

The role of speech planning in the articulation of pauses^{a)}

Jelena Krivokapić,^{1,b)} Will Styler,² and Dani Byrd³

¹Department of Linguistics, University of Michigan, Lorch Hall, 611 Tappan Ave, Ann Arbor, Michigan 48109, USA

²Department of Linguistics, University of California San Diego, 9500 Gilman Drive #0108, La Jolla, California 92093, USA

³Department of Linguistics, University of Southern California, 3601 Watt Way GFS 301, Los Angeles, California 90089, USA

ABSTRACT:

Extensive research has found that the duration of a pause is influenced by the length of an upcoming utterance, suggesting that speakers plan the upcoming utterance during this time. Research has more recently begun to examine articulation during pauses. A specific configuration of the vocal tract during acoustic pauses, termed pause posture (PP), has been identified in Greek and American English. However, the cognitive function giving rise to PPs is not well understood. The present study examines whether PPs are related to speech planning processes, such that they contribute additional planning time for an upcoming utterance. In an articulatory magnetometer study, the hypothesis is tested that an increase in upcoming utterance length leads to more frequent PP occurrence and that PPs are longer in pauses that precede longer phrases. The results indicate that PPs are associated with planning time for longer utterances but that they are associated with a relatively fixed scope of planning for upcoming speech. To further examine the relationship between articulation and speech planning, an additional hypothesis examines whether the first part of the pause predominantly serves to mark prosodic boundaries while the second part serves speech planning purposes. This hypothesis is not supported by the results. © 2022 Acoustical Society of America.

<https://doi.org/10.1121/10.0009279>

(Received 14 December 2021; accepted 20 December 2021; published online 24 January 2022)

[Editor: Anders Löfqvist]

Pages: 402–413

I. INTRODUCTION

The goal of this study is to examine the relationship between speech planning and articulation during pauses at prosodic boundaries. A long line of research, starting in the 1960s (see overviews in Goldman-Eisler, 1968; Butterworth, 1980), has established a relationship between speech planning and pauses. For example, Goldman-Eisler (1968), in a series of spontaneous speech studies, finds that pausing behavior varies depending on the task, such that descriptions of a cartoon contain shorter pauses than interpretations of it, and that repetition (of the cartoon descriptions and of the cartoon interpretation) leads to shorter pauses than the first description or interpretation, showing that cognitively complex tasks require more planning. Many subsequent studies have elaborated on our understanding of pauses as indicators of speech planning. Speakers are known to pause longer before structurally complex utterances compared to structurally simpler utterances (Grosjean *et al.*, 1979; Cooper and Paccia-Cooper, 1980; Ferreira, 1991; Strangert, 1991; 1997; Krivokapić, 2007; 2012). Pauses are also longer preceding longer compared to shorter utterances (Sternberg *et al.*, 1978; Ferreira, 1991; Zvonik and Cummins, 2002; 2003; Watson and Gibson, 2004; Krivokapić, 2007; 2012; Fuchs *et al.*, 2013; Krivokapić

et al., 2020). The implication of these findings is that speakers use the pause interval to plan the upcoming utterance, and that the longer and/or more complex an utterance is, the more time speakers need to plan it (Cooper and Paccia-Cooper, 1980; Ferreira, 1991; Strangert, 1997; Watson and Gibson, 2004; Krivokapić, 2007; Fuchs *et al.*, 2013).

The term speech planning as used here encompasses a wide range of processes taking place during pauses. Sternberg *et al.* (1978) examine the execution of the motor programs in a delayed speech task and suggest that the motor program for an upcoming string of words is stored in a buffer before being executed, and that the time it takes to retrieve the program (manifest as a pause before speech onset) increases with each stressed syllable stored in the buffer. In a similarly practiced task, using memorized sentences, Ferreira (1991) extends this line of investigation to examine the effect of the planning of syntactic structure and finds evidence that an increase in syntactic complexity leads to an increase in pause time; she suggests that it is the process of phonological encoding of the complex utterances that leads to the increase in pause time. Many other studies use read speech tasks (e.g., Cooper and Paccia-Cooper, 1980; Watson and Gibson, 2004; Fuchs *et al.*, 2013) to reach similar findings. Regardless of the task and specific type of planning examined, pauses have been recognized as reflecting cognitive activity related to speech planning.

It is generally assumed that speakers do not plan a complete (multi-word) utterance before they start speaking. Rather they plan in increments, and they continue planning

^{a)}An earlier and partial version of this study was presented at the 12th International Seminar on Speech Production, December, 2020.

^{b)}Also at Haskins Laboratories, New Haven, CT 06511, USA. Electronic mail: jelenak@umich.edu

as they speak, i.e., they speak and plan simultaneously (Kempen and Hoenkamp, 1987; Levelt, 1989). The size of the planning increments varies depending on a number of factors such as speakers' individual strategies and cognitive resources, the complexity of the utterance, and the task (Ferreira and Swets, 2002; Krivokapić, 2007; Swets *et al.*, 2007; Wagner *et al.*, 2010; Konopka, 2012; Swets *et al.*, 2013; 2014; Bishop and Intlekofer, 2020; Swets *et al.*, 2021). Importantly, however, existing evidence shows that at least some amount of planning takes place before the start of an utterance, during the pause interval preceding it.

More recently, another line of research has started examining articulation during acoustic pauses. A series of studies of speakers' movement of their articulators during pauses has shown that this articulatory action can start quite early, up to 3s before acoustic speech onset in Krause and Kawamoto (2021) (see also Drake and Corley, 2015; Tilsen *et al.*, 2016; Krause and Kawamoto, 2019; 2020; Tilsen, 2020). This is suggestive of rich effects of cognitive processes and planning. Articulation can also differ depending on the cognitive processes taking place during articulation. Ramanarayanan *et al.* (2009) find differences in articulation during pauses in spontaneous speech for grammatical pauses as compared to non-grammatical pauses. They argue that some of the observed differences—namely differences in variability in speed of movement—can be understood to arise due to the fact that grammatical pauses (but not non-grammatical ones) are cognitive units, i.e., they are linguistically controlled. Ramanarayanan *et al.* (2013) examine pauses during read and spontaneous speech, pauses occurring at the beginning and end of a data acquisition interval (“absolute rest positions”), and pauses occurring directly

prior to speech onset. They identify a number of differences and suggest that various pause types are structurally controlled to a different degree, with pauses occurring during speech in read speech showing most evidence of control, while the absolute rest positions are least linguistically controlled. This is related to the idea that planning pauses and structure pauses might be entirely distinct from each other, as has been discussed in Goldman-Eisler (1968) and Ferreira (2007). Relatedly, Butterworth (1980) finds some evidence that pauses at the end of clauses (which he refers to as “juncture pauses”) in fluent parts of speech are less sensitive to planning needs than pauses in periods of less fluent speech.

The spatiotemporal trajectories of articulators during pauses have also been examined, and specific trajectories that occur at grammatical pauses (specifically, at prosodic boundaries) have been identified (Katsika *et al.*, 2014; Krivokapić *et al.*, 2020; see also Rasskazova *et al.*, 2018 for related findings). Following Katsika *et al.* (2014), we will refer to these as *pause postures*. A pause posture (PP) can be described as a movement that occurs between the final pre-boundary phonological gesture and the anticipatory movement for the upcoming post-boundary gesture, and crucially, it is a movement *that departs from* a straight interpolation between the pre-boundary gesture and post-boundary anticipatory position (see Fig. 1). Katsika *et al.* (2014) identify stable relative timing patterns between the PP and the vowel gesture and boundary tone. They further show spatial stability for the pause postures. The presence of PPs, their systematic timing patterns, and their spatial stability have also been found for American English (Krivokapić *et al.*, 2020). Together, the findings for pause postures indicate

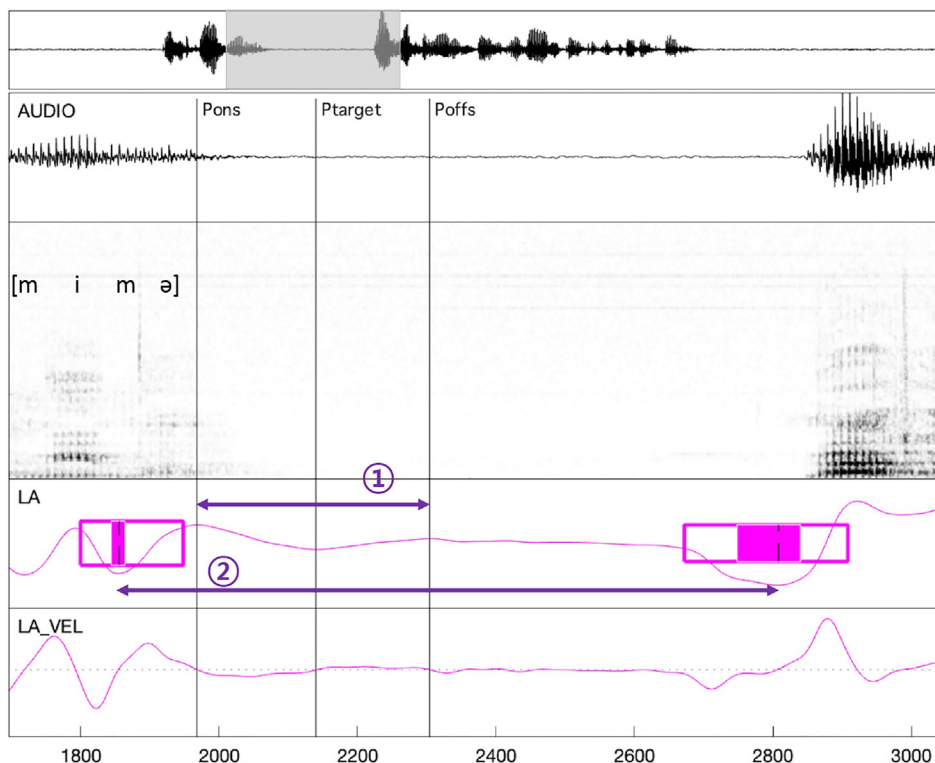


FIG. 1. (Color online) Labeling for the sentence, “They surprised *Mima*. Matt helped enormously with every aspect of the two-day party.”, showing, for the lip aperture, the labeling for the phrase-final bilabial consonant, the pause posture (PP), and the phrase-initial bilabial consonant. Boxes indicate consonant 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 pause posture onset, target (maximum constriction) and offset. 1 = PP duration; 2 = boundary duration; LA: lip aperture trajectory and velocity.

that these are likely to be linguistically controlled movements of the vocal tract.

While previous research has shown that different types of pauses can have different articulatory properties, it remains less clear what cognitive functions underlie these articulations. Krivokapić *et al.* (2020) examine this question for pause postures and suggest that they are related to planning—specifically, that PPs serve to provide additional planning time for an upcoming utterance, i.e., speakers plan the upcoming utterance throughout a pause but deploy a pause posture if they need additional time. However, the study by Krivokapić *et al.* (2020) was not designed to examine planning. The present study builds on this work and collects new data designed specifically to test the hypotheses that: (1) longer upcoming utterances lead to more frequent pause postures, and (2) longer upcoming utterances lead to longer pause postures.

We further evaluate how planning proceeds throughout the boundary interval, specifically if there are differences in *where* during the boundary that planning takes place. It is known that both the phrase preceding the pause and the phrase following the pause have an effect on pause duration, but it has been suggested that conceptually the effects differ. The preceding phrase contributes to pause length as a way of marking prosodic structure, and possibly it provides time for preceding material to be deactivated (Watson and Gibson, 2004; Ferreira, 2007; Krivokapić, 2007). As discussed earlier, the phrase after the boundary is assumed to contribute to pause length because speakers need time to plan the material that will appear in this upcoming phrase. Thus, pauses have multiple cognitive processes as their source, and it is conceivable that the different processes might be active at different points in time and/or with different time courses. While planning may be taking place throughout the boundary interval, we are interested in this study if there is a predominance in the first (earlier) or in the second (later) part of the boundary-related interval.

We start with the strongest version of this hypothesis, that no planning takes place in the first part of the boundary interval (i.e., that it serves for marking prosodic structure only), and instead planning takes place in the second part of the boundary related interval. Some evidence for this hypothesis comes from Ferreira (1991) who finds that an increase in structural complexity of an upcoming phrase has an effect on pause duration but does not have an effect on the duration of the last word of the pre-pausal phrase. (NB: She interprets this finding in combination with other results as evidence that structural pauses are different from planning pauses). We test this hypothesis by examining the release of the phrase-final constriction gesture (the gesture immediately leading into the pause) as an indicator of the planning activity in the first part of the boundary interval and the constriction forming duration of the first phrase-initial consonant (the gesture immediately ending the pause) as an index of the planning activity in the second portion of the boundary interval. If planning indeed takes place in the latter part of this interval and not in the first, we expect to

see upcoming utterance length to have no effect on the duration of the release gesture of the phrase-final consonant but to result in a longer phrase-initial constriction forming gesture for longer upcoming utterances.

A question that arises in relation to pauses regards their cognitive status (see also Krivokapić *et al.*, 2020; Byrd and Krivokapić, 2021). Specifically, if PPs indeed add additional planning time, do speakers explicitly insert a new gesture to allow for this planning time or does the pause posture emerge as a result of articulators returning to their default position once articulation is complete? Our study does not explicitly test this question, but we evaluate the issue in the discussion.

II. METHOD

To examine these three hypotheses, an electromagnetic articulometry experiment was conducted.

A. Stimuli and participants

Eight speakers (four male and four female) participated in the experiment. They were all native speakers of American English with no known speech or hearing disorders. The participants were students at the University of Michigan.

Four sets of sentences were designed, each containing three utterances, and containing the names *Mima* or *miMA*, for a total of 24 sentences (four sets \times three utterances \times two names). The target pause in each utterance was between two phrases. In each set, the pre-boundary phrase was identical (and was five or six syllables) while the post-boundary phrase varied in length (short: four syllables; medium: 10 syllables, and long: 17 syllables). This allowed testing for an effect of the length of an upcoming phrase on PP occurrence and duration. The pre-boundary phrase ended in one of the three words (*miMA*, *Mima*, or *Ema*; capitalization showing stress), and the post-boundary phrase always started with a bilabial (the immediate post-boundary word was *Bob*, *Mike*, *Mima*, *miMA*, or *Matt*). The specific target words and their position as phrase-final or phrase-initial were introduced for the purposes of another study, but crucially, the consonants preceding and following the pause were always bilabial, allowing us to track lip aperture in a controlled manner. The stimuli are shown in Table I.

The stimuli were pseudorandomized in blocks of 24 sentences. Eight repetitions were recorded for each of the speakers, for a total of 192 utterances per speaker. Participants were asked to read the sentences as if reading a story to someone. The instructions also contained information on how to produce the new names introduced in the stimuli (*Mima* and *miMA*, which were pronounced as [ˈmimə] and [mɪˈmɑ]). For purposes of another experiment, some of the included sentences had further contextual instructions as to how to speak the sentences (e.g., as if the speaker was uncertain, or as if they were making a suggestion). The participants familiarized themselves with the stimuli by reading the sentences aloud one or two times

TABLE I. Stimuli sets, shown with the target word “MIma”. The same utterances were also recorded with “miMA”. The instructions on how to produce the sentences and the context sentences are given in italics.

Condition	Stimuli	Number of syllables before/after the boundary
Set 1		
<i>(uncertain)</i>		
short	1. I think it was MIma. # Bob told me so.	6/4
medium	<i>(uncertain)</i>	6/10
2. I think it was MIma. # Bob told me about her marriage last week.		
long	<i>(uncertain)</i>	6/17
3. I think it was MIma. # Bob just talked to me about her upcoming marriage and honeymoon.		
Set 2		
<i>Context: There is no one who knows this!</i>		
short	<i>I can't ask anyone!</i>	5/4
<i>(contradicting/suggesting)</i>		
4. You could ask MIma. # Mike always does.		
<i>Context: There is no one who knows this!</i>		
medium	<i>I can't ask anyone!</i>	5/10
<i>(contradicting/suggesting)</i>		
5. You could ask MIma. # Mike always asks her for help with physics.		
<i>Context: There is no one who knows this!</i>		
long	<i>I can't ask anyone!</i>	5/17
<i>(contradicting/suggesting)</i>		
6. You could ask MIma. # Mike regularly asks her for help with his mom's obnoxious parrot.		
Set 3		
short	7. They surprised MIma. # Matt helped a lot.	5/4
8. They surprised MIma. # Matt helped them a lot with the arrangements.		
medium		5/10
long	9. They surprised MIma. # Matt helped enormously with every aspect of the two-day party.	5/17
Set 4		
short	10. Do you know Emma? # MIma knows her.	5/4
medium	11. Do you know Emma? # MIma ran into her at the market.	5/10
long	12. Do you know Emma? # MIma ran into her while shopping with her husband the other day.	5/17

before the recording. During the recordings, the experimenter monitored the productions and asked the participant to read a sentence again in case of suspected error.

B. Data collection

Articulatory kinematic data were recorded using electromagnetic articulometry (EMA) (Carstens Articulograph AG 501), at a sampling rate of 400 Hz. Sensors were placed midsagittally on the tongue tip, body, and dorsum, on the upper and lower lips, on the jaw, and on right and left eyebrows. Three reference sensors (right and left mastoid and upper incisors) were used to correct for head movement. Acoustic data were acquired simultaneously using a Sennheiser shotgun microphone at a sampling rate of 16

kHz. In post-processing, articulatory data were corrected for head movement, rotated to the occlusal plane, and smoothed with a 3rd-order Butterworth low pass filter with a cut-off frequency of 20 Hz. Velocity signals were calculated as the central difference of the filtered position data, approximating the first derivative.

C. Data analysis

A research assistant naive to the purposes of the experiment listened to all the productions to ascertain the target words were produced with the correct stress and that the sentences did not have disfluencies. Tokens where this was not the case were excluded. All utterances that were produced correctly were kept in the analysis, even if it exceeded the targeted eight repetitions (Participants were occasionally asked to repeat an utterance, in case of suspected error. If the original repetition was later found to be produced without an error, both the first and the second repetition were kept in the analysis). Utterances that had more than one pause posture during the pauses (in each case, it was two pause postures) were also excluded (see below in this section on how pause postures were labeled). Due to experimental error, F1 had 186 instead of 192 utterances collected. The total number of utterances included in the analysis is 1446. Table II shows the data collected and excluded for each participant.

The lip aperture gestures for the bilabial consonants surrounding the pause were semi-automatically labeled using mview (custom software written by Mark Tiede at Haskins Laboratories, New Haven, CT) using velocity criteria. Specifically, for the consonant gestures, we label gesture onset (20% of onset peak velocity), peak velocity of the constriction forming movement, maximum constriction (velocity minimum), peak velocity of the constriction release movement, and gesture offset (20% of offset peak velocity). Utterances in which the constriction gestures could not be identified reliably were excluded from further analysis (see Table I). Pause postures (PP) were identified on lip aperture (LA) as well, using mview.

Pause postures are considered to be movements during the acoustic pause that deviate from a linear interpolation between the pre-boundary and post-boundary consonant constrictions (Fig. 1) (Katsika *et al.*, 2014; Krivokapić *et al.*, 2020). For the PPs, the following points along the LA trajectory were identified: the onset of the PP was defined as the velocity zero-crossing preceding a change in direction of movement towards the pause posture, the PP offset was defined as the velocity zero-crossing before a change of movement direction or plateau, and the target of the PP was defined as the maximum constriction of the lips (i.e., minimum LA). Many PPs were straightforwardly identifiable, as the movement clearly deviated from the straight interpolation line; that said, we took as a threshold for PP identification if the LA for target of the PP was at least 1 mm smaller (lips more closed) than the PP onset and PP offset. The measure of 1 mm follows from Krivokapić *et al.*, 2020, where a human annotator estimated, based on visual inspection of

TABLE II. Data excluded due to disfluencies, prosodic errors, and unlabelability.

Participant	Total number of utterances collected	Disfluent/prosodic errors	Could not be labeled	Pauses with two PPs	Total number of utterances in the analysis
F1	186	0	0	0	186
F2	208	33	4	0	171
F3	202	20	1	0	181
F4	217	26	0	0	191
M1	200	31	2	0	167
M2	212	23	3	0	186
M3	220	38	3	3	176
M4	210	20	0	2	188

the data, that 1 mm deviation from a straight line could be a meaningful movement in the sense that it seemed to exclude random jitter. This measure was further evaluated in a machine learning model (Krivokapić *et al.*, 2020), which indicated that although pause postures show complexity of curvature, a machine learning model trained on curvature alone and a human annotator using this 1 mm criterion reached a Cohen's Kappa agreement of 0.89, affirming this as a useful heuristic. From the labeled landmarks, we calculate PP duration (onset to offset of PP) and boundary duration (from maximum constriction of the LA of the pre-boundary consonant to maximum constriction of the LA of the post-boundary consonant) (Fig. 1). To test the third hypothesis regarding early and late phases of the interphrase interval, we calculated (1) the duration of the release of the phrase-final constriction gesture (as an indicator of the first part of the boundary), from gesture target to gesture offset of the phrase-final consonant, and (2) the constriction formation duration of the phrase-initial consonant (as the indicator of the second part of the boundary) from gesture onset to gesture target.

D. Statistical analysis and results

Pause postures occurred in 393 out of 1446 utterances (27.18%). Table III shows the number and percentages of pause postures for each speaker. As has been found in Krivokapić *et al.* (2020), speakers vary substantially in the number of PPs.

In considering the effects of utterance length on pause postures, we examined the three conditions in our stimuli—short: four syllables occurring after the pause; medium: 10

syllables occurring after the pause, and long: 17 syllables occurring after the pause. We test the hypotheses (1) that upcoming utterance length has an effect on PP occurrence, such that longer upcoming utterances lead to more frequent PPs, and (2) that upcoming utterance length has an effect on PP duration, such that longer utterances lead to longer PPs. Significance for all tests was assessed at $p < 0.05$.

All data analysis was conducted using the R Statistics Software (R Core Team, 2021, Vienna, Austria, <https://www.R-project.org>). To test the first hypothesis, a generalized linear model (GLM) was fitted using “glm” in R, testing the effect of upcoming utterance length (short, medium, long). Results show that there is a significant effect of upcoming phrase length on PP occurrence, such that longer upcoming phrases lead to more PPs for all speakers pooled, $p < 0.0001$ (Fig. 2) and for all speakers individually except speaker F3 (Fig. 3). For the speakers pooled, *post hoc* Tukey analyses show that for long and medium upcoming phrases, PPs are more likely than they are for short upcoming phrases (in both comparisons $p < 0.0001$). For individual speakers, for F1, the effect is such that for long phrases, there is a higher likelihood of PPs than for short and medium phrases; for F2 and M2, there is a higher likelihood for PPs for medium than for short phrases; for F4, M1 and M4, there is a higher likelihood of PPs for long and for medium phrases than for short phrases; and for M3, there is a higher

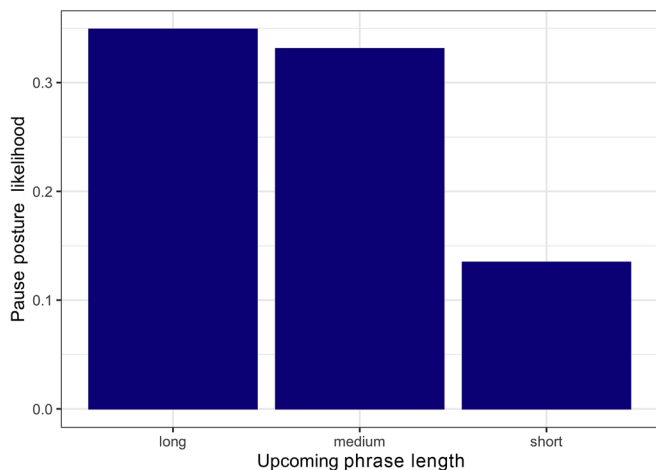


FIG. 2. (Color online) The effect of upcoming phrase length on pause posture occurrence, all speakers pooled.

TABLE III. Total number of pause postures and percentages.

Participant	Total number of pause postures and percentage
F1	4 (2.15)
F2	47 (27.48)
F3	13 (7.18)
F4	50 (26.32)
M1	35 (20.96)
M2	81 (43.55)
M3	78 (44.32)
M4	85 (45.21)

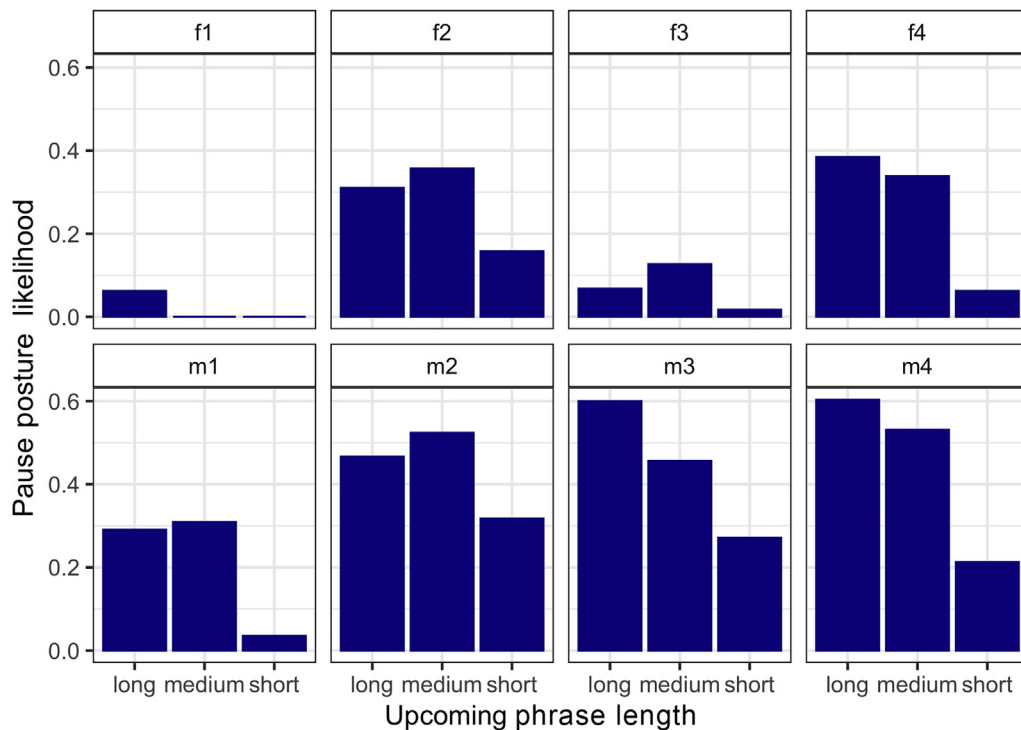


FIG. 3. (Color online) The effect of upcoming phrase length on pause posture occurrence, individual speakers.

likelihood for PP occurrence for long than for short upcoming phrases (see Table IV).

Given that PPs are part of the boundary interval, it needs to be ascertained that the observed effect of upcoming phrase length on PP occurrence is independent of boundary duration, as upcoming phrase length is known to have an effect on pause and boundary duration (e.g., Ferreira, 1991; Krivokapić, 2007; Krivokapić *et al.*, 2020; Sternberg *et al.*, 1978; Zvonik and Cummins, 2003; Watson and Gibson, 2004). To examine this, we first fitted a GLM, testing the effect of boundary duration on PP occurrence. Boundary duration was z-scored by speaker for pooled analysis. The effect is significant ($p < 0.0001$), with longer boundaries leading to more PPs. This is the case both for speakers pooled (Fig. 4) and for individual speakers (Fig. 5). To ensure that the observed effect of upcoming phrase length on PP occurrence is independent of boundary duration, model comparisons (using “anova” in R) compared models that included z-scored boundary duration as well as upcoming utterance length as predictors of PP occurrence to models that included only z-scored boundary duration. We see a significant model improvement (in terms of minimizing error) when adding upcoming utterance length to the model.

This result indicates that the effect of upcoming utterance length on PP occurrence is different from and independent of the effect of boundary duration. A model comparison between this two-parameter model (both boundary duration and upcoming phrase length) and a nested one-parameter model (only boundary duration) found that the two-parameter model has a better fit compared to the one-parameter model ($p < 0.0001$). Thus, both boundary duration and upcoming utterance length are significant factors in PP occurrence, and upcoming phrase length has an independent significance effect on PP occurrence.

To test the second hypothesis that longer upcoming utterances lead to longer pause postures, a correlation analysis examined the relationship between upcoming utterance length and PP duration (using the *lm* function in R). No significant overall correlation was found (with neither raw nor z-scored-by-speaker pause posture duration). Separate analyses were run for speakers with more than 50 PP tokens (as most speakers did not have enough PPs for such a model to converge; three speakers had more than 50PPs); again, no significant correlations were found. Thus, hypothesis two associating upcoming utterance length with longer PP duration failed to be confirmed.

TABLE IV. Results of *post hoc* analyses of PP occurrence by speaker (only significant effects shown).

F1	F2	F3	F4	M1	M2	M3	M4
long > short ($p = 0.037$)	medium > short ($p = 0.043$)	No significant effect	long > short ($p < 0.001$)	long > short ($p = 0.002$)	medium > short ($p < 0.048$)	long > short ($p < 0.001$)	long > short ($p < 0.001$)
long > medium ($p = 0.04$)			medium > short ($p < 0.001$)	medium > short ($p < 0.001$)			medium > short ($p < 0.001$)

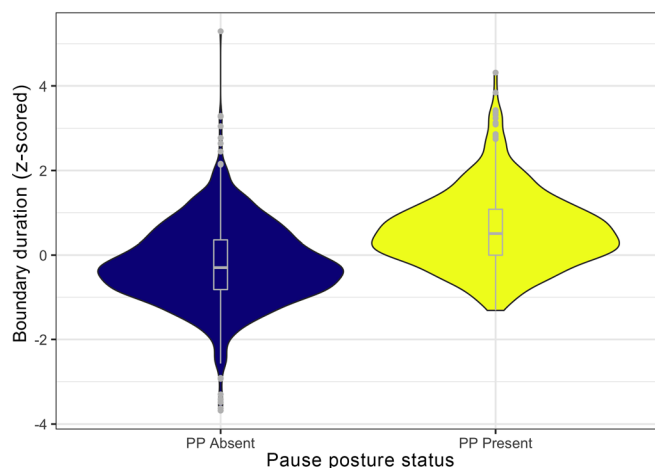


FIG. 4. (Color online) The effect of z-scored boundary duration on the occurrence of pause posture, all speakers pooled.

To get a further qualitative sense of the relationship between PP duration and boundaries, and PP duration and upcoming phrase length, Fig. 6 shows the distribution of pause posture length, and Fig. 7 shows the percentage of the duration of the boundary that is occupied by the PP. What we see from these figures is that PPs are not all of equal duration, nor do they occupy the same proportion of the boundary interval in each utterance. We will return to this in the discussion. To investigate the relationship of upcoming utterance length and planning in more depth, we conducted further correlation tests examining the effect of upcoming utterance length on boundary duration for boundaries without PPs and for boundaries with PPs, and on boundary duration for boundary intervals overall (i.e., boundaries with and without PPs combined). Boundary durations were z-scored by speaker, based on all data (PP and non-PP combined).

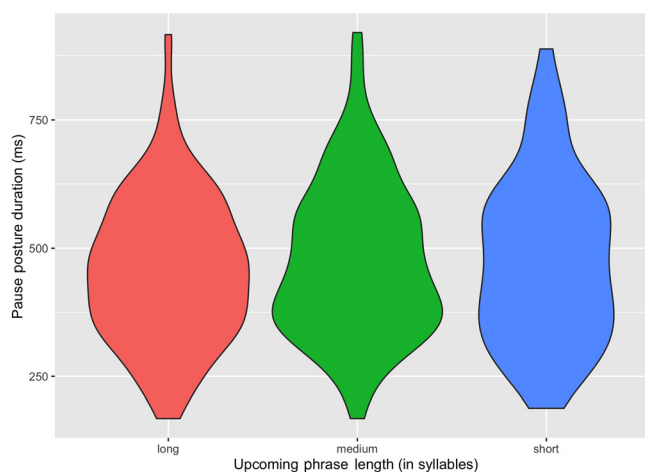


FIG. 6. (Color online) The effect of upcoming utterance length on pause posture duration.

Based on previous studies that found an effect of upcoming utterance length on pause duration, we expected to find a significant correlation between overall boundary duration and upcoming utterance length, and, based on Krivokapić *et al.* (2020), we expected to see an effect of upcoming utterance length and boundaries with PPs and boundaries without PPs. There was a positive correlation between upcoming utterance length and boundary duration for boundaries overall (Fig. 8), such that longer upcoming phrases are associated with longer boundaries than short upcoming phrases (with each additional syllable of length predicted to add 4 ms to the boundary's length in a non-z-scored version of the model, both $p < 0.005$). No other correlations were significant. To probe the absence of effect (given that the effect of upcoming phrase length on pause duration is very well established), separate by-speaker

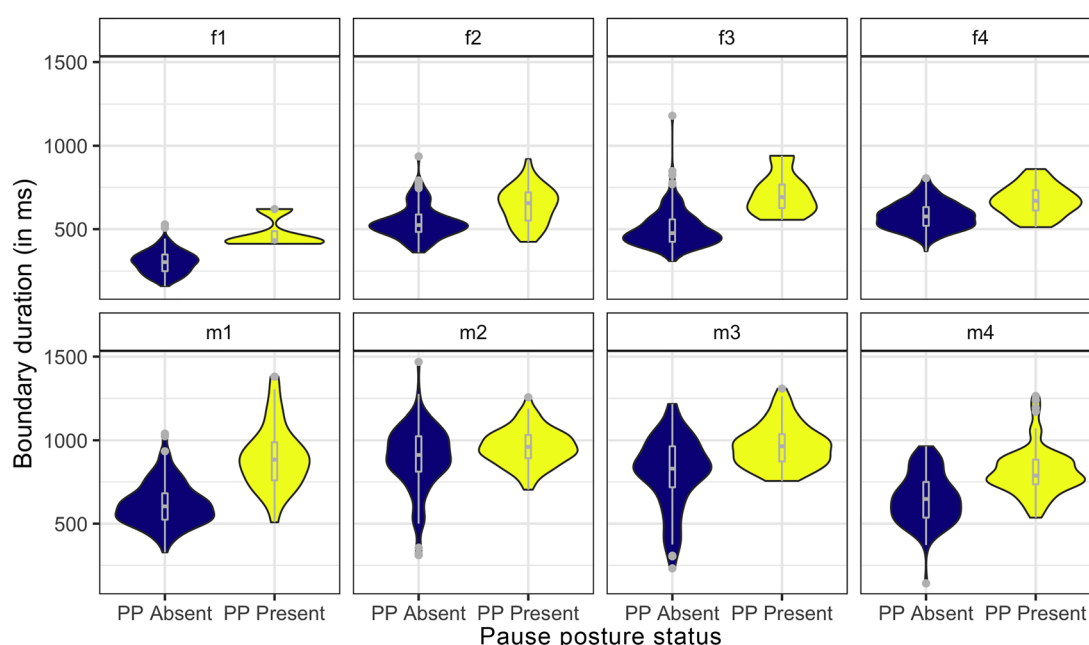


FIG. 5. (Color online) The effect of boundary duration on the occurrence of pause posture, individual speakers.

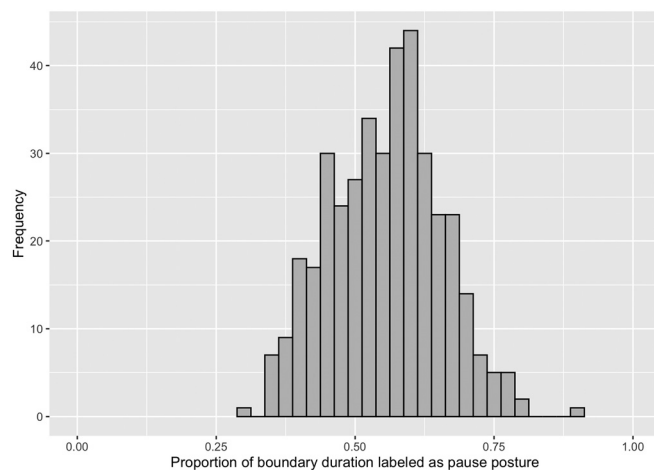


FIG. 7. The percentage of the duration of the boundary that is occupied by the PP.

analyses were run for data with and without pause postures. For boundaries with PPs, the analyses were run for the three speakers with more than 50 PP tokens (as most speakers did not have enough PPs for such a model to converge), and no effect was found even for these speakers. In boundaries without pause postures, F3, F4, and M1 all showed significant effects ($p < 0.05$) consistent with the overall finding, with greater syllable-count length of upcoming phrase predicting longer boundaries (see Fig. 9). That said, the separate pooled analyses for boundaries with PPs and boundaries without PPs did not reach significance, likely due to the smaller amounts of data involved and the lack of uniformity across speakers.

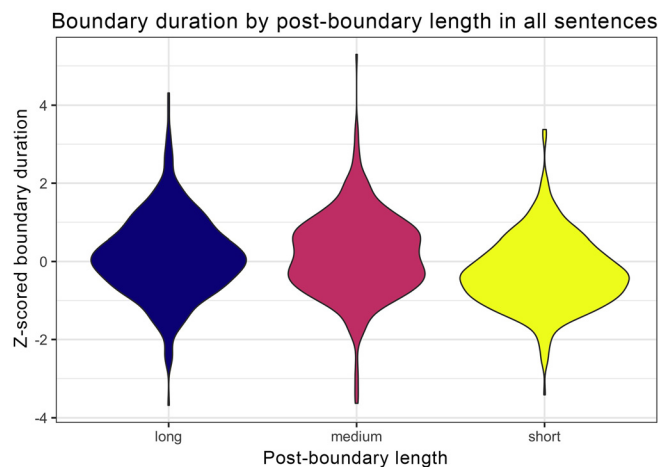


FIG. 8. (Color online) The effect of upcoming phrase length on z-scored boundary duration.

The third hypothesis tested whether planning occurs in the second part of the boundary and not in the first. If so, we expect there to be no effect of upcoming utterance length on the duration of the release gesture of the phrase-final consonant, and we expect that longer upcoming utterances will lead to longer phrase-initial constriction forming gestures. Linear models tested the effect of upcoming utterance length on by-speaker z-scored constriction release and constriction formation for individual speakers. For all speakers pooled, there was a main effect such that the shortest upcoming phrase leads to longer z-scored constriction forming duration ($p = 0.024$). In by-speaker tests using non-z-scored

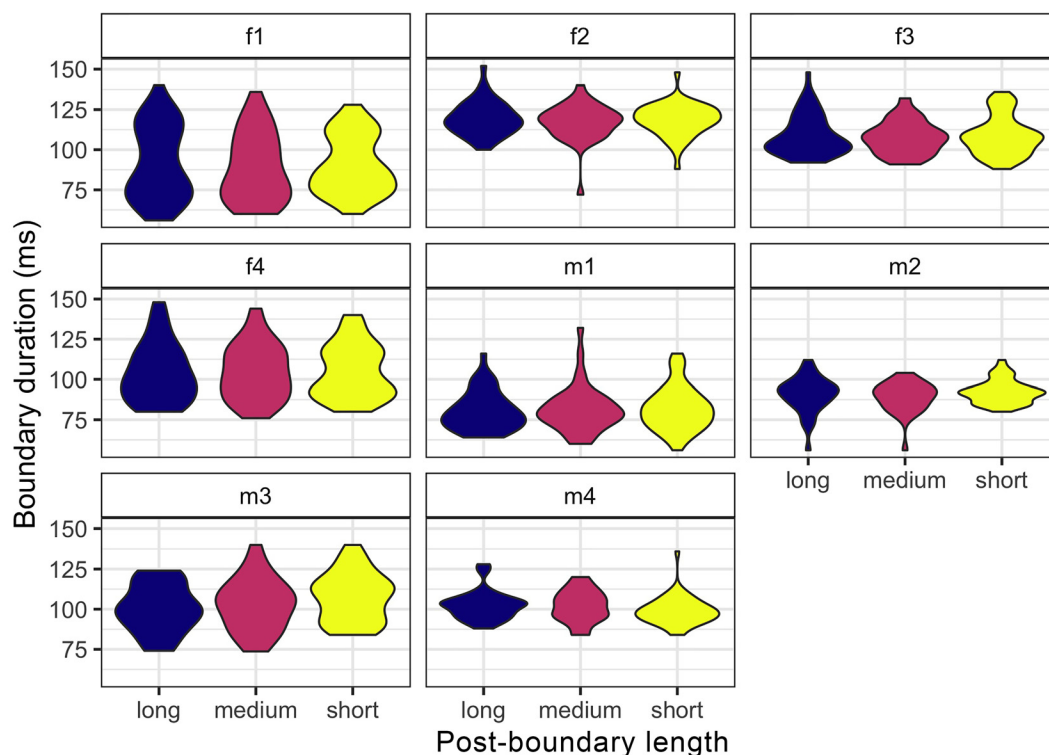


FIG. 9. (Color online) The effect of upcoming phrase length on boundary duration for sentences without PPs.

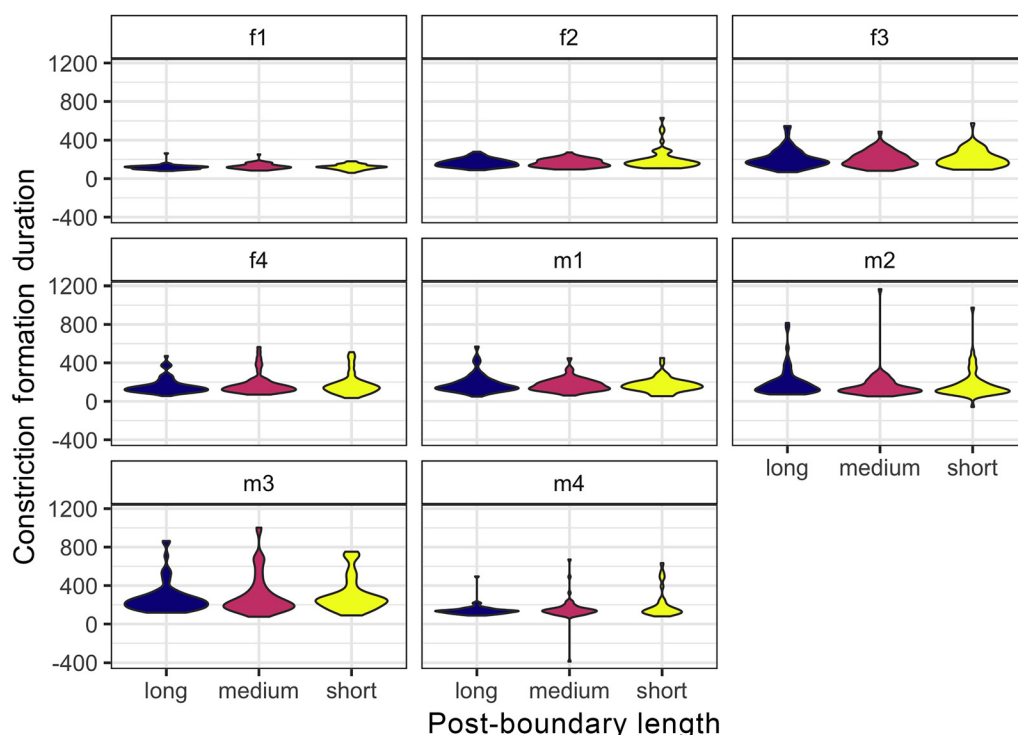


FIG. 10. (Color online) The effect of upcoming phrase length on constriction formation duration.

data (Fig. 10), three speakers showed an effect of constriction formation, with speaker F1 showing that medium upcoming lengths lead to longer duration ($p = 0.0322$), and speakers F2 ($p = 0.007$), and M4 ($p = 0.001$) mirroring the main effect of shortest upcoming lengths leading to longer constriction formation duration. The findings for all speakers pooled and for F2 and M4 was in the direction opposite of what our hypothesis stated, while the result for F1 could be interpreted either way.

A second series of linear models were conducted to examine the effect of upcoming utterance length on constriction release. The prediction is that there will be no effect of upcoming utterance length on constriction release. There was no effect of upcoming utterance length on constriction release duration when all speakers were pooled (release durations were z-scored by speaker for the pooled analysis). When releases were examined individually by speaker, only speaker M3 shows an effect, again with short upcoming phrases predicting longer constriction releases ($p = 0.023$). Thus, hypothesis 3 is not supported.

III. DISCUSSION

The study examined three hypotheses aimed at illuminating the relationship between articulation and speech planning. The first two, the primary focus of the study, build on work by Krivokapić *et al.* (2020) and examined if and how pause postures are related to the planning of an upcoming utterance, specifically whether PPs provide additional planning time for speakers to plan an upcoming utterance. Pause postures occurred in 27.18% of the utterances, a slightly

lower percentage than has been found in Krivokapić *et al.* (2020), where 31% of the utterances had PPs. As in that study, we see large individual differences in the number of PPs, ranging from 2% for speaker F1 to 45% for speaker M4. Given well-known differences in how individual speakers plan upcoming utterances, this variability is not surprising if PPs are related to planning. Hypothesis 1 examined whether longer upcoming utterances lead to more frequent PP occurrence, and hypothesis 2 tested whether PPs are longer when upcoming utterances are longer. Our findings support hypothesis 1, but evidence is not found supporting hypothesis 2. The third hypothesis examined speech planning in early and later portions of the pausal interval, specifically testing the hypothesis that planning predominantly takes place during the latter part of a pause. We do not find evidence supporting this hypothesis.

The study's findings are consistent with the view that planning takes place throughout the prosodic boundary, as argued in Krivokapić *et al.* (2020). Further evidence for this comes from the present study in that boundary duration increases with an increase in upcoming utterance length and does not specifically increase in specific parts of the boundary related interval—e.g., the PP, phrase-final consonant release, phrase-initial constriction formation—as indicated by the results for hypothesis 2 and 3.

What, then, do pause postures specifically contribute to speech planning? We propose that they provide additional planning time for the upcoming utterance, as also argued for in Krivokapić *et al.* (2020). The time they contribute varies (Fig. 6), as does the proportion of the inter-phrase interval

that this time occupies (Fig. 7). These findings, together with the fact that PP duration does not vary with upcoming utterance length (contra hypothesis 2), indicates that there is a relatively fixed scope of planning for the upcoming utterance. We do *not* suggest that there is a specific planning increment or unit for all speakers, only that speakers plan a relatively fixed amount of material before they start speaking, whatever that amount is. Speakers can use a PP to accommodate additional planning time that they might need above and beyond that available from the inter-phrase boundary interval (that would occur without a PP). We postulate that speakers aim to speak fluently and that to do so, they need to plan only a chunk of speech ahead of time and then can continue remaining planning as they are speaking (see, e.g., Ferreira and Swets, 2002; Griffin, 2003). Especially for read speech, that chunk can be quite small, as a lot of information is already provided from the text for the speaker (for example, lexical access demands are minimal). The initial planning chunk can also be assumed to be relatively stable in size since the conditions of the production are stable across the experiment (for example, there is no possibility of an interrupting co-speaker or of not knowing what to say). Thus, speakers can presumably often smoothly plan this chunk during the inter-phrase boundary interval, but if not, they may use a PP to afford themselves additional planning time.

It is worth noting that results of the current study regarding PP duration differ somewhat from those in Krivokapić *et al.*, 2020, and it might be that this is due to our particular setup; speakers were familiar with the relatively “templatic” sentences they were reading and thus could produce them relatively easily, without the need for extensive planning. In contrast, Krivokapić *et al.* (2020) had more widely varying utterances. Further, it is also known that one of the factors influencing planning time is the first word in an utterance. It was important in the present study’s experimental method to keep that word constant across the length conditions; that said, the fact that the first word varied very little probably also led to a more fixed planning time. In sum, the present experiment was optimized to control the length and composition of the sentences, which likely promoted a relatively stable planning increment, but future work could introduce more variation into the upcoming phrases to evaluate the role of predictability in modulating these planning effects.

Finally, we turn to the cognitive status of these PPs (see also Krivokapić *et al.*, 2020; Byrd and Krivokapić, 2021). Two possibilities are put on the table. First, pause postures may be cognitive units akin to gestures that are added to the speech specifically with the purpose of providing additional planning time when a speaker needs it. Or alternatively, pause postures may “merely” arise or emerge as the active articulator returns to its default position, no longer under active control for creating a speech constriction gesture. In the framework of Articulatory Phonology, which adopts a point attractor model for constriction tasks in the vocal tract, a neutral attractor has been proposed specifically for this

purpose, which draws each articulator back to its equilibrium or rest position when it is not under the active control of a constriction gesture. This way, articulators becoming de-active constrictions at any particular point in time are actively reciprocally attracted toward a default position in the vocal tract, preventing them from remaining in the posture of the executed constriction (setting aside the question of active constriction release gestures). It has been suggested that default articulatory “settings” of the vocal tract, including distinct cross-linguistic settings, might arise this way (Saltzman and Munhall, 1989; see also Ramanarayanan *et al.*, 2013). Under this account, speakers producing a long pause to allow extra planning time would be in a situation where a neutral attractor would be at work and, thereby, lead to the consequent emergence of a pause posture.

These two possibilities can potentially be distinguished. A PP required for planning needs (possibility 1) would (all else equal) make demands on *all* vocal tract articulators. Such a pause posture would, by hypothesis, be identifiable on all kinematic trajectories at the same time.¹ Possibility 2—that PPs emerge due to a neutral attractor or default setting—would be supported if evidence of PPs is found at separate times for separate articulators depending on when different constriction gestures are deactivated.

A further consideration related to possibility 2 is proffered in the Katsika *et al.* (2014) account of pause postures; this is based on the Byrd and Saltzman (2003) dynamic account of local slowing at phrase edges—the π -(or prosodic)-gesture model. The π -gesture models temporal properties of prosodic boundaries by instantiating a local slowing of the timeflow of gestural activation functions. These cognitive elements—prosodic gestures—are driven by phrasal structure and do not have a specific articulator associated with them; rather, they act *upon* articulatory gestures that are co-active at the same time, leading, among other things, to the well-known effect of final lengthening and pausing. Particularly relevant for the present discussion is that the strength of a π -gesture varies—and in turn, its effect on concurrent articulatory actions varies—depending on the strength of the prosodic juncture. This means that stronger boundaries will yield greater local slowing and, if sufficiently strong, longer pauses. Katsika *et al.* (2014), based on stable relative timing patterns of PPs to gestures of the preceding utterance, argue that PPs are triggered by particularly, i.e., sufficiently, strong π -gestures. This accounts both for the fact that these postures occur during pauses (since structural pauses only occur at strong boundaries) and for their stable timing pattern. Support for this approach is also seen in Krivokapić *et al.* (2020), where the same stable relative timing patterns are identified. If this is further verified, the pause posture could be a specific gesture (that can occur across vocal tract constriction subsystems) with a specific planning function (à la option 1), *and* it could be implemented or realized as a robust neutral attractor for any de-activating constriction’s articulator(s)—this action would be in the same “vicarious” spirit as the π -gesture in that its sole vocal tract action is *an effect upon* articulatory gestures

(Byrd and Saltzman, 2003). Relatedly, Byrd and Saltzman (2003) discuss the possibility of two π -gestures, one at the end of a prosodic phrase and one at the beginning of the next prosodic phrase. This would exist in the case when there is a substantial amount of time between the two phrases, typically a long pause. A neutral attractor of the sort we are theorizing could be robustly active in just such a case. If this approach is correct, possibilities 1 and 2 are not mutually exclusive and would be difficult to distinguish (though perhaps the context of a full stop in speech with no upcoming material would allow a window into the distinction between planning versus a “simple” return to a default setting).

IV. CONCLUSION

To summarize, our study finds evidence for the existence of articulatory pause postures (PPs) in American English, supporting earlier like findings. The frequency of PPs increases with an increase in upcoming phrase length (indexed by syllable count), but PP duration is not affected by upcoming phrase length. We interpret this to mean that the emergence of PPs at boundaries is associated with a need to increment additional planning time for longer utterances. The lack of effect of upcoming phrase size on PP duration may indicate a relatively fixed scope of planning for upcoming speech in this study, regardless of its actual length, with a relatively stable PP duration sufficing to allow sufficient planning time for the utterances to be produced fluently. Finally, planning for the upcoming utterance appears to proceed in both the earlier and later portions of the boundary-related inter-phrase interval. In sum, this study adds novel articulatory grounding to the body of evidence that pauses are related to cognitive speech planning.

ACKNOWLEDGMENTS

We are grateful to Jiseung Kim, Stephen Tobin, and Mariko Ito for data collection, labeling, and prosodic verification. We also thank the JASA editor for his thoughtful guidance. This work was supported by NSF Grant No. 1551513 to Krivokapić, and by the University of Michigan Phonetics Laboratory.

¹Alternatively we cannot rule out the possibility that only the articulators of the last active gesture of the preceding phrase or the first of the upcoming phrase would participate in the PP; for example, as the need for additional planning time becomes evident, the PP is activated specifically to currently active articulators, such that the Pause Posture is only specified for the articulator of the immediate gesture. We consider such articulator-specific PPs less likely and will not discuss them further but cannot rule-out the possibility.

Bishop, J., and Intlekofer, D. (2020). “Lower working memory capacity is associated with shorter prosodic phrases: Implications for speech production planning,” in *Proceedings of 10th International Conference on Speech Prosody*, pp. 2020–2039.

Butterworth, B. (1980). “Evidence from pauses in speech,” in *Speech and Talk*, edited by B. Butterworth (Academic Press, London, New York, Toronto, Sydney, San Francisco), pp. 155–176.

Byrd, D., and Krivokapić, J. (2021). “Cracking prosody in articulatory phonology,” *Annu. Rev. Linguist.* **7**, 31–53.

Byrd, D., and Saltzman, E. (2003). “The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening,” *J. Phon.* **31**, 149–180.

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

Drake, E., and Corley, M. (2015). “Articulatory imaging implicates prediction during spoken language comprehension,” *Mem. Cogn.* **43**, 1136–1147.

Ferreira, F. (1991). “Effects of length and syntactic complexity on initiation times for prepared utterances,” *J. Mem. Lang.* **30**, 210–233.

Ferreira, F. (2007). “Prosody and performance in language production,” *Lang. Cogn. Proc.* **22**, 1151–1177.

Ferreira, F., and Swets, B. (2002). “How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums,” *J. Mem. Lang.* **46**, 57–84.

Fuchs, S., Petrone, C., Krivokapić, J., and Hoole, P. (2013). “Acoustic and respiratory evidence for utterance planning in German,” *J. Phon.* **41**, 29–47.

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

Griffin, Z. M. (2003). “A reversed word length effect in coordinating the preparation and articulation of words in speaking,” *Psychon. Bull. Rev.* **10**, 603–609.

Grosjean, F., Grosjean, L., and Lane, H. (1979). “The patterns of silence: Performance structures in sentence production,” *Cogn. Psychol.* **11**, 58–81.

Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., and Goldstein, L. (2014). “The coordination of boundary tones and its interaction with prominence,” *J. Phon.* **44**, 62–82.

Kempen, G., and Hoenkamp, E. (1987). “An incremental procedural grammar for sentence formulation,” *Cogn. Sci.* **11**, 201–258.

Konopka, A. E. (2012). “Planning ahead: How recent experience with structures and words changes the scope of linguistic planning,” *J. Mem. Lang.* **66**, 143–162.

Krause, P. A., and Kawamoto, A. H. (2019). “Anticipatory mechanisms influence articulation in the form preparation task,” *J. Exp. Psychol. Hum. Percept. Perform.* **45**, 319–335.

Krause, P. A., and Kawamoto, A. H. (2020). “Nuclear vowel priming and anticipatory oral postures: Evidence for parallel phonological planning?,” *Lang. Cogn. Neurosci.* **35**, 106–123.

Krause, P. A., and Kawamoto, A. H. (2021). “Predicting One’s Turn With Both Body and Mind: Anticipatory Speech Postures During Dyadic Conversation,” *Front. Psychol.* 684248.

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

Krivokapić, J. (2012). “Prosodic planning in speech production,” *Speech planning dynamics in Speech planning and dynamics* (Peter Lang, Frankfurt, Berlin, Bern, Bruxelles, New York, Oxford, Vienna), pp. 157–190.

Krivokapić, J., Styler, W., and Parrell, B. (2020). “Pause postures: The relationship between articulation and cognitive processes during pauses,” *J. Phon.* **79**, 100953.

Levelt, W. (1989). *JM 1989. Speaking: From Intention to Articulation*, “(The MIT Press, Cambridge, MA).

R Core Team. (2021). “R: A Language and Environment for Statistical Computing.” available at <https://www.R-project.org>.

Ramanarayanan, V., Bresch, E., Byrd, D., Goldstein, L., and Narayanan, S. S. (2009). “Analysis of pausing behavior in spontaneous speech using real-time magnetic resonance imaging of articulation,” *J. Acoust. Soc. Am.* **126**, EL160–EL165.

Ramanarayanan, V., Goldstein, L., Byrd, D., and Narayanan, S. S. (2013). “An investigation of articulatory setting using real-time magnetic resonance imaging,” *J. Acoust. Soc. Am.* **134**, 510–519.

Rasskazova, O., Mooshammer, C., and Fuchs, S. (2018). “Articulatory settings during inter-speech pauses,” in *Proceedings of the Conference on Phonetics & Phonology in German-Speaking Countries (P&P 13)*, 28–29 September, Berlin, Germany, pp. 161–164.

Saltzman, E. L., and Munhall, K. G. (1989). “A dynamical approach to gestural patterning in speech production,” *Ecol. Psychol.* **1**, 333–382.

Sternberg, S., Monsell, S., Knoll, R. L., and Wright, C. E. (1978). “The latency and duration of rapid movement sequences: Comparisons of speech and typewriting,” in *Information Processing in Motor Control and Learning* (Elsevier, New York), pp. 117–152.

Strangert, E. (1991). “Pausing in texts read aloud,” in *Proceedings of the XIIIth international congress of phonetic sciences* (Université de Provence, Service des Publications), 19–24 August, Aix-en-Provence, France, Vol. 4, pp. 238–241.

- Strangert, E. (1997). "Relating prosody to syntax: Boundary signalling in Swedish," in *Proceedings of the 5th European Conference on Speech Communication and Technology*, 22–25 September, Rhodes, Greece, pp. 239–242.
- Swets, B., Desmet, T., Hambrick, D. Z., and Ferreira, F. (2007). "The role of working memory in syntactic ambiguity resolution: A psychometric approach," *J. Exp. Psychol.* **136**, 64–81.
- Swets, B., Fuchs, S., Krivokapić, J., and Petrone, C. (2021). "A cross-linguistic study of individual differences in speech planning," *Front. Psychol.* **12**, 1439.
- Swets, B., Jacovina, M. E., and Gerrig, R. J. (2013). "Effects of conversational pressures on speech planning," *Discourse Process.* **50**, 23–51.
- Swets, B., Jacovina, M. E., and Gerrig, R. J. (2014). "Individual differences in the scope of speech planning: Evidence from eye-movements," *Lang. Cogn.* **6**, 12–44.
- Tilsen, S. (2020). "Detecting anticipatory information in speech with signal chopping," *J. Phon.* **82**, 100996.
- Tilsen, S., Spincemaille, P., Xu, B., Doerschuk, P., Luh, W.-M., Feldman, E., and Wang, Y. (2016). "Anticipatory posturing of the vocal tract reveals dissociation of speech movement plans from linguistic units," *PloS one* **11**, e0146813.
- Wagner, V., Jescheniak, J. D., and Schriefers, H. (2010). "On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access," *J. Exp. Psychol. Learn. Mem. Cogn.* **36**, 423–440.
- Watson, D., and Gibson, E. (2004). "The relationship between intonational phrasing and syntactic structure in language production," *Lang. Cogn. Proc.* **19**, 713–755.
- Zvonik, E., and Cummins, F. (2002). "Pause duration and variability in read texts," in *Proceedings of the 2002 International conference on spoken language processing (ICSLP '02)*, 16–20 September, Denver, CO, pp. 1109–1112.
- Zvonik, E., and Cummins, F. (2003). "The effect of surrounding phrase lengths on pause duration," in *Proceedings of Eurospeech 2003*, 1–4 September, Geneva, Switzerland, pp. 777–780.