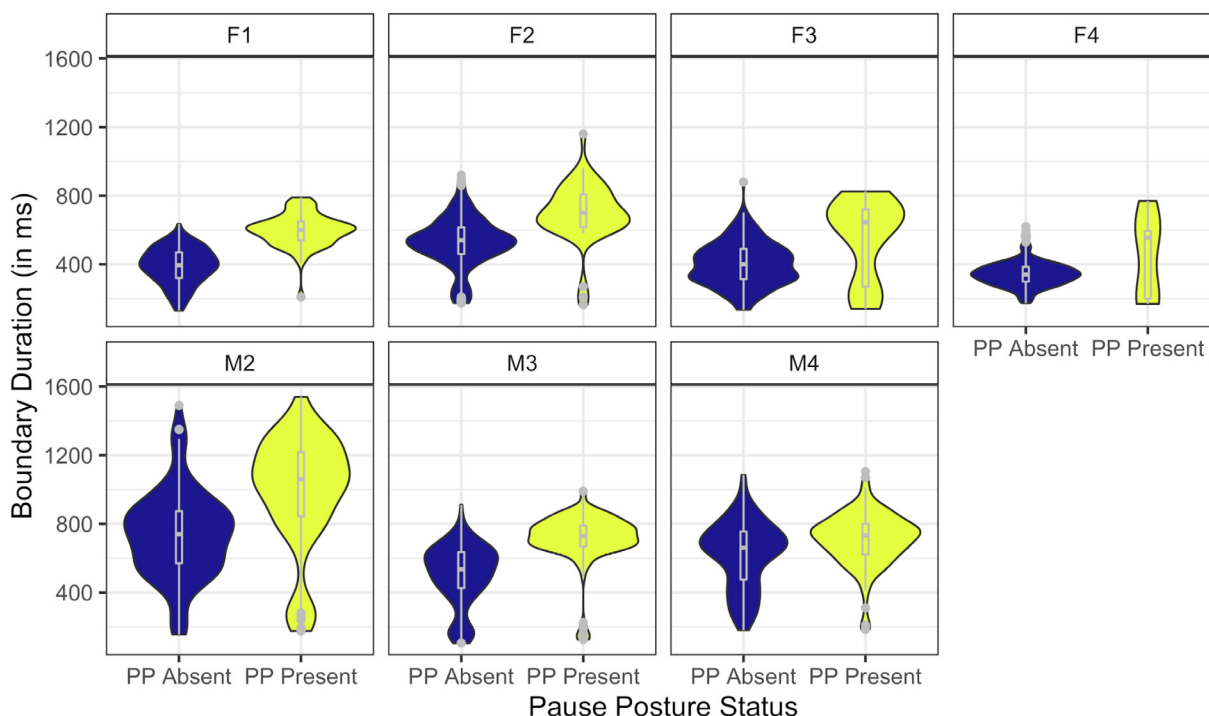


Fig. 11. The effect of boundary duration on the occurrence of pause posture, individual speakers.

tion. This result also provides further support for Katsika's claim that PP are triggered by strong  $\pi$ -gestures and therefore occur at strong boundaries (Katsika et al., 2014).

**3.3.2.2. The effect of upcoming phrase length on pause posture occurrence.** We test the hypothesis that pause postures are more likely to occur when the upcoming phrase is longer, to allow for additional planning time. A GLM was fitted to the data testing the effect of number of upcoming (post-boundary) syllables (shown in Table 1) of the first upcoming prosodic phrase on pause posture likelihood (see Appendix I, 4. Upcoming Phrase Length for full model output). Note that the 2-syllable, 3-syllable, and the 8-syllable condition consisted of one sentence type only. The results for all speakers are shown in Figs. 12 and 13. The model indicates that larger upcoming phrases increase the likelihood of PP occurrence. This is significant for the pooled data ( $p < 0.0001$ ) as well as for four speakers (F1, M2, M3, M4) individually (at  $p < 0.05$ ). The result is not as clear as other results, as in particular the longest upcoming phrase, which had eight syllables, consistently did not have the effect that would be expected, namely the highest number of PPs. However, the trend for the other sentences is relatively stable. There is some speaker variability, in that F2, F3, and F4 do not show a significant effect, but, while this in itself would not be surprising given that individual differences in planning are to be expected (Petrone, Fuchs, & Krivokapić, 2011; Swets, Desmet, Hambrick, & Ferreira, 2007; Swets, Jacovina, & Gerrig, 2014), this might also be the result of these speakers having many fewer PPs than other speakers.

While the results of upcoming phrase length indicate that planning has an effect on pause posture occurrence, we have also found pause posture occurrence to be associated with strong boundaries (long boundary duration), and these are known to be more likely to occur when the preceding phrase

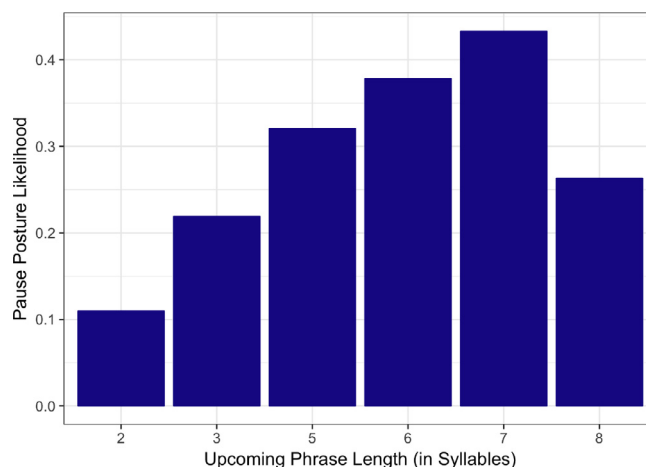


Fig. 12. The effect of upcoming phrase length (in syllables) on the occurrence of pause posture, all speakers pooled.

is long, and when the following phrase is long (e.g., Fuchs et al., 2013; Krivokapić 2007a, 2007b; Zvonik & Cummins, 2003). To evaluate the relative contributions of preceding phrase length (number of syllables preceding the boundary), boundary duration (interval 4 in Fig. 2), and upcoming phrase length (number of syllables of the first prosodic phrase following the boundary), and to test to ensure that the upcoming phrase length effect is not simply a function of boundary duration, we ran a series of nested model comparisons.<sup>7</sup>

<sup>7</sup> Studies examining planning have typically examined the effect of linguistic structure on pause duration. We instead examine boundary duration, an interval that consists of the acoustic pause and the pre-boundary and post-boundary articulations. We take this interval to be at least equally well suited for examining the effect of planning as acoustic pauses are, if not better, given that including the pre-boundary and post-boundary articulations more accurately reflects the boundary.

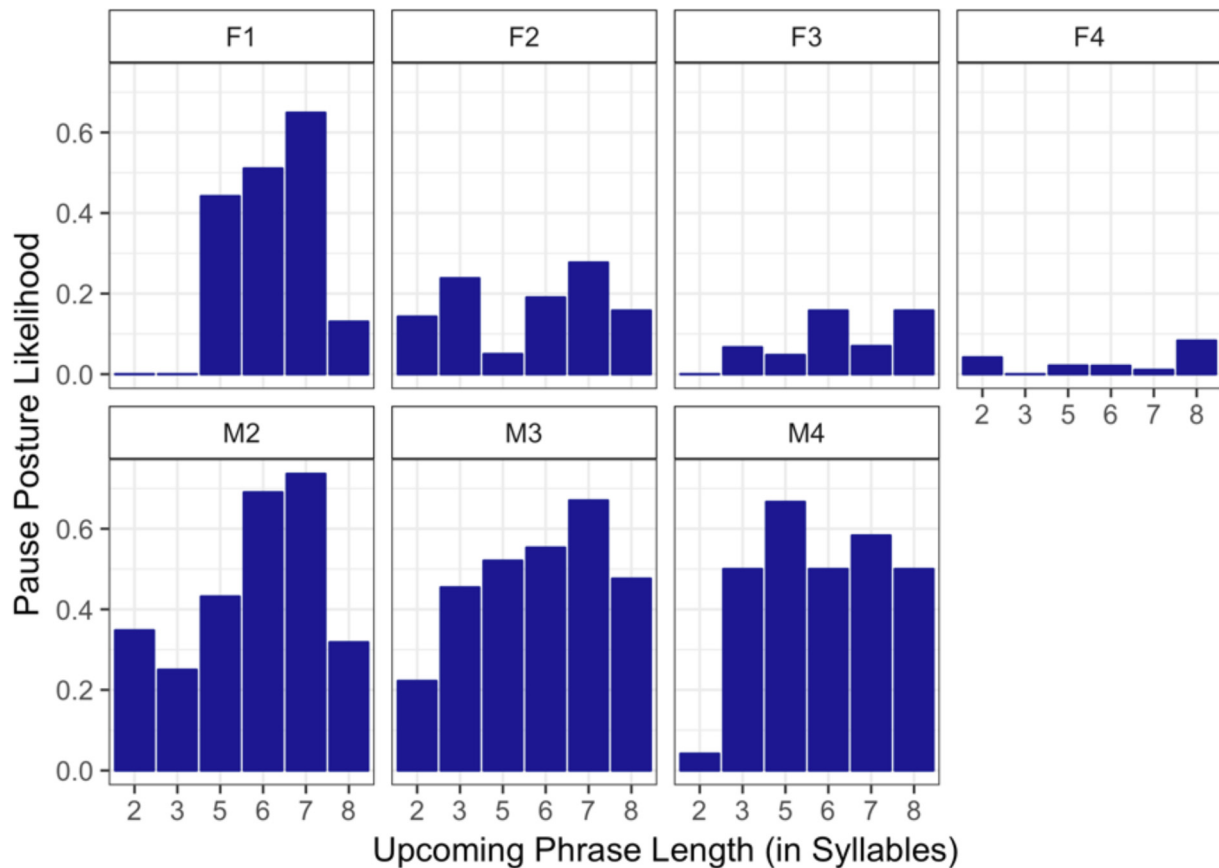


Fig. 13. The effect of upcoming phrase length (in syllables) on the occurrence of pause posture, individual speakers.

To compare all three components that we hypothesized could contribute to increased boundary duration and pause posture occurrence, we fit a GLM to evaluate pause posture likelihood by boundary duration and upcoming and preceding phrase length (Appendix I, Table 5a). Pause postures were more likely at longer boundaries ( $p < 0.0001$ ) and before longer upcoming prosodic phrases ( $p = 0.013$ ). No significant effect of pre-boundary length on pause posture likelihood was found. For model comparison purposes, a second, nested model, including only boundary duration and upcoming length (Appendix I, Table 5b) was fitted and found similar results (boundary duration,  $p < 0.0001$  and upcoming phrase length,  $p = 0.005$ ). A model including only boundary duration (Appendix I, Table 5c) also shows that pause postures are more likely as boundary duration increases ( $p < 0.0001$ ). These results indicate that boundary duration and upcoming length are both significant in determining pause posture occurrence.

To further test this, two model comparisons are conducted using the *anova()* function in R with a chi-square test. Model comparison between the three-parameter model (boundary duration, pre-boundary phrase length and upcoming phrase length) and the two-parameter model (boundary duration and upcoming phrase length) (Appendix I, Tables 5d and e, respectively) shows that there is no difference in fit ( $p(\text{difference}) = 0.50$ ) indicating that, when boundary duration and upcoming phrase length are present, pre-boundary phrase length does not contribute to the fit when accounting for the increased complexity of the model. However, a similar model comparison

between the two-parameter model (boundary duration and upcoming phrase length) and a one-parameter model (boundary duration alone), found that there is a significant improvement in fit associated with the two-parameter model ( $p = 0.005$ ). This indicates that even taking into account the increased complexity, the inclusion of upcoming phrase length improves fit substantially in comparison to boundary duration alone.

These results indicate that boundary duration has a strong effect on pause posture occurrence, but upcoming phrase length also has a meaningful and independent effect.

**3.3.2.3. The effect of upcoming phrase length on pause posture duration.** If planning takes place during PPs, the upcoming phrase length will have an effect on PP duration, such that PPs will be longer if the upcoming phrase is longer. We conducted a series of correlations (using the *lm* function in R) to examine this question.

We also need to take into account the possibility that any effect of upcoming phrase length on pause posture duration is due to a potential effect of upcoming phrase length on boundary duration. We therefore tested the correlation between boundary duration and upcoming phrase length (Appendix I, Table 6a). As expected, given the results reported here and in previous research, the overall boundary length is positively correlated with the length of the upcoming phrase in tokens with pause postures phrase ( $R^2 = 0.008$ ,  $p = 0.013$ ) and in tokens without pause postures (Appendix I,

Table 6b,  $R^2 = 0.029$ ,  $p < 0.0001$ ). Although this is a small effect, it remains significant despite large variability in the presence of pause postures within and across speakers, the fact that these were read rather than spontaneously produced sentences (and so had less planning demands), and the fact that the stimuli used were not explicitly designed to test the effects of planning. Overall, this result indicates that we need to take the relationship between pause posture duration and boundary duration into account as we examine the relationship between upcoming phrase length and PP duration.

We further tested the correlation between pause posture duration and upcoming phrase length (Appendix I, Table 6c). Pause posture duration is also positively correlated with upcoming phrase length ( $R^2 = 0.007$ ,  $p = 0.022$ ). The nature of this relationship, with regression line, is shown in Fig. 14.

For this correlation to be meaningfully interpretable, we must address the concern that the relationship between pause posture duration and planning may be an artifact of the boundary duration effect, as increasing the length of the boundary might increase the length of the sub-intervals within that boundary. Fig. 15 shows a histogram of the proportions of each boundary occupied by the pause posture across all tokens. It can be seen here that the relationship between the two is not the same across all boundary tokens and that the pause posture does not necessarily occupy most of the boundary. Therefore, the effect of the upcoming phrase might have an independent effect on the pause posture duration.

Although differences in the slope coefficients associated with upcoming length in the analysis above are suggestive of independence of these features (16.45 for boundary duration, 11.02 for pause posture duration), to further test for the independence of pause posture duration from overall boundary length, we conducted two additional correlations, examining the correlation between the pre-pause-posture duration (duration between the start of the boundary and the start of the pause posture, output in Appendix I, Table 6d), and the post-pause-posture duration (the duration between the end of the pause posture and the end of the boundary, output in Appendix I, Table 6e) with upcoming phrase length.

There was no link ( $R^2: 0$ ,  $p = 0.906$ ) between the length of the pre-pause-posture period and upcoming phrase length.

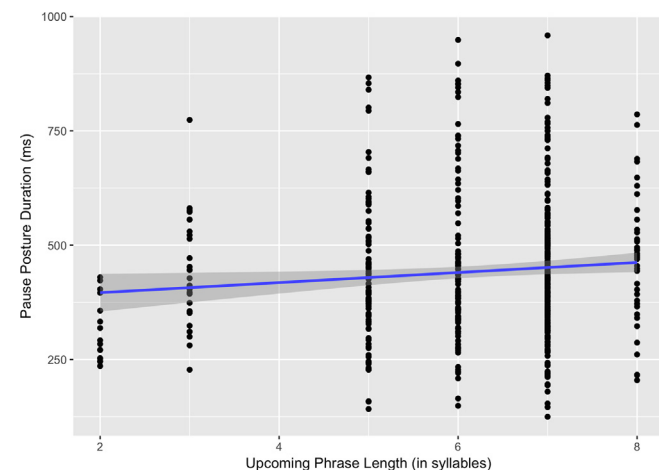


Fig. 14. Pause Posture Duration vs. Upcoming Phrase Length, with lm-based regression line and model-derived 95% confidence intervals.

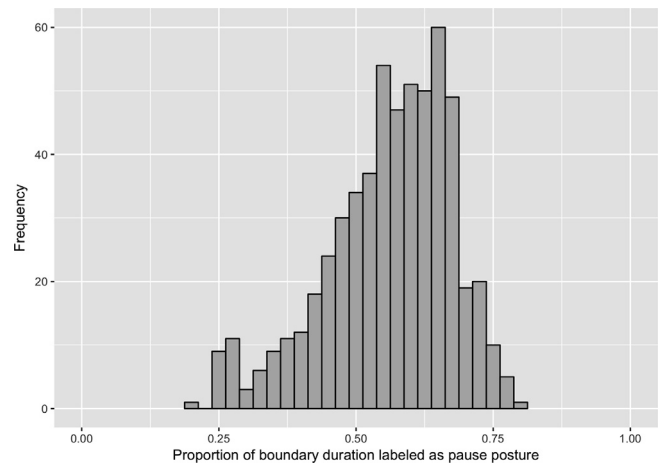


Fig. 15. The percentage of boundary duration occupied by the pause posture.

Given that the timing of the pause posture has been argued to be at a stable relationship with the boundary tone, it is not surprising that the interval is not related to upcoming phrase length. More importantly, we also found no link between the post-pause-posture duration and upcoming phrase length ( $R^2 = 0.002$ ,  $p = 0.244$ ). The fact that only the pause posture, of the three sub-intervals measured, shows any effect of the upcoming phrase length suggests that pause postures are lengthened to accommodate additional time for increased planning demands. The duration of the movements into and out of the pause posture on the other hand are relatively stable and unrelated to increase in cognitive demands.

3.3.2.4. *The effect of repetition on pause posture occurrence and duration.* Finally, the question arises whether properties of planning change in the course of the experiment.<sup>8</sup> We do not have predictions but the goal is to contribute empirical data to this matter. While it might be expected that planning pauses will decrease with repetition of the same material (e.g., Goldman Eisler, 1968, see also Ferreira & Karimi, 2015), reading these sentences is not a trivial task—it is a bit boring and the participants have a large number of sensors on their tongue and face. These take some time to attach, so the experiment is also long. Thus, while the default expectation might be that with repetition the amount of planning time decreases, it is equally likely that the task itself becomes harder over time (due to, for example, boredom). We examined this question both in terms of the relative likelihood of pause postures, and in terms of their overall duration. In terms of likelihood, using a GLM model, we found no overall effect of repetition (modeled as number) on the likelihood of PP ( $p = 0.143$ ) (see Appendix I, 7. Repetition vs. Pause Posture Likelihood for full model output). For two of the speakers individually, we do see effects of repetition, but in conflicting directions, with F2 showing a negative trend (i.e., later repetitions had fewer PPs) and M2 showing a positive trend (both  $p < 0.05$ ). These two speakers, along with the other speakers, are displayed in Fig. 16.

<sup>8</sup> We thank an anonymous reviewer and the associate editor, Marianne Pouplier, for bringing this point up.



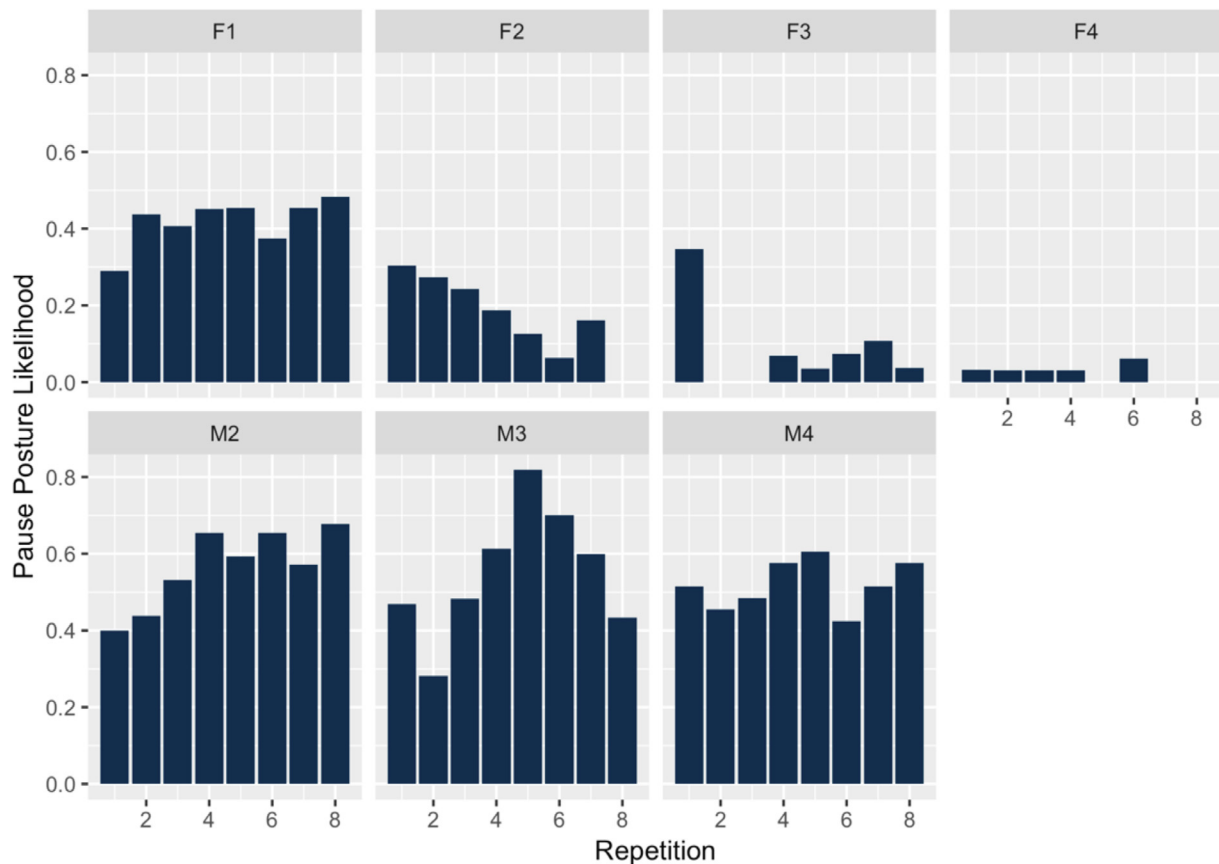


Fig. 16. Pause posture likelihood by repetition across speakers.

For the effect of repetition on pause posture duration, again modeled using a GLM, there was no overall effect ( $p = 0.437$ ), and again, there were contradictory effects among four of the seven speakers (see Appendix I, 8. *Repetition vs. Pause Posture Duration* for full model output). Two speakers (F2 and F4) showed a negative trend (i.e., later repetitions had shorter PPs), and the other two (M2 and M4) showed positive trends (all at  $p < 0.05$ ). These by-speaker trends are displayed in Fig. 17. Note that the results for F2 and M2 go in the same direction in the two comparisons, indicating that PP occurrence and duration are affected in the same manner by repetition.

The lack of overall effect suggests that, where present, the effect of repetition is idiosyncratic, with some speakers showing increases in PP duration and likelihood, others showing the opposite pattern, and others showing no effect at all. Thus, although it's clear that repetition has some effects on the phenomenon, those effects are neither predictable nor uniform.

#### 4. Discussion

##### 4.1. The existence and properties of pause postures

The analyses of the seven speakers found that pause postures occur in American English. All speakers had some number of PPs, but the frequency with which they occur is speaker dependent (Table 4), as has been reported in other studies (Schaeffler et al., 2008; Rasskazova et al., 2018). The target words differed in the frequency of pause postures, with *biBU*

having more pause postures than the other two target words. This is likely related to the articulatory properties of the final vowel: while we can't tell what the exact target for the lips during a PP is, the current results suggest that it is a position towards lip closure, and thus the final vowel of *biBU* is closer to the target position for PPs (with LA moving towards closed lips) than the final vowels in *Mlma* and *miMA*. Therefore it is more likely that the LA approaches the target for the PP after *biBU* than after other target words. While it is unclear how this would lead to more PPs, since PPs were identified in part as a deviation from an interpolation line (and not by how closely they approached a target), we expect the explanation to be related to this, as no other explanation seems likely.

As a step to examine the cognitive status of pause postures, we examined the timing of PPs to other linguistic events at the boundary (following the analysis in Katsika et al. (2014) for Greek). As reported for Greek, we find for American English that stress had an effect on the pause posture formation duration and on the boundary-tone to V-target interval duration, such that the intervals are longer when stress is on the second (*miMA*) than when it is on the first syllable (*Mlma*). Additionally, there was no effect of stress on the boundary-tone to PP-target interval. We also found that longer boundaries lead to more pause postures. The results of this study thus parallel the findings in Katsika et al. (2014) and provide further evidence for the account developed there. More importantly for our study, they show that the pause posture in American English behaves predictably in relationship to other linguistic events, providing

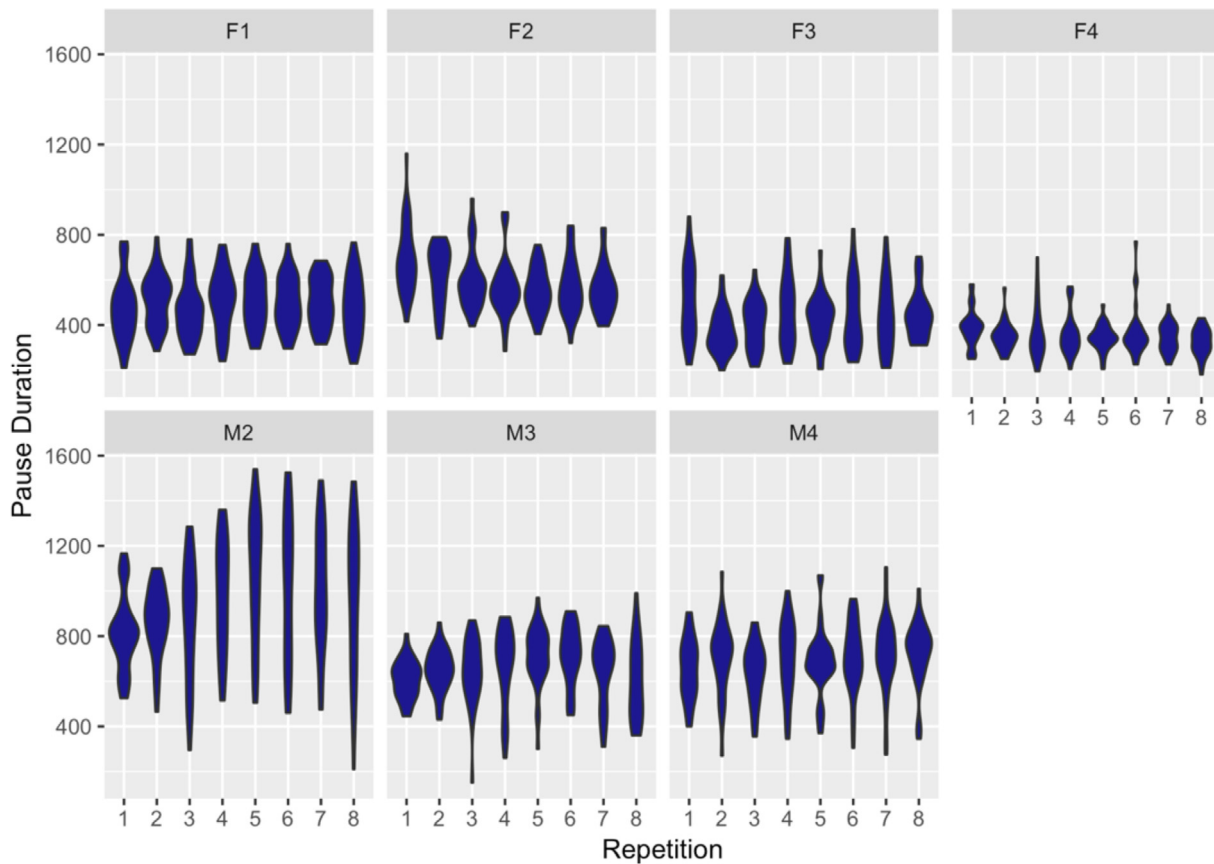


Fig. 17. Pause Posture Duration by Repetition.

support for the hypothesis that the pause posture is a cognitive, controlled gesture.

The spatial variability of pause postures is comparable to that of other controlled targets (specifically C1) for longer PPs. However, PPs show more variability than C1 when they are shorter. We suggest that this is related to undershoot, i.e., that the articulators do not reach the target in shorter PPs. Large spatial variability is also plausible if the PP occurrence and duration are related to planning, since the more planning time is needed, the longer the PP will be and the more time it will have to reach its target—this variability in needed planning time will thus necessarily lead to spatial variability. We also found that the PP following *biBU* is somewhat more stable than for the other two target words. The explanation for this difference is also related to undershoot: since the final vowel of *biBU* is closer to the target for the PP than the final vowels in *Mima* and *miMA*, the PP after *biBU* is more likely to achieve the target, and therefore there will be less variability. Overall then, the results suggest that PPs have a controlled target.

Finally, we see evidence that pause postures are not just recognizable by human annotators, but that these kinds of targeted gestures can be recovered and classified using fPCA curve-modeling and a Support Vector Machine classifier, without any intermediate steps or gesture-specific modeling. This not only bolsters our claim as to the existence of the phenomenon, but provides a relatively fast and lightweight method

by which other gestures can be detected and measured without relying solely on human annotation.

To conclude, the results of the first set of analyses strongly suggest that PPs exist in American English and that they are cognitive units.

#### 4.2. Factors determining the occurrence of pause postures

The main question the study addressed was what causes the occurrence of PPs, in other words, how are they related to cognitive processes. We examined the effects of discourse and of planning.

There was an effect of sentence for five out of seven speakers but the effects did not pattern in a similar way, indicating that there is no consistent effect of the discourse-pragmatic factors tested here on pause postures. However, the significant effect for most speakers individually suggests that different discourse boundaries do differentially affect the frequency of pause postures, but that individuals differ in the specifics of this implementation. These results are consistent with previous work that showed similar between-speaker variability in kinematic effects of prosodic boundaries (Byrd & Saltzman, 1998; Byrd, Krivokapić, & Lee, 2006; Parrell et al., 2013). One caveat is that the utterances were read, not spoken in a dialogue. There thus remains the possibility that there is an effect of discourse, but that our study was not designed to test it, as the sentences were read without a

co-speaker. Further studies will need to address this question in more detail.

One interesting finding is that PPs occur both utterance-initially and utterance-finally. These PPs differ kinematically from the utterance-medial ones (due to the lack of preceding or following material), and phonetic evidence in [Ramanarayanan et al. \(2013\)](#) suggest that articulatory settings before speech onset have different properties than utterance-medial ones. These PPs also might have different cognitive functions. The present study thus provides further evidence that articulation does not just start or stop with the first or last audible gesture, and future work will need to investigate the factors conditioning these specific utterance-initial and utterance-final articulations and their cognitive functions.

Finally, the role of planning in the occurrence and duration of PPs was examined, starting from the well-supported idea in the literature that longer upcoming phrases lead to an increase in pause duration because speakers need more time to plan the upcoming phrase. The data in our study show that upcoming phrase length has an effect on PP occurrence, such that, in general, longer upcoming phrases lead to more PPs, and this held true even when controlling for overall boundary duration, as model comparisons showed. One caveat, as mentioned before, is that the longest upcoming phrase (eight syllables long) did not have the highest number of pause postures. As only one sentence had eight syllables, this might be a confound with some other property of the sentence, though it is not clear what that could be. The length of the preceding phrase did not have an effect on the occurrence of PPs. Overall, these results suggest that the occurrence of PPs may be related to planning.

The question arises in what manner the PP can be related to planning, and our hypothesis states that speakers plan the upcoming utterance during the PP. Given that not all pauses have PPs, we must assume that planning does not have to take place exclusively during PPs, but instead, that speakers produce a PP *when they need additional planning time*. We found some evidence for this planning hypothesis, in that the duration of the PP was significantly correlated with upcoming phrase length, while the other parts of the boundary (the part before the PP and the part after the PP) were not correlated with upcoming phrase length. Thus, the boundary duration overall does not increase in all parts equally, and it is only the PP that is correlated with upcoming phrase length. This effect is small, which is not surprising given that our present data were not designed to examine this question specifically, and that these were read sentences, but it also indicates that other factors might also affect the duration of the PP. Nevertheless, the results provide evidence that pause postures might occur as a part of the upcoming utterance planning process under some circumstances – perhaps, as we suggest, when extra planning time is needed. It is unclear whether this means that planning occurs solely during the PP when one is present, or that planning takes place during the whole boundary and the PP just provides additional time. The first option would mean that there are different planning strategies even within a speaker, such that they plan differently depending on whether the PP is present or absent, a possibility we consider unlikely. If planning were concentrated in one part of the boundary when a PP is present, this would provide some support for Fer-

reira's strong hypothesis that pauses are partitioned into parts with different cognitive functions, though the partitioning would be different than suggested in her work, with only the PP serving as a predominantly planning part, rather than the whole second part of the boundary. However, given that not all speakers have PPs and that, as expected, boundary duration in utterances without PPs are also correlated with upcoming phrase length, it is clear that planning does not have to take place during PPs only, and future research needs to investigate the time-course of planning during boundaries in more detail. Note that we are not specifying what aspect of the upcoming utterance is planned during the PP—while we tested for syllable number, indicating phonetic and phonological planning, any type of planning could in principle take place here.

One question that arises is why, if speakers need additional planning time, i.e., if the upcoming phrase is complex enough to require extra time, speakers would *add* another gesture, which presumably is in some sense costly, in that it for example requires extra planning.<sup>9</sup> One reason might be that, in lip aperture, this is a visible gesture to listeners, and thus, in addition to providing the planning time, it could signal to listeners that this will be a longer pause, and could be a “hold the floor” indicator. Alternatively, the increased pause duration needed for planning may allow for the emergence of the pause posture via the “neutral attractor” which draws the articulators towards a neutral position in the absence of linguistic constriction gestures. Which of these views is accurate is an open question, as the predictions in timing relations are similar. In one view, the PP is the emergence of the neutral attractor (which is always active in the absence of other gestures) at long pause durations used for speech planning, in the other it is an additional gesture. At the moment, without additional evidence, not assuming an additional gesture might be a better option, but this is an open question. Results from [Ramanarayanan et al. \(2013\)](#) indicate that articulatory behavior at boundaries differs from articulatory behavior during other silences (the beginning of an utterance for example) and this might be evidence for a gesture different from a setting resulting from a neutral attractor.

Note that the interval between the offset of the pause posture and the onset of the post-boundary consonant is not affected by the post-boundary utterance, potentially indicating that the PP has a stable temporal relationship to the onset of the post-boundary consonant. This could indicate that the  $\pi$ -gesture for the next phrase is coordinated with the PP. Or, in a similar vein, it could be the first lexical item of the following phrase that is coordinated with the PP (possibly the PP could be coordinated with the first stressed syllable for example). This would allow flexibility in planning and would also account for the large variability in the production of boundaries: Depending on the utterance, the planning strategies of a speaker, the amount of planning they do at boundaries, and other factors influencing planning, boundaries could have a longer or shorter PP and the following utterance would be timed to the PP. This, as well, requires further research.

To summarize: The present study examined the potential cognitive factors underlying the occurrence and properties of pause postures. We have established that pause postures, as have been identified for Greek in [Katsika et al. \(2014\)](#), occur

<sup>9</sup> We thank an anonymous reviewer for bringing this point up.

in American English as well, and are timed to other gestures at boundaries in a manner that supports the model suggested in Katsika et al. (2014). These results provide evidence for the suggestion that PPs are controlled, cognitive units. In addition, we observe that pause postures occur utterance-initially and utterance-finally as well. We also established that PPs are identifiable using basic machine learning and curve description techniques. The occurrence of PPs is speaker specific, but not utterance specific across speakers, and their occurrence utterance-medially appears related to the planning of the upcoming utterance. While the evidence is far from conclusive, we suggest that PPs might provide additional time for speakers to plan the upcoming utterance, and suggest that additional research into the interaction of planning, pause postures, and prosodic boundaries is merited.

#### Declaration of Competing Interest

The authors declare that they have no conflict of interest.

#### CRediT authorship contribution statement

**Jelena Krivokapić:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Will Styler:** Conceptualization, Methodology, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision. **Benjamin Parrell:** Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing.

#### Acknowledgments

We are grateful to Sungbok Lee, Jiseung Kim, Dolly Goldenberg, Anna Stone, Hayley Heaton, Kate Sherwood, Argyro Katsika, Stephen Tobin, Mariko Ito, members of the Michigan Phonetics Laboratory, and especially Dani Byrd for help with this study. Two anonymous reviewers and the Associate Editor, Marianne Pouplier, have also provided insightful comments that have greatly improved the paper. This work was supported by NIH DC003172 to Dani Byrd, NIH DC002717 to Doug Whalen, and NSF 1551513 to Krivokapić, and the Michigan Phonetics Laboratory.

#### Appendix I. Full regression outputs for all models

##### 1. Quantile regression

Using quantreg:  $\text{rq}(\text{spatial\_mm} \sim \text{constriction} * \text{targetword} * \text{zdur}, \text{tau} = \text{c}(0.25, 0.75), \text{data} = [\text{dataset}], \text{method} = \text{"fn"})$

1a: Quantile Regression at 25th Percentile:

	Value	Std. error	t value	Pr(> t )
(Intercept)	-0.392	0.0805	-4.868	0
constrictionpp_mm	-1.2	0.1314	-9.1329	0
targetwordmiMA	-0.061	0.1067	-0.5714	0.5679
targetwordMIma	-0.3372	0.1103	-3.0564	0.0023
zdur	0.0781	0.0744	1.0485	0.2947
constrictionpp_mm:targetwordmiMA	-0.1766	0.2168	-0.8146	0.4155
constrictionpp_mm:targetwordMIma	0.1567	0.2447	0.6407	0.5219
constrictionpp_mm:zdur	-0.1135	0.1209	-0.9384	0.3482
targetwordmiMA:zdur	0.012	0.1074	0.1119	0.911
targetwordMIma:zdur	-0.0163	0.1094	-0.1491	0.8815
constrictionpp_mm:targetwordmiMA:zdur	0.0547	0.2262	0.2419	0.8089
constrictionpp_mm:targetwordMIma:zdur	-1.369	0.2195	-6.2381	0

(Note:  $p = 0$  represents model internal rounding)

1b: Quantile Regression at 75th Percentile:

	Value	Std. error	t value	Pr(> t )
(Intercept)	0.6713	0.0799	8.4072	0
constrictionpp_mm	-0.4969	0.1944	-2.5557	0.0107
targetwordmiMA	-0.1053	0.1064	-0.9896	0.3226
targetwordMIma	-0.6158	0.1045	-5.8915	0
zdur	0.0121	0.0736	0.1645	0.8693
constrictionpp_mm:targetwordmiMA	3.1841	0.2649	12.021	0
constrictionpp_mm:targetwordMIma	2.4523	0.3736	6.5649	0
constrictionpp_mm:zdur	-0.5759	0.1732	-3.3254	9e-04
targetwordmiMA:zdur	-0.0475	0.1017	-0.467	0.6406
targetwordMIma:zdur	0.0536	0.1052	0.5098	0.6103
constrictionpp_mm:targetwordmiMA:zdur	-0.3184	0.2554	-1.2466	0.2128
constrictionpp_mm:targetwordMIma:zdur	-0.426	0.3115	-1.3673	0.1718



## 2. Sentence effect

Unless otherwise stated, models will be presented first as overall (pooled) results, then by speaker.

`glm(svm_pausebinary ~ sentence, family = binomial, data = [dataset])`

(by speaker models are identical, but run on single-speaker subsets of the data)

Overall:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−0.9483	0.1796	−5.2795	0
sentence2	0.6162	0.2414	2.5523	0.0107
sentence3	−0.0813	0.2573	−0.316	0.752
sentence4	−0.3221	0.2689	−1.1981	0.2309
sentence5	−1.1457	0.3136	−3.6537	3e-04
sentence6	0.1561	0.2501	0.6242	0.5325
sentence8	0.5746	0.2422	2.3728	0.0177
sentence9	0.3763	0.2444	1.5398	0.1236
sentence10	0.7195	0.2407	2.9897	0.0028
sentence11	0.5325	0.2412	2.2073	0.0273
sentence12	0.9861	0.2396	4.1147	0
sentence13	0.0408	0.2506	0.1627	0.8707
sentence14	−0.1256	0.2549	−0.4926	0.6223

F1:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−0.8473	0.488	−1.7364	0.0825
sentence2	0.6802	0.6371	1.0677	0.2857
sentence3	−1.0498	0.7883	−1.3317	0.1829
sentence4	−17.7188	1331.4281	−0.0133	0.9894
sentence5	−17.7188	1331.4281	−0.0133	0.9894
sentence6	−0.4336	0.7026	−0.6172	0.5371
sentence8	0.7603	0.6421	1.184	0.2364
sentence9	1.1097	0.6442	1.7225	0.085
sentence10	2.1823	0.7005	3.1153	0.0018
sentence11	2.4567	0.7335	3.3491	8e-04
sentence12	2.4054	0.7353	3.2712	0.0011
sentence13	−1.5041	0.8864	−1.6968	0.0897
sentence14	0.1542	0.6524	0.2363	0.8132

M3:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	0.1671	0.4097	0.4078	0.6834
sentence2	0.4616	0.5996	0.7698	0.4414
sentence3	−0.2624	0.599	−0.438	0.6614
sentence4	−0.3494	0.5926	−0.5896	0.5555
sentence5	−1.4198	0.6995	−2.0298	0.0424
sentence6	0.0953	0.5872	0.1623	0.8711
sentence8	1.3911	0.6859	2.0281	0.0425
sentence9	−0.1671	0.5784	−0.2888	0.7727
sentence10	0	0.5794	0	1
sentence11	−0.1671	0.5913	−0.2825	0.7775
sentence12	0.6596	0.6109	1.0798	0.2802
sentence13	−1.2657	0.6245	−2.0265	0.0427
sentence14	−2.5649	0.8443	−3.038	0.0024

F2:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−18.5661	1458.5063	−0.0127	0.9898
sentence2	17.1191	1458.5064	0.0117	0.9906
sentence3	16.8921	1458.5065	0.0116	0.9908
sentence4	17.4029	1458.5064	0.0119	0.9905
sentence5	16.7743	1458.5065	0.0115	0.9908
sentence6	0	2037.9363	0	1
sentence8	17.6498	1458.5064	0.0121	0.9903
sentence9	16.3688	1458.5065	0.0112	0.991
sentence10	18.0806	1458.5064	0.0124	0.9901
sentence11	17.4675	1458.5064	0.012	0.9904
sentence12	18.0806	1458.5064	0.0124	0.9901
sentence13	16.7743	1458.5065	0.0115	0.9908
sentence14	18.4607	1458.5064	0.0127	0.9899

F4:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−3.1355	1.0215	−3.0695	0.0021
sentence2	−17.4306	3619.1967	−0.0048	0.9962
sentence3	0.7376	1.2605	0.5852	0.5584
sentence4	−17.4306	3619.1967	−0.0048	0.9962
sentence5	0	1.4446	0	1
sentence6	−17.4306	3619.1967	−0.0048	0.9962
sentence8	0	1.4446	0	1
sentence9	−17.4306	3697.0378	−0.0047	0.9962
sentence10	0	1.4446	0	1
sentence11	−17.4306	3619.1967	−0.0048	0.9962
sentence12	−17.4306	3780.1277	−0.0046	0.9963
sentence13	1.6314	1.1615	1.4046	0.1601
sentence14	2.2482	1.1159	2.0147	0.0439

M2:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−0.5596	0.4432	−1.2627	0.2067
sentence2	1.6582	0.647	2.5628	0.0104
sentence3	−0.2025	0.6371	−0.3179	0.7506
sentence4	−0.539	0.7278	−0.7405	0.459
sentence5	−0.069	0.623	−0.1107	0.9118
sentence6	1.0451	0.6312	1.6559	0.0977
sentence8	2.6997	0.869	3.1066	0.0019
sentence9	0.5596	0.615	0.9099	0.3629
sentence10	1.7228	0.6774	2.543	0.011
sentence11	0.7267	0.6035	1.204	0.2286
sentence12	1.8946	0.6701	2.8273	0.0047
sentence13	1.6582	0.647	2.5628	0.0104
sentence14	−1.8383	0.8611	−2.1349	0.0328

M4:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	0.5108	0.4216	1.2115	0.2257
sentence2	0.1823	0.6044	0.3017	0.7629
sentence3	−0.5108	0.5869	−0.8704	0.3841
sentence4	−0.5108	0.5869	−0.8704	0.3841
sentence5	−3.6463	1.1049	−3.3001	0.001

## Appendix 1 (continued)

	Estimate	Std. error	z value	Pr(> z )
sentence6	−0.3438	0.5879	−0.5848	0.5587
sentence8	−0.8473	0.5909	−1.4338	0.1516
sentence9	0.3765	0.616	0.6112	0.5411
sentence10	−0.6779	0.5879	−1.1531	0.2489
sentence11	−0.1744	0.5909	−0.295	0.768
sentence12	0.1823	0.6044	0.3017	0.7629
sentence13	−1.5523	0.635	−2.4444	0.0145
sentence14	−2.8134	0.8531	−3.2979	0.001

F3:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−19.5661	2404.6704	−0.0081	0.9935
sentence2	17.4866	2404.6705	0.0073	0.9942
sentence3	17.8921	2404.6705	0.0074	0.9941
sentence4	16.927	2404.6706	0.007	0.9944
sentence5	0	3359.9889	0	1
sentence6	18.3133	2404.6705	0.0076	0.9939
sentence8	0	3287.955	0	1
sentence9	17.2635	2404.6705	0.0072	0.9943
sentence10	17.3688	2404.6705	0.0072	0.9942
sentence11	16.475	2404.6706	0.0069	0.9945
sentence12	17.7202	2404.6705	0.0074	0.9941
sentence13	18.7394	2404.6704	0.0078	0.9938
sentence14	19.399	2404.6704	0.0081	0.9936

## 3. Boundary duration

```
glm(svm_pausebinary ~ pausedur,family = binomial,data = [dataset])
```

(by speaker models are identical, but run on single-speaker subsets of the data)

Overall:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−4.0128	0.1936	−20.7282	0
pausedur	0.0056	3e-04	18.537	0

F1:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−11.8175	1.4348	−8.2365	0
pausedur	0.0227	0.0028	8.2026	0

M3:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−5.3871	0.7981	−6.75	0
pausedur	0.0087	0.0012	7.1911	0

F2:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−6.3823	0.9351	−6.8254	0
pausedur	0.008	0.0014	5.6728	0

F4:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−6.7142	1.2621	−5.3201	0
pausedur	0.0087	0.0028	3.0587	0.0022

M2:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−2.0212	0.4218	−4.7917	0
pausedur	0.0027	5e-04	5.8119	0

M4:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−1.8286	0.5038	−3.6297	3e-04
pausedur	0.0028	7e-04	3.8069	1e-04

F3:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−5.426	0.8353	−6.4956	0
pausedur	0.0069	0.0015	4.494	0

## 4. Upcoming phrase length

```
glm(svm_pausebinary ~ upcoming_phrase_length,family = binomial,data = [dataset])
```

(by speaker models are identical, but run on single-speaker subsets of the data)

Overall:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−1.9554	0.2005	−9.7514	0
upcoming_phrase_length	0.2202	0.0323	6.8101	0

F1:

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	−0.1304	0.0979	−1.3323	0.1839
upcoming_phrase_length	0.0961	0.0163	5.8853	0

M3:

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	0.2257	0.1119	2.0175	0.0447
upcoming_phrase_length	0.0562	0.0186	3.0251	0.0028

F2:

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	0.0906	0.0892	1.0156	0.3109
upcoming_phrase_length	0.0182	0.0149	1.2211	0.2233



F4:

	Estimate	Std. error	<i>t</i> value	Pr(>  <i>t</i>  )
(Intercept)	0.0112	0.0315	0.3564	0.7218
upcoming_phrase_length	0.0021	0.0053	0.3905	0.6965

M2:

	Estimate	Std. error	<i>t</i> value	Pr(>  <i>t</i>  )
(Intercept)	0.1986	0.1092	1.8195	0.0701
upcoming_phrase_length	0.0629	0.018	3.4857	6e-04

M4:

	Estimate	Std. error	<i>t</i> value	Pr(>  <i>t</i>  )
(Intercept)	0.2022	0.1029	1.9651	0.0505
upcoming_phrase_length	0.0553	0.0172	3.2197	0.0014

F3:

	Estimate	Std. error	<i>t</i> value	Pr(>  <i>t</i>  )
(Intercept)	-0.0128	0.064	-0.1997	0.8419
upcoming_phrase_length	0.0163	0.0106	1.5364	0.1259

## 5. Boundary Duration vs. Upcoming vs. Preceding Length

```
threelm = glm(svm_pausebinary ~ pausedur + preboundary + upcoming_phrase_length,family = binomial,data = [dataset])
```

5a: Boundary Duration vs. Upcoming vs. Preceding Length:

	Estimate	Std. error	<i>z</i> value	Pr(>  <i>z</i>  )
(Intercept)	-6.0099	0.4163	-14.4367	0
pausedur	0.0073	4e-04	18.5775	0
preboundary_length	0.0256	0.0385	0.6648	0.5062
upcoming_phrase_length	0.1043	0.0421	2.4758	0.0133

5b: Nested Model: Boundary Duration vs. Upcoming Length:

```
twolm = glm(svm_pausebinary ~ pausedur + upcoming_phrase_length,family = binomial,data = [dataset])
```

	Estimate	Std. error	<i>z</i> value	Pr(>  <i>z</i>  )
(Intercept)	-5.8542	0.342	-17.1173	0
pausedur	0.0073	4e-04	18.6036	0
upcoming_phrase_length	0.1116	0.0406	2.753	0.0059

5c: Nested Model: Boundary Duration Only:

```
onelm = glm(svm_pausebinary ~ pausedur,family = binomial,data = [dataset])
```

	Estimate	Std. error	<i>z</i> value	Pr(>  <i>z</i>  )
(Intercept)	-5.2673	0.2554	-20.6224	0
pausedur	0.0074	4e-04	18.9937	0

5d: Model Comparison ANOVA (3-Way vs. 2-Way):  
 anova(threelm,twelm,test="Chisq")

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	1706	1535.7111	-	-	-
2	1707	1536.1519	-1	-0.4408	0.5067

5e: Model Comparison ANOVA (2-Way vs. Boundary Duration):  
 anova(twelm,onelm,test="Chisq")

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	1707	1536.1519	-	-	-
2	1708	1543.9438	-1	-7.7919	0.0052

#### 6. Boundary Duration vs. Upcoming Phrase Length

6a: Boundary Duration vs. Upcoming Phrase Length (in Sentences with Pause Postures):  
 lm(pausedur ~ upcoming\_phrase\_length,data = [dataset])

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	684.6877	41.7584	16.3964	0
upcoming_phrase_length	16.4539	6.6227	2.4845	0.0133

6b: Boundary Duration vs. Upcoming Phrase Length (in sentences WITHOUT Pause Postures):  
 lm(pausedur ~ upcoming\_phrase\_length,data = [dataset\_no\_pps])

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	404.738	16.6963	24.2411	0
upcoming_phrase_length	16.7744	2.8537	5.8781	0

6c: Pause Posture Duration vs. Upcoming Phrase Length:  
 lm(ppdur ~ upcoming\_phrase\_length,data = [dataset])

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	376.7806	30.8174	12.2262	0
upcoming_phrase_length	11.1603	4.8714	2.291	0.0223

6d: Pre-Pause-Posture Period vs. Upcoming Phrase Length:  
 lm(preppdur ~ upcoming\_phrase\_length,data = [dataset])

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	138.3371	16.5912	8.338	0
upcoming_phrase_length	0.3085	2.6201	0.1177	0.9063

6e: Post-Pause-Posture Period vs. Upcoming Phrase Length:  
 lm(postppdur ~ upcoming\_phrase\_length,data = [dataset])

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	-184.068	22.6541	-8.1251	0
upcoming_phrase_length	-4.1741	3.581	-1.1656	0.2443

## 7. Repetition vs. Pause Posture Occurrence

```
glm(svm_pausebinary ~ repetitionfamily = binomial,data = [dataset])
```

(by speaker models are identical, but run on single-speaker subsets of the data)

Overall:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−0.8111	0.1135	−7.1465	0
repetition	0.0331	0.0226	1.4647	0.143

F1:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−0.6212	0.2855	−2.1761	0.0295
repetition	0.0654	0.0559	1.1685	0.2426

M3:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−0.2887	0.2808	−1.0282	0.3039
repetition	0.111	0.0567	1.9585	0.0502

F2:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−0.5731	0.3484	−1.6447	0.1
repetition	−0.229819.8375	0.0889	−2.5844	0.0098

F4:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−3.0252	0.8119	−3.7262	2e-04
repetition	−0.1764	0.1895	−0.9312	0.3518

M2:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−0.3864	0.2848	−1.3566	0.1749
repetition	0.1454	0.058	2.5085	0.0121

M4:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−0.0478	0.2714	−0.1759	0.8603
repetition	0.0275	0.0538	0.5106	0.6097

F3:

	Estimate	Std. error	z value	Pr(> z )
(Intercept)	−1.5041	0.4861	−3.0939	0.002
repetition	−0.2272	0.1165	−1.95	0.0512

### 8. Repetition vs. Pause Posture Duration

lm(ppdur ~ upcoming\_phrase\_length, data = [dataset])

(by speaker models are identical, but run on single-speaker subsets of the data)

Overall:

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	586.6951	12.7168	46.1354	0
repetition	1.9874	2.5589	0.7767	0.4375

F1:

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	489.7403	17.3385	28.2459	0
repetition	1.0219	3.4287	0.2981	0.7659

M3:

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	643.5199	19.5889	32.8512	0
repetition	2.8133	3.9061	0.7202	0.4721

F2:

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	672.7244	18.8064	35.7711	0
repetition	−19.8375	4.2406	−4.678	0

F4:

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	379.9173	11.3842	33.3723	0
repetition	−5.763	2.2426	−2.5698	0.0107

M2:

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	825.709	36.8525	22.4058	0
repetition	26.7642	7.3489	3.6419	3e-04

M4:

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	653.7987	20.2526	32.2822	0
repetition	8.5877	4.0106	2.1412	0.0332

F3:

	Estimate	Std. error	t value	Pr(> t )
(Intercept)	436.7095	20.7279	21.0687	0
repetition	0.2653	4.0954	0.0648	0.9484



## References

- Bannert, R., Botinis, A., Gawronska, B., Katsika, A., & Sandblom, E. (2003). Discourse structure and prosodic correlates Barcelona, Spain, August 3–9. In *Proceedings of the XVth international congress of phonetic sciences* (pp. 1229–1232). Adelaide: Causal Productions..
- Beckman, M. E., & Ayers Elam, G. (1997). Guidelines for ToBI labeling. (version 3.0). Unpublished ms.
- Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic typology. The phonology of intonation and phrasing* (pp. 9–54). Oxford University Press.
- Boersma P, & Weenink D. (2017). Praat: Doing phonetics by computer [Computer program]. <http://www.praat.org/>.
- Butcher, A. (1981). Aspects of the speech pause: Phonetic correlates and communicative functions. Kiel: Arbeitsberichte, 15. Institut für Phonetik.
- Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English.. *Phonology Yearbook*, 3, 255–309.
- Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of phrase boundary effects. *Journal of the Acoustical Society of America*, 120, 1589–1599.
- Byrd, D., & Saltzman, E. (1998). Intragerural dynamics of multiple phrasal boundaries. *Journal of Phonetics*, 26, 173–199.
- Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31, 149–180.
- Choi, J.-Y. (2003). Pause length and speech rate as durational cues for prosody markers. *Journal of the Acoustical Society of America*, 114, 2395.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37–46.
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.
- Den Ouden, H., Noordman, L., & Terken, J. (2009). Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports. *Speech Communication*, 51, 116–129.
- Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., & Weingessel, A. (2008). Misc functions of the Department of Statistics (e1071), TU Wien. *R package*, 1, 5–24.
- Drake, E., & Corley, M. (2015). Articulatory imaging implicates prediction during spoken language comprehension. *Memory & Cognition*, 43, 1136–1147.
- Ferreira, F. (1988). *Planning and timing in sentence production: The syntax-to-phonology conversion* PhD thesis. Amherst, MA: University of Massachusetts.
- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, 30, 210–233.
- Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100, 233.
- Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes*, 22, 1151–1177.
- Ferreira, F., & Karimi, H. (2015). Prosody, performance, and cognitive skill: Evidence from individual differences. In *Explicit and implicit prosody in sentence processing* (pp. 119–132). Cham: Springer.
- Fletcher, J. (1987). Some micro and macro effects of tempo change on timing in French. *Linguistics*, 25, 951–967.
- Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (pp. 523–602). Hoboken, NJ: Wiley-Blackwell Publishing.
- Fuchs, S., Petrone, C., Krivokapić, J., & Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics*, 41, 29–47.
- Gao, M. (2008). *Mandarin tones: An articulatory phonology account* PhD thesis. New Haven, CT, USA: Yale University.
- Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411–458.
- Gick, B., Wilson, I., Koch, K., & Cook, C. (2004). Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica*, 61, 220–233.
- Goldman Eisler, F. (1968). *Psycholinguistics. Experiments in spontaneous speech*. London and New York: Academic Press.
- Gollrad, A. (2013). *Prosodic cue weighting in sentence comprehension: Processing German case ambiguous structures* Ph.D. Dissertation. Potsdam University.
- Grosjean, F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, 11, 58–81.
- Gustafson-Čapková, S., & Megyesi, B. (2002). Silence and discourse context in read speech and dialogues in Swedish. *Speech prosody*.
- Hirschberg, J., & Nakatani, C. H. (1996). A prosodic analysis of discourse segments in direction-giving monologues. In *Proceedings of the 34th annual meeting on Association for Computational Linguistics* (pp. 286–293). Association for Computational Linguistics.
- Honikman, B. (1964). Articulatory settings. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, & J. L. M. Trim (Eds.), *Honour of Daniel Jones* (pp. 73–84). London: Longman.
- Horne, M., Strangert, E., & Heldner, M. (1995). Prosodic boundary strength in Swedish: Final lengthening and silent interval duration Stockholm, Sweden, August 13–19, 1995. *Proceedings of the XIIIth international congress of phonetic sciences* (Vol. 1, pp. 170–173). Stockholm: KTH and Stockholm University.
- Jenner, B. (2001). Genealogies of articulatory settings: Genealogies of an idea. *Historiographia Linguistica*, 28, 121–141.
- Katsika, A. (2012). *Coordination of prosodic gestures at boundaries in Greek* PhD thesis. New Haven, CT, USA: Yale University.
- Katsika, A. (2016). The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics*, 55, 149–181.
- Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., & Goldstein, L. (2014). The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics*, 44, 62–82.
- Kawamoto, A. H., Liu, Q., Mura, K., & Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language*, 58, 347–365.
- Kentner, G. (2007). Length, ordering preference and intonational phrasing: Evidence from pauses. *Eighth annual conference of the international speech communication association*.
- Kim, J. (2019). Individual differences in the production of prosodic boundaries in American English. *International Congress of Phonetic Sciences, 4–10 August 2019 2019 Melbourne Australia*.
- Koenker, R. (2018). quantreg: Quantile Regression. R package version 5.36.
- Krivokapić, J. (2007a). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics*, 35, 162–179.
- Krivokapić, J. (2007b). *The planning, production, and perception of prosodic structure* PhD thesis. Los Angeles, CA: University of Southern California.
- Krivokapić, J. (2012). Prosodic planning in speech production. In S. Fuchs, M. Wehrich, D. Pape, & P. Perrier (Eds.), *Speech planning and dynamics* (pp. 157–190). Peter Lang.
- Krivokapić, J. (2014). Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Communicative rhythms in brain and behaviour. Theme Issue of the Philosophical Transactions of the Royal Society B (Biology)*, 369, 20130397.
- Krivokapić, J. (2020). Prosody in articulatory phonology. In S. Shattuck-Hufnagel & J. Barnes (Eds.), *Prosodic theory and practice*. MIT Press.
- Krivokapić, J., & Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics*, 40, 430–442.
- Lane, H., & Grosjean, F. (1973). Perception of reading rate by listeners and speakers. *Journal of Experimental Psychology*, 97, 141–147.
- Lehiste, I. (1975). The phonetic structure of paragraphs. In *Structure and process in speech perception* (pp. 195–206). Springer, Berlin, Heidelberg.
- Laver, J. (1978). The concept of articulatory settings: An historical survey. *Historiographia Linguistica*, 5, 1–14.
- Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. Cambridge, MA: MIT Press.
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia medica: Biochemia medica*, 22, 276–282.
- Mooshammer, C., Goldstein, L., Nam, H., McClure, S., Saltzman, E., & Tiede, M. (2012). Bridging planning and execution: Temporal planning of syllables. *Journal of Phonetics*, 40(3), 374–389.
- Mücke, D., Nam, H., Hermes, A., & Goldstein, L. (2012). Coupling of tone and constriction gestures in pitch accents. In P. Hoole, M. Pouplier, L. Bombien, C. Mooshammer, & B. Kühnert (Eds.), *Consonant clusters and structural complexity* (pp. 205–230). Berlin/New York: Mouton de Gruyter.
- Öhman, S. (1967). Peripheral motor commands in labial articulation, Speech Transmission Laboratory Quarterly Progress Status Report No. 4/, 30–63, Royal Institute of Technology (KTH), Stockholm.
- Parrell, B., Lee, S., & Byrd, D. (2013). Evaluation of prosodic juncture strength using functional data analysis. *Journal of Phonetics*, 41, 442–452.
- Perkell, J. S. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. MIT Research Monograph. No 53.
- Petrone, C., Fuchs, S., & Krivokapić, J. (2011). Consequences of working memory differences and phrasal length on pause duration and fundamental frequency. In *Proceedings of the 9th International Seminar on Speech Production (ISSP)*, (pp. 393–400). Montréal, Canada.
- Petrone, C., Truckenbrodt, H., Wellmann, C., Holzgreffe-Lang, J., Wartenburger, I., & Höhle, B. (2017). Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics*, 61, 71–92.
- R C Team. (2013). R: A language and environment for statistical computing.
- Ramanarayanan, V., Bresch, E., Byrd, D., Goldstein, L., & Narayanan, S. (2009). Analysis of pausing behavior in spontaneous speech using real-time magnetic resonance imaging of articulation. *Journal of the Acoustical Society of America Express Letters*, 126, EL160–EL165.
- Ramanarayanan, V., Goldstein, L., Byrd, D., & Narayanan, S. (2013). An investigation of articulatory setting using real-time magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 134(1), 510–519.
- Rasskazova, O., Mooshammer, C., & Fuchs, S. (2018). Articulatory settings during inter-speech pauses. In M. Belz, C. Mooshammer, S. Fuchs, S. Jannedy, O. Rasskazova, & M. Žygis (Eds.), *Proceedings of the Conference on Phonetics & Phonology in German-speaking countries (P&P13)* (pp. 161–164). Berlin: HU & ZAS.
- Riley, M. A., & Turvey, M. T. (2002). Variability and determinism in motor behavior. *Journal of Motor Behavior*, 34(2), 99–125.
- Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1, 333–382.
- Saltzman, E., Nam, H., Krivokapić, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of the 4th international conference on speech prosody* (pp. 175–184). Brazil: Campinas.
- Sanderman, A. A., & Collier, R. (1995). Prosodic phrasing at the sentence level. In F. Bell-Berti & L. J. Raphael (Eds.), *Producing speech: Contemporary issues. For Katherine Safford Harris* (pp. 321–332). New York: American Institute of Physics.
- Schaeffer, S., Scobbie, J. M., & Mennen, I. (2008). An evaluation of inter-speech postures for the study of language-specific articulatory settings. *Proceeding of the 8th ISSP*. Strasbourg.

- Shaw, J. A., & Kawahara, S. (2018). Assessing feature specification in surface phonological representations through simulation and classification of phonetic data. *Phonology*, 35, 481–522.
- Smith, C. (2004). Topic transitions and durational prosody in reading aloud: Production and modeling. *Speech Communication*, 42, 247–270.
- Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152). New York: Academic Press.
- Strangert, E. (1991). Pausing in texts read aloud. In Proceedings of the XIIIth international congress of phonetic sciences, Aix-en-Provence, France, August 19–24, 1991, Vol. 4 (pp. 238–241). Université de Provence, Service des Publications.
- Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, 37, 21–43.
- Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, 136, 64–81.
- Swets, B., Jacovina, M. E., & Gerrig, R. J. (2014). Individual differences in the scope of speech planning: Evidence from eye movements. *Language and Cognition*, 6, 12–44.
- Tilsen, S., Spincemille, P., Xu, B., Doerschuk, P., Luh, W. M., Feldman, E., & Wang, Y. (2016). Anticipatory posturing of the vocal tract reveals dissociation of speech movement plans from linguistic units. *PLoS One*, 11, e0146813.
- Tyler, J. (2013). Prosodic correlates of discourse boundaries and hierarchy in discourse production. *Lingua*, 133, 101–126.
- Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19 (6), 713–755.
- Whalen, D. H., Chen, W. R., Tiede, M. K., & Nam, H. (2018). Variability of articulator positions and formants across nine English vowels. *Journal of phonetics*, 68, 1–14.
- Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37, 356–381.
- Wilson, I., & Gick, B. (2014). Bilinguals use language-specific articulatory settings. *Journal of Speech, Language, and Hearing Research*, 57, 361–373.
- Yang, X., Xu, M., & Yang, Y. (2014). Predictors of pause duration in read-aloud discourse. *IEICE Transactions on Information and Systems*, 97, 1461–1467.
- Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (pp. 41–62). Chichester: John Wiley.
- Zvonik, E., & Cummins, F. (2003). The effect of surrounding phrase lengths on pause duration. In *Proceedings of Eurospeech 2003* (pp. 777–780).