



Predicting One's Turn With Both Body and Mind: Anticipatory Speech Postures During Dyadic Conversation

Peter A. Krause^{1,2*} and Alan H. Kawamoto²

¹ Department of Psychology, California State University Channel Islands, Camarillo, CA, United States, ² Department of Psychology, University of California, Santa Cruz, Santa Cruz, CA, United States

OPEN ACCESS

Edited by:

Martine Grice,
University of Cologne, Germany

Reviewed by:

Francisco Torreira,
McGill University, Canada
Jelena Krivokapic,
University of Michigan, United States

*Correspondence:

Peter A. Krause
peakraus@ucsc.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 23 March 2021

Accepted: 18 June 2021

Published: 13 July 2021

Citation:

Krause PA and Kawamoto AH
(2021) Predicting One's Turn With
Both Body and Mind: Anticipatory
Speech Postures During Dyadic
Conversation.
Front. Psychol. 12:684248.
doi: 10.3389/fpsyg.2021.684248

In natural conversation, turns are handed off quickly, with the mean downtime commonly ranging from 7 to 423 ms. To achieve this, speakers plan their upcoming speech as their partner's turn unfolds, holding the audible utterance in abeyance until socially appropriate. The role played by prediction is debated, with some researchers claiming that speakers predict upcoming speech opportunities, and others claiming that speakers wait for detection of turn-final cues. The dynamics of articulatory triggering may speak to this debate. It is often assumed that the prepared utterance is held in a response buffer and then initiated all at once. This assumption is consistent with standard phonetic models in which articulatory actions must follow tightly prescribed patterns of coordination. This assumption has recently been challenged by single-word production experiments in which participants partly positioned their articulators to anticipate upcoming utterances, long before starting the acoustic response. The present study considered whether similar anticipatory postures arise when speakers in conversation await their next opportunity to speak. We analyzed a pre-existing audiovisual database of dyads engaging in unstructured conversation. Video motion tracking was used to determine speakers' lip areas over time. When utterance-initial syllables began with labial consonants or included rounded vowels, speakers produced distinctly smaller lip areas (compared to other utterances), prior to audible speech. This effect was moderated by the number of words in the upcoming utterance; postures arose up to 3,000 ms before acoustic onset for short utterances of 1–3 words. We discuss the implications for models of conversation and phonetic control.

Keywords: articulation, motor control, speech planning, timing prediction, turn-taking

INTRODUCTION

Successful spoken communication requires navigating two overlapping sets of temporal constraints. On the one hand, there is what might be called phonological timing: how the flow of articulatory events gives rise to intelligible speech. Without proper phonological timing, the intended utterance “dab” might be distorted to “bad” (Browman and Goldstein, 1995). On the

other hand, there is what might be called situational timing: how phonetic events are timed against the background grid of the environment, including others' speech. Situational timing is key to inter-speaker coordination. For example, inter-turn gaps at changes of floor are quite short, with mean gap time varying from 7 to 423 ms across several languages (Stivers et al., 2009). As we will outline below, most extant speech models assume that phonological and situational timing are governed by distinct cognitive mechanisms. We will argue that understanding inter-speaker coordination requires re-evaluating this assumption. Such coordination may arise when speakers apply situational timing mechanisms to aspects of the utterance traditionally viewed as the domain of phonological timing.

Traditional models assume that utterance initiation is controlled by an online decision mechanism sensitive to situational factors like a "go" signal (Sternberg et al., 1978), or, when adapted to the context of conversation, another speaker's communicative cues (e.g., Levinson and Torreira, 2015; Levinson, 2016). However, once initiated, an utterance's internal timing is assumed to follow a prefabricated motor plan. In Levelt et al.'s (1999) influential model, this plan is a programmatic gestural score produced by the phonetic encoding mechanism. In Articulatory Phonology with Task Dynamics (AP/TD), this plan comprises the parameterized constriction gestures (Saltzman and Munhall, 1989), the phase couplings between gestural planning oscillators (Saltzman and Byrd, 2000), and the π -gestures implementing prosodic adjustments at phrase boundaries (Byrd and Saltzman, 2003).

There is empirical support for the separability of speech planning and speech triggering, both in pure laboratory tasks and in conversation tasks. For example, delayed naming tasks (e.g., Sternberg et al., 1978) attempt to isolate the speech triggering process by informing the participant of what they will say ahead of time, and then providing a secondary "go" signal to cue speech onset. The assumption is that participants will withhold the articulatory response until the "go" signal. Contrariwise, speeded naming tasks (e.g., Meyer, 1990, 1991) attempt to isolate the planning process by asking participants to respond as quickly as possible after the content of the next utterance is revealed. The assumption is that participants will complete planning and then initiate articulation as quickly as possible afterward. In delayed naming, participants generally reserve acoustic onset until after the "go" signal, and produce shorter acoustic latencies compared to speeded naming tasks. Both phenomena fit with the claim that triggering has been isolated. Similarly, in conversational tasks, EEG evidence suggests that relevant speech planning begins long before a partner finishes their current utterance (Bögels et al., 2015). However, as mentioned above, inter-turn gaps are short; further, Levinson and Torreira (2015) indicate that acoustically overlapped speech comprises less than 5% of total conversation time. In aggregate, this suggests that speakers in conversation plan upcoming utterances and then acoustically withhold them while awaiting the next speech opportunity. Not only does the evidence converge to the conclusion that planning and triggering are separable, but it also implies possible parallels between delayed naming and conversational speech initiation. We will return to this point later.

The separability of speech planning and speech triggering does not, on its own, entail the strict encapsulation of phonological from situational timing. Instead, we propose that these ideas have been accidentally conflated, partly because of the classical "motor program" concept. Work in delayed naming has revealed evidence that speakers preferentially "chunk" their utterances during the final moments of preparation. Sternberg et al. (1978) found acoustic response latency following the "go" signal to be a linear function of the number of stress-bearing syllables in the utterance. This work was later replicated and extended by Wheeldon and Lahiri (1997). The finding has been offered as evidence that stress-bearing syllables are the "subprograms" of speech, terminology which certainly implies fixed movement timing. Arguably, however, this interpretation reflects a preexisting commitment to the computer metaphor, as much as it reflects the specific empirical evidence.

The notion of "soft" movement plans is already well-ensconced in the phonetics literature, in the form of AP/TD's articulatory gestures. In that theory, planned gestures do not uniquely determine the spatial trajectories of articulators. Each gesture corresponds to an articulatory synergy (e.g., Browman and Goldstein, 1989); if the motion of one articulator is impeded, other articulators in the synergy can move differently to compensate. This affords the flexibility to adapt to unexpected situational events (such as perturbation of jaw motion during a bilabial closure, e.g., Folkins and Abbs, 1975; Kelso et al., 1984). Specific kinematic trajectories are emergent from the intersection of the gesture with the (dynamic, evolving) embedding context. It seems possible, at least in principle, that a plan for phonological timing could similarly comprise constraints (e.g., on the ordering and/or permissible overlap of actions) rather than a rigid specification of the behavioral time course. [Note that this is admittedly not the case in AP/TD itself; timing in that theory is prescribed by gestures' stiffness parameters, combined with the stable phasing relationships of coupled planning oscillators. But other touchstones exist. See, for example, Jordan, 1986; Liu and Kawamoto, 2010; Tilsen, 2016, 2018. This narrative was recently reviewed in detail by Krause and Kawamoto (2020b)].

These issues are highly relevant to the triggering of speech in conversation. Conversational utterance timing is precise. This is true not only for canonical turns (as represented by extremely short inter-turn gaps), but also for backchannels, which tend to be acoustically initiated following similar syntactic and prosodic cues as floor transitions (e.g., Koiso et al., 1998; Ward and Tsukahara, 2000) and which have a proper timing that is both perceptible and non-random (Poppe et al., 2011). One possibility is that this precision is aided by mechanisms that predict opportunities for speech onset. Most of the relevant evidence comes from the turn-taking literature. De Ruiter et al. (2006) found that participants could predict the timing of turn ends from lexico-syntactic cues. Magyari and de Ruiter (2012) found that listeners partly predicted turn-end phrasings and suggested this prediction could be used as a proxy estimate of remaining turn length. Rühlemann and Gries (2020) gave evidence that speakers progressively slow speech rate over most of the turn, implying that listeners might use prosodic cues in turn-end prediction. However, contrary to

the above, Bögels et al. (2015) reported that accurate turn-end detection required participants to hear turn-final intonational phrase boundaries.

Often overlooked is that the debated role of prediction in speech triggering is entangled with the issue of whether planning uniquely determines phonological timing. Before producing the earliest sounds of the utterance, speakers must first establish the initial constrictions in the vocal tract. Estimates and measurement practices vary, but Rastle et al. (2005) found the mean delay between articulatory and acoustic onset to range from 223 to 302 ms across syllable onset types. If speakers precisely time acoustic onset (e.g., by targeting no acoustic gap and no acoustic overlap at turn transitions, Sacks et al., 1974), they must work around this lead time. If this lead time is fixed by prior planning, a tricky problem arises. If a speaker waits to be certain a speech opportunity has arisen, they may initiate their utterance too late. The (un-compressible) lead time will then compound the delay preceding their acoustic response. If a speaker initiates their utterance from the predicted timing of a speech opportunity, then error in this prediction may lead them to start too early. The (un-expandable) lead time will then inexorably unfold to the point of an acoustic interruption. On this latter basis, Torreira et al. (2015) have argued that articulation is not initiated from predicted timing. Levinson and Torreira's (2015) and Levinson (2016) model of turn taking asserts that utterances are largely planned on a partner's turn but held in abeyance until the end of that turn is definitively detected.

To our knowledge, only one study has directly evaluated the late-initiation assumption using a conversational task. It therefore warrants specific consideration here. Torreira et al. (2015) analyzed breathing patterns of dyads during question-answer sequences. Specifically, they inspected the distribution of inbreath timings following the start of a question. This distribution was highly variable, but the mode fell 15 ms after question end. The authors reported this as evidence for late articulatory initiation. This study is an important first step in the area but has some critical limitations. One may question how well its restricted focus on question-answer sequences generalizes to both other kinds of turn exchanges and utterance types, such as backchannels, deliberately omitted from the turn-taking tradition. Further, we wonder whether the large variability in inbreath timings arose because a wider range of breathing strategies was in use than the study recognized. Finally, inbreath timing is not likely to index a fixed coordinative sequence for speech. For example, Mooshammer et al. (2019) found that acoustic response times were later for naming targets presented mid-inbreath, compared to ones presented mid-outbreath, suggesting speakers finished in-progress inhalations before initiating verbal responses. However, speakers also did not take new inbreaths, when presented with the target during exhalation.

Moreover, in naming research, articulatory kinematics often tell a different story from other measures. We earlier noted that conversational speech triggering invites comparison to the delayed naming paradigm. Classical findings in delayed naming,

based on acoustic measures, appeared to indicate that speech was not initiated until the "go" signal. However, when delayed naming experiments have added articulatory measures, the results have suggested a different interpretation, one seemingly incompatible with the fixed-time-course narrative. Both Kawamoto et al. (2014) and Tilsen et al. (2016) presented participants with monosyllables to be read aloud upon "go" signal presentation, while measuring articulator positions using either video or structural MRI. The "go" signal followed stimulus presentation by a variable delay. Participants postured their vocal tracts to anticipate form-specific requirements of the utterance. They formed and maintained these postures during the unpredictable period separating stimulus onset from "go" signal, while nonetheless delaying the acoustic response until appropriate.

This suggests that speakers in conversation may have heretofore unrecognized degrees of freedom for coordinating acoustic onset timing. The silent interval during which initial constrictions are formed may in fact be compressible or expandable, even after movement has started. This leads us to the following general hypothesis motivating this study: We propose that speakers in conversation can initiate the earliest articulatory movements from predicted timing, at least under some conditions. We further suggest that they compensate for prediction error online, by slowing or speeding the articulatory time course as it unfolds. A second general hypothesis follows by implication from the first: The earlier that pre-acoustic articulation is initiated (with respect to the eventual acoustic onset), the lower the peak velocity of that motion.

This mechanism may not be equally utilized (or available) across all contexts. Laboratory work examining articulatory strategies for speech suggests they respond to many factors. Consider studies examining incrementality (i.e., speakers' choices to produce speech by small chunks as they are planned, versus waiting and then producing large chunks all at once). Propensity to incremental speech can reflect individual differences (Kawamoto et al., 2014), subtleties of task (such as whether instructions were to begin speaking as soon as possible or to speak as briefly as possible, Holbrook et al., 2019), and even language spoken (Swets et al., 2021). Speaker's use of predictive initiation with adjustment may therefore vary with several factors. These factors might include how early (with respect to the targeted moment of acoustic onset) the initial words of the utterance are planned, how much of the utterance can be held in working memory, and/or the speaker's willingness to produce those early words incrementally. Overall, then, it may be that particularly short, stereotyped utterances are the most likely to be prepared in this manner, when disregarding other contextual factors. Knudsen et al. (2020) have recently suggested that speakers use such "forgotten little words" to mitigate conversational costs. The present study makes use of this observation.

We should emphasize that this tendency for articulatory preparation to arise more in some contexts than others is a core theme of the present study. It is not intended *per se* as a refutation of Levinson and Torreira's (2015) proposed rule for articulatory initiation. It is more so intended as a fundamental reframing of

the point, away from assertions of hard rules or typical cases, and toward consideration of the range of strategies available and their domains of application.

In the present study, we sought evidence for anticipatory speech postures in natural conversation. The study utilized data from the Cardiff Conversation Database (Aubrey et al., 2013), an audiovisual database of dyads engaging in unscripted conversations. Motion-tracked video of speakers' faces was used to assess changes in lip area over time. We looked for contrasts between utterances beginning with smaller lip area due to closure and/or rounding (the *labially constrained* condition) versus other oral configurations (the *labially unconstrained* condition). Further, we examined how these contrasts were moderated by the number of words in the utterance. We specifically predicted that lip areas for labially constrained vs. unconstrained utterances would be discriminable earlier, relative to acoustic onset, when utterances were very short (1–2 words long).

MATERIALS AND METHODS

This study used data drawn from the Cardiff Conversation Database (Aubrey et al., 2013). This database is available by request from <https://ccdb.cs.cf.ac.uk/signup.html>. The authors asked several dyads to engage in 5-min unscripted conversations while their faces were video-recorded (at 30 frames per second) and their speech audio-recorded. While the authors suggested possible conversation topics, these topics were not actively enforced. Full data coding and transcription has been completed for eight dyads; these are the dyads analyzed in the current study.

Participants

The eight analyzed dyads included six speakers. The following demographic information was provided by P. Rosin (personal communication, February 4, 2021; April 29, 2021). All speakers were Caucasian males. Ages ranged from 27 to 47 years ($M = 36.33$). Two speakers spoke English with a Welsh accent, one with a Scottish accent, one with a German accent, and two with an English accent (one having lived throughout Southern England, and one having grown up in Essex).

Each speaker participated in at least two dyads (equating to at least 10 min of recorded video footage). Details about the number of dyads each speaker participated in, plus additional information about their utterance distributions, appears in **Table 1**.

TABLE 1 | Speaker-specific information.

Speaker	Number of dyads	Mins. of recordings	Constrained utterances	Unconstrained utterances
P1	2	10	12	29
P2	3	15	12	50
P3	3	15	50	59
P4	2	10	19	23
P5	2	10	14	36
P6	3	15	19	29

Number of utterances refers only to retained data.

Elan Annotations

The database includes detailed, time-aligned behavioral annotations performed in the Elan software (Wittenburg et al., 2006) for all eight dyads. These annotations include temporal markings for the acoustic onsets and offsets of all verbal utterances, as well as full transcriptions of utterance content. Acoustic onsets for utterances were specifically marked at the first notable swell in audio intensity leading into an identifiable speech sound (P. Rosin, personal communication, April 29, 2021).

OpenFace Tracking Outputs

The database authors processed each facial video using OpenFace 2.0 (Baltrušaitis et al., 2018). The database contains the resulting output files. OpenFace detects the most prominent face in a video and tracks its pose in six degrees of freedom relative to camera origin, as well as tracking the three-dimensional positions of 128 key points on the face. We have previously described how to extract linguistically useful information about oral configuration from OpenFace output (Krause et al., 2020). OpenFace's positional estimates hew closest to the true values when OpenFace is provided with the camera's intrinsic lens parameters. OpenFace was not calibrated in this way before processing the facial videos in the database (P. Rosin, personal communication, February 3, 2021). However, since all statistical inference will be based on within-speaker comparisons, the lack of a pure correspondence to real-world units is incidental.

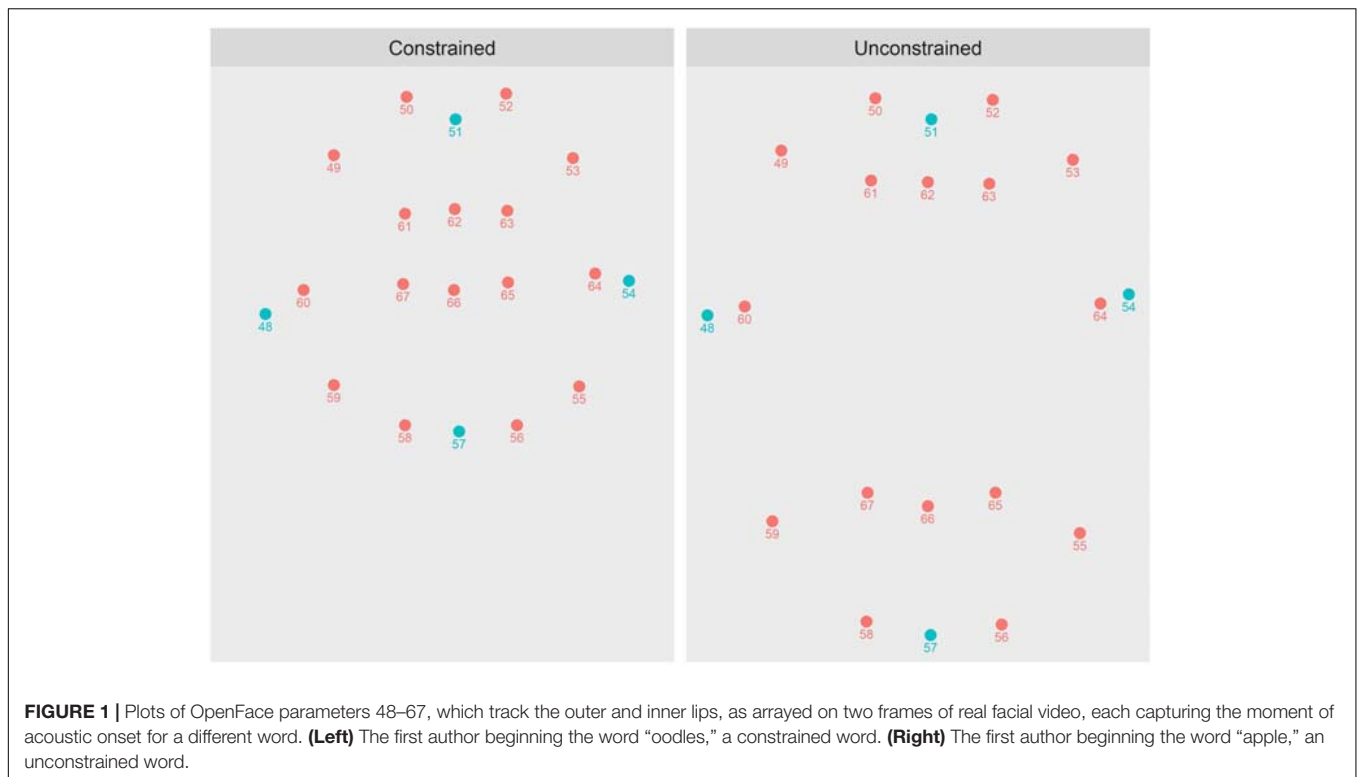
For readers unfamiliar with the OpenFace system, we provide **Figure 1** as illustration. The panels of **Figure 1** depict just those parameters of OpenFace that track the outer and inner lips (i.e., parameters 48–67). The dots colored in blue depict those parameters used in the lip-area computations described below. The panels specifically depict how OpenFace differentially tracks the lips when they are in different configurations (as in the *labially constrained* versus *labially unconstrained* utterance types, described in more detail below). To produce these plots, the first author spoke two example utterances on digital video that was later processed by OpenFace: the word “oodles” (a constrained utterance, left) and the word “apple” (an unconstrained utterance, right). The depicted tracking is from the moment of acoustic onset for both utterances. To facilitate comparison, both plots have been centered with respect to parameter 51, which marks the notch at the top of the outer lips.

Data Preparation

In total, the Python script described below identified 509 distinct utterances in the annotated Elan data.

Utterance Types and Inter-Utterance Gaps

A specialized Python script processed the Elan output for each speaker. For each annotated utterance, the script determined the most recently initiated prior utterance (even if this prior utterance was not yet concluded). If that recently initiated utterance was by the same speaker, the present utterance was labeled as a restart and marked to be dropped from the final dataset (our interest being in inter-speaker coordination). Utterances were also marked as restarts if the most recent



initiation was from the other speaker but had occurred within 500 ms of a prior initiation from the current speaker. Otherwise, utterances were marked as responses, and an inter-utterance gap time was computed by subtracting the onset time of the current utterance from the offset time of the most recently initiated prior utterance.

Similarly to Heldner and Edlund (2010), responses were further labeled as gaps (if the current utterance followed by previous one by a positive gap time), between-overlaps (if the current utterance started before the prior utterance completed, but finished afterward), and within-overlaps (if the current utterance lay completely within the acoustic boundaries of the prior utterance).

Lip Area Trajectories

By referencing the OpenFace outputs, the Python script computed lip area at each of 90 frames (3,000 ms) preceding the acoustic onset of each annotated turn. The script used the estimated x - and y -coordinates of four key points: The left-hand and right-hand corners of the lips (OpenFace parameters 48 and 54, respectively), and the external points at the center-top and center-bottom of the lips (OpenFace parameters 51 and 57, respectively). Area was computed as described by Liu et al. (accepted).

The strategy creates a perimeter of line segments running clockwise around the points, each conceived as the hypotenuse of a right triangle. Lip area is the sum of the areas of all four right triangles, plus the area of a residual central rectangle. Our formulas assume that one labels the left corner as (X_1, Y_1) and then increments the X - and Y -values while moving clockwise

around the set. The following formula produces the areas of each right triangle i (which has one corner at (X_i, Y_i)):

$$A_i = \frac{|X_i - X_{i+1}| \times |Y_i - Y_{i+1}|}{2}$$

The following formula produces the area of the residual central rectangle j :

$$A_j = |X_1 - X_3| \times |Y_2 - Y_4|$$

Identification of Key Predictor Values

The Python script also counted the number of words in the transcription for each turn and extracted the first transcribed word for reference.

We manually coded each turn as either *labially constrained* (in which case we would expect a comparatively small lip area at acoustic onset) or as *labially unconstrained*. We made this assessment based on the first syllable of the first annotated word. Specifically, we considered both the initial consonant (if applicable) and the nuclear vowel. Turns with initial consonants were classified as labially constrained if the consonant was bilabial, labiodental, or rounded, i.e., a member of the set /b, f, m, p, ɸ, v, w/. Regardless of initial consonant, turns were classified as labially constrained if their first nuclear vowel was rounded, i.e., a member of the set /ɔ, o, ʊ, u/. Turns not fitting either of these criteria were classified as labially unconstrained.

RESULTS

Utterance Types and Inter-Utterance Gap Times

Restarts and the initial utterances of a conversation were expunged from the data, leaving 352 utterances in the set. Gap utterances comprised 47% of retained data, between-overlaps 22%, and within-overlaps 30%.

Because the nominally intended start times for gap utterances and between-overlaps are relatively clear (i.e., the acoustic offset of the prior utterance) their coordination can be further characterized by considering their inter-utterance gap times (i.e., their floor transfer offsets, Levinson and Torreira, 2015). Figure 2 depicts a histogram of the floor transfer offsets in the final data. The mean floor transfer offset was 258.37 ms (SD = 914.96).

The coordination of within-overlaps cannot be so easily characterized, since these are typically backchannels or other short utterances whose timing targets are a syntactic or prosodic closure within an utterance that is acoustically continued. However, spot-checking of the transcript suggested that within-overlaps were coordinated reasonably.

For example, the most extreme within-overlap marked in the dataset arises in a conversation between P3 and P4. At this point, P4 has just been asked his favorite biography. Comparison of the Elan annotations with the acoustic waveforms reveals this rough pattern of coordination (bolded text indicates the specific point of overlap):

P4: The, the one on Copeland one Aaron Copeland one was a good one, **uhh**, a large volume. . .

P3: Okay.

P4's turn continues for some time after without an acoustic break, resulting in the inter-utterance gap time for P3's "Okay"

being computed as -99 s, despite it having been realized shortly after the resolution of P4's statement "Aaron Copeland was a good one."

All within-overlaps were retained in the final dataset so as not to compromise statistical power for detecting anticipatory postures.

Lip Area Trajectories

Statistical analysis of lip area was carried out in R version 4.1.0 ("Camp Pontanezen"), using the "lme4," "lmerTest," "interactions," "effects," and "bootMer" packages. We fit linear mixed-effects models to the data at 15-frame (500-ms) intervals, starting at 3,000 ms prior to acoustic onset, and ending at acoustic onset. This resulted in 7 fit models. The dependent variable in each case was lip area, with predictor variables of interest being utterance word count (log10-transformed to correct for extreme positive skewness), labial constraint (constrained vs. unconstrained), and their two-way interaction. We used Speaker ID and the first word of the utterance as clustering variables. We determined the random effects structure for each model via a backward selection procedure that started with the maximal structure (Barr et al., 2013) and then simplified until lme4 returned no warnings about convergence failure or model singularity. To facilitate hypothesis testing, we estimated denominator degrees of freedom via Satterthwaite's method (see Luke, 2017 for a justification).

Our labial constraint predictor was dummy-coded "0" for constrained utterances and "1" for unconstrained utterances. Therefore, positive slope values would indicate larger lip area for unconstrained than constrained utterances, signaling the likely presence of speech postures. This gives the slopes of the word count \times labial constraint interaction terms a straightforward interpretation. Negative values for these slopes would indicate

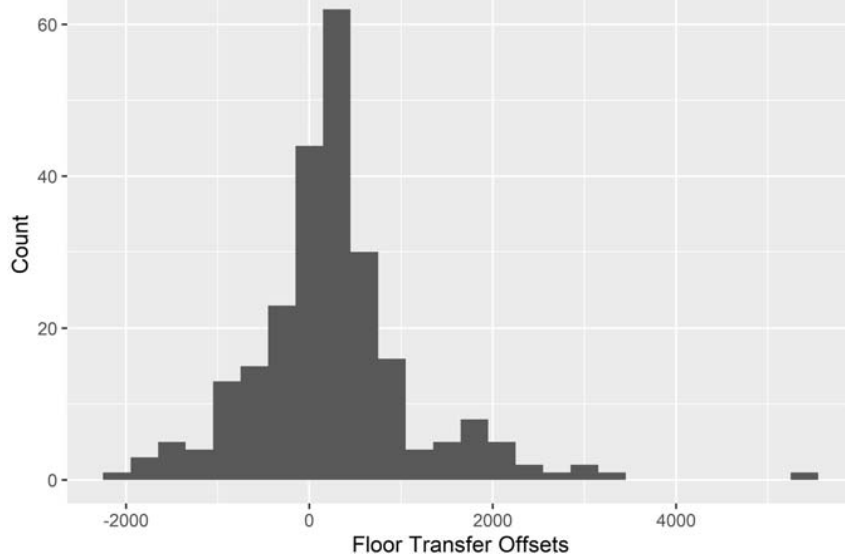


FIGURE 2 | A histogram depicting the floor transfer offsets (i.e., the inter-utterance gaps for gap utterances and between-overlaps) in the final dataset.

decreasing probability and/or magnitude of speech postures with increasing word count, consistent with our hypothesis. When the interaction term was statistically reliable, we followed up by computing Johnson-Neyman intervals revealing the specific ranges of word count over which the labial constraint predictor differed from 0.

Table 2 reports the statistical results in detail, including the random effects structures of the final fitted models. Where we report the results of Johnson-Neyman intervals, we give ranges of actual word counts (as opposed to their log-10 transformations), and report only those ranges where the slope of labial constraint was both positive and within the observed span of the data.

The pattern of results in Table 2 is consistent with our prediction that anticipatory postures would emerge earlier (resulting in earlier distinctions between labially constrained vs. unconstrained turns) for shorter utterances. The word count \times labial constraint interaction was statistically reliable from $-3,000$ ms through $-1,000$ ms, with Johnson-Neyman intervals suggesting that postures were most likely for utterances of three words or fewer at the earliest time points. At the end of the time course this interaction (unsurprisingly) disappears, leaving only a reliable main effect of labial constraint at 0 ms.

Figure 3 presents a more visual illustration of these effects. For each of the seven models it plots the predicted lip area values (with bootstrapped 95% confidence intervals) for labially constrained and unconstrained utterances of two and eight words in length. (These values were chosen because two words is decidedly within the span at which the earliest anticipatory postures arose, while eight words is decidedly outside it, without being substantially larger).

Maximum Lip Movement Speed

In the Introduction, we suggested that speakers might increase articulatory movement speed to partly offset delays from planning difficulty, meaning faster movement speeds would be expected for utterances of more words.

We analyzed lip movement speeds for all utterances. We did this by computing the change in area at each of the final 15 steps of the analysis window (when lips should be approaching final configuration targets) and multiplying each discrete change by 30 so that the units were given as mm^2/s . We determined the maximum lip movement speed by selecting the resulting value with the largest absolute magnitude. We then fit a linear mixed-effects model as described above, with this absolute magnitude as the dependent variable. Log10-transformed word count, labial constraint, and their two-way interaction were the predictors. The final fitted model included random intercepts for Speaker ID and first word, and a random slope for labial constraint at the Speaker ID level. The only reliable effect was a main effect of word count ($\beta = 455.22$, $SE = 185.15$), $t(64.73) = 2.50$, $p = 0.02$. This effect suggests that, as predicted, lip movement speeds increase as word count increases. Figure 4 depicts the regression line for word count (with 95% confidence ribbon).

Content of One-Word Utterances

The results suggest that the earliest anticipatory speech postures arose for utterances of three or fewer words. Although

TABLE 2 | Reports of linear mixed models fit to lip area.

Time point (ms)	Model details	
-3,000	Random effects	(1 speaker) + (1 first word)
	Constraint \times log10(word count)	$\beta = -64.48$, $SE = 24.99$, $t(109.35) = -2.58^*$
	Main effect of constraint Johnson-Neyman interval for probable speech postures	$\beta = 68.73$, $SE = 25.96$, $t(20.81) = 2.65^*$ [0, 3.09]
-2,500	Random effects	(log10(word count) speaker)
	Constraint \times log10(word count)	$\beta = -61.86$, $SE = 24.14$, $t(291.77) = -2.56^*$
	Main effect of constraint Johnson-Neyman interval for probable speech postures	$\beta = 66.33$, $SE = 23.62$, $t(251.79) = 2.81^{**}$ [0, 3.72]
-2,000	Random effects	(log10(word count) speaker) + (1 first word)
	Constraint \times log10(word count)	$\beta = -58.07$, $SE = 25.09$, $t(113.91) = -2.32^*$
	Main effect of constraint Johnson-Neyman interval for probable speech postures	$\beta = 71.78$, $SE = 25.83$, $t(20.80) = 2.78^*$ [0, 4.47]
-1,500	Random effects	(log10(word count) speaker)
	Constraint \times log10(word count)	$\beta = -90.33$, $SE = 25.56$, $t(299.03) = -3.53^{***}$
	Main effect of constraint Johnson-Neyman interval for probable speech postures	$\beta = 105.23$, $SE = 24.96$, $t(253.49) = 4.22^{***}$ [0, 6.45]
-1,000	Random effects	(log10(word count) speaker) + (1 first word)
	Constraint \times log10(word count)	$\beta = -70.37$, $SE = 27.22$, $t(191.49) = -2.59^*$
	Main effect of constraint Johnson-Neyman interval for probable speech postures	$\beta = 82.24$, $SE = 29.40$, $t(51.24) = 2.80^{**}$ [0, 3.89]
-500	Random effects	(log10(word count) speaker) + (1 first word)
	Constraint \times log10(word count)	$\beta = -42.84$, $SE = 28.79$, $t(277.68) = -1.49$
	Main effect of constraint Johnson-Neyman interval for probable speech postures	$\beta = 62.23$, $SE = 32.81$, $t(97.14) = 1.90^\dagger$ N/A
0	Random effects	(log10(word count) speaker) + (1 first word)
	Constraint \times log10(word count)	$\beta = -35.16$, $SE = 24.35$, $t(267.60) = -1.44$
	Main effect of constraint Johnson-Neyman interval for probable speech postures	$\beta = 97.83$, $SE = 27.63$, $t(89.50) = 3.54^{***}$ N/A

Random effects were first extracted as numeric vectors. Double-bar notation indicates uncorrelated random effects. Although Word Count was log10-transformed, Johnson-Neyman intervals are given here as number of words, for ease of interpretation.

$^\dagger p < 0.1$.

$^* p < 0.05$.

$^{**} p < 0.01$.

$^{***} p < 0.001$.

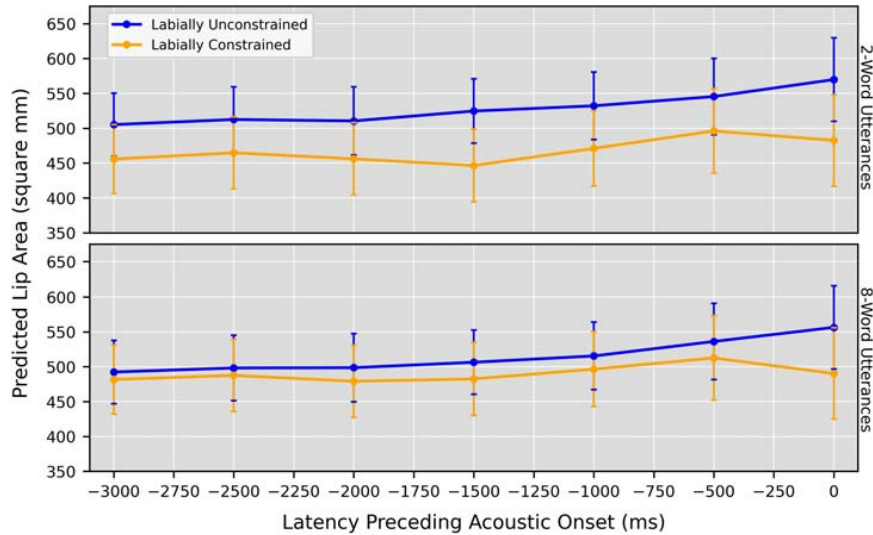


FIGURE 3 | Predicted lip area values produced by the linear mixed models, when setting the word count predictor to 2 words and 8 words. Models were fit to junctures at 15-frame (500-ms) intervals, starting at 90 frames (3,000 ms) preceding acoustic onset. Error bars: Bootstrapped 95% CI.

our hypothesis proceeded from the assumption that shorter utterances are easier to plan, there are multiple confounded reasons that this might be. For example, the smaller number of words might *per se* lower planning complexity, but the communicative content itself might also be simpler or higher in frequency. For this reason, some readers may wish to get a sense of the content of these shorter utterances. Table 3 presents a frequency-ordered list of every type of one-word utterance in the final dataset (one-word utterances representing 62% of all utterances of three or fewer words).

anticipatory oral postures well in advance of starting their acoustic utterance. In this study, statistical detection of these postures was facilitated when utterances were especially short (1–3 words). This may suggest that the articulatory planning and/or control of these postures is made easier when the overall utterance is low in complexity. As such, although this study extends the findings of a previous study of utterance timing in conversation (Torreira et al., 2015), it may not directly contradict that prior study’s finding that utterances were

DISCUSSION

The main finding of the present study is that speakers awaiting their turn in natural conversation sometimes produce

TABLE 3 | Content and frequency of 1-word utterances.

Word	Count	Labial constraint
Yeah	37	Unconstrained
Alright	13	Constrained
Right	9	Constrained
No	8	Constrained
Yep	5	Unconstrained
Yes	4	Unconstrained
Really	3	Constrained
So	2	Constrained
Excellent	1	Unconstrained
Mmhmm	1	Constrained
Next	1	Unconstrained
Nice	1	Unconstrained
Oh	1	Constrained
Ok	1	Constrained
Thanks	1	Unconstrained
that’s	1	Unconstrained
Very	1	Constrained
Well	1	Constrained
What	1	Constrained
Which	1	Constrained

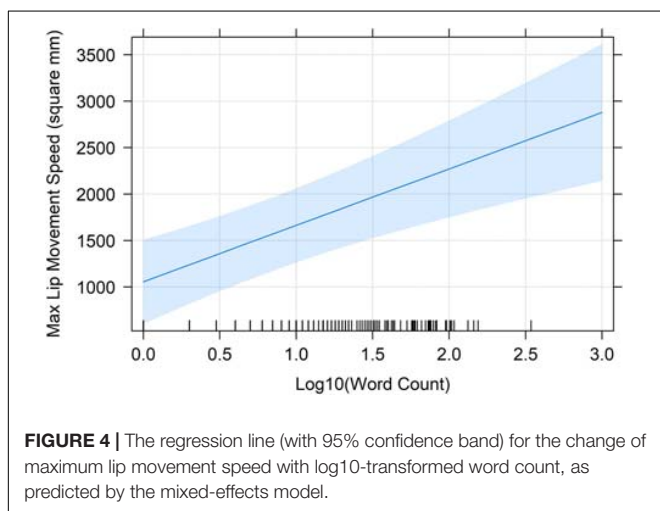


FIGURE 4 | The regression line (with 95% confidence band) for the change of maximum lip movement speed with log10-transformed word count, as predicted by the mixed-effects model.

initiated late, since the utterances in the earlier study were of higher complexity.

Prior studies have reported similar effects for isolated word production tasks (Kawamoto et al., 2014; Drake and Corley, 2015; Tilsen et al., 2016; Krause and Kawamoto, 2019, 2020a; Tilsen, 2020). We believe, however, that this is the first report of such effects in an ecological conversation task. In addition, while Tilsen (2020) recently found that articulatory postures could arise as long as 500 ms before canonical articulatory onset, the current study is remarkable in finding that very short utterances could see postures arise as long as 3,000 ms before acoustic onset.

Limitations

The non-experimental nature of our approach leaves the causal underpinnings of these postures underdetermined. We assumed that the short utterances would be faster to plan, leading to more cases of delay between phonological availability and acoustic response opportunity. However, the observed pattern may instead arise from some correlate of utterance length.

The analysis reflects 352 utterances produced by six male Caucasian speakers of English; generality may be limited. However, because anticipatory posturing has not previously been reported in an ecological task, the results are intrinsically important. They provide a valuable existence proof for the narrative of flexible articulatory control outlined in the Introduction, whereby speakers can independently manipulate utterance initiation and the time course of articulation. Further, the results suggest fertile possibilities for experimental follow-up.

Theoretical Implications

We propose that, under some conditions, speakers can initiate articulation from ongoing prediction of the next speech opportunity, while using online control to finesse the moment of acoustic onset. Outstanding questions at this juncture include why postures are more likely under some conditions than others, and whether the postures are strategically functional.

Why Does the Emergence of Postures Vary?

As noted earlier, Torreira et al. (2015) concluded that articulation was initiated just after a partner's utterance ended. Although this may reflect the specific dependent measure used (inbreaths, as indexed via inductive plethysmography), it may also be that the specific utterances analyzed did not facilitate early articulation. In the present study, anticipatory posturing only verifiably arose for utterances of 1–3 words in length. Although we predicted that postures should be variable, in accordance with the strategic flexibility observed in past articulatory studies, it remains uncertain exactly how this variability is structured.

We presumed that number of words in the utterance indexed planning complexity. If we are correct, then the phonology of more complex utterances might become available later, relative to the targeted moment of acoustic onset, leaving a shorter span inside which anticipatory postures could emerge and reduce movement speeds. Admittedly, however, “planning complexity” is ambiguous here. It could be that having fewer words to plan places less strain on the phonological system. It could also be that utterances that perform certain communicative functions

or comprise certain high-frequency phrases are planned more easily, and that these utterances incidentally tend to be shorter. Further, the emergence of postures for certain kinds of utterances might reflect, not simpler planning, but some difference in either conventional timing or the need to visually signal intent (see below).

It might also be that the number of words in an utterance somehow moderates the readiness with which phonological plans are conferred into action. In their classic delayed-naming study, Sternberg et al. (1978) found that longer prepared utterances had longer acoustic latencies following the go signal. Although the present study is one of several to show that articulatory motion can be partly de-coupled from acoustic onset, perhaps this de-coupling becomes more difficult to manage as the upcoming utterance grows in length.

Are Postures Strategically Functional?

Possibility 1: intention leaks into articulation

One possibility is that the emergence of postures is not strategic but is instead a passive, incidental consequence of how the planning and motor systems are coupled. We base this possibility on Tilsen's (2019) explanation of the anticipatory postures observed in laboratory tasks. In this account, the postures arise when partly activated but unselected speech gestures cascade their influences into current articulatory targets, across a partly permeable threshold.

This proposal can account for anticipatory speech postures in which the phonologically relevant constrictions are only partly formed (which Tilsen et al., 2016, found to be common). However, not all anticipatory postures are partial. In video recordings of their delayed naming task, Kawamoto et al. (2014) observed speakers who both closed their lips and accumulated intra-oral pressure when preparing /p/-initial utterances. (A comparable observation is not viable for the present study, owing to the small number of spontaneously produced utterances happening to start with bilabial plosives).

Possibility 2: reduction of movement costs

Starting early (i.e., lengthening posture duration) may minimize movement costs, as would be predicted by optimal control theories of speech behavior (e.g., Nelson et al., 1984). We observed that maximum lip movement speed was relatively lower over the ranges of word counts at which anticipatory postures arose. For a frictionless system, peak velocity is an estimate of the integral of force applied per unit mass with respect to time (Nelson, 1983). Such force integrals are one candidate approach for quantifying the energy and/or effort costs of motor function (see, e.g., Turk and Shattuck-Hufnagel, 2020).

Possibility 3: social signaling

Testing of Levinson and Torreira's (2015) model has largely emphasized listeners' abilities to predict and prepare for floor yielding. However, the anticipatory postures observed in this study may reflect the listener's agency in effecting speech opportunities. It is possible that listeners deliberately use them to indicate readiness to speak. Irrespective of whether the behavior is deliberate, it may be perceived as a social signal by the current speaker. It might also assist the current speaker in

managing attention, such that the incoming utterance is a less surprising event.

This proposal has some precedence. Kendon (1967) reported that as a speaker approached a possible floor transition, they tended to shift their gaze to the listener; this gaze continued for a while all the way through the transition. Listeners sometimes visually signaled a readiness to take the floor prior to the gaze shift. Similarly, Bavelas et al. (2002) found that brief windows of mutual gaze between speaker and listener often preceded backchannels. Kendrick and Holler (2017) presented evidence that listeners' gaze patterns could influence the end of speakers' turns. Averted gazes signaled that listeners were planning dispreferred responses; in some cases these gazes yielded last-minute repairs by the speaker intended to eliminate the dispreferred response.

Concluding Remarks

The current study presented preliminary evidence that speech planning and articulation may flexibly overlap in natural conversation. This suggests that articulation may at times be initiated based on the predicted timing of speech opportunities, without obligating an acoustic interruption. Future work will be necessary to determine exactly what mechanisms speakers use to regulate the time course of articulation, how much

deliberate strategy is involved, and whether this phenomenon carries social/pragmatic implications.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://doi.org/10.17605/OSF.IO/JKR96>.

AUTHOR CONTRIBUTIONS

PK carried out the data encoding and analysis and wrote the entire manuscript. Both authors jointly conceived and designed this study.

FUNDING

Publication fees for the present article were funded in part by the Department of Psychology at California State University Channel Islands.

REFERENCES

- Aubrey, A. J., Marshall, D., Rosin, P. L., and Vandeventer, J. (2013). "Cardiff Conversation Database (CCDb): a database of natural dyadic conversations," in *Proceedings of the V & L Net Workshop on Language for Vision, 2013*, Portland, OR.
- Baltrušaitis, T., Zadeh, A., Lim, Y. C., and Morency, L. (2018). "OpenFace 2.0: facial Behavior Analysis Toolkit," in *Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)(FG)*, Xi'an, 59–66. doi: 10.1109/FG.2018.00019
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278.
- Bavelas, J., Coates, L., and Johnson, T. (2002). Listener responses as a collaborative process: the role of gaze. *J. Commun.* 52, 566–580. doi: 10.1111/j.1460-2466.2002.tb02562.x
- Bögels, S., Magyari, L., and Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Sci. Rep.* 5:12881. doi: 10.1038/srep12881
- Browman, C. P., and Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology* 6, 201–251. doi: 10.1017/S0952675700001019
- Browman, C. P., and Goldstein, L. (1995). "Dynamics and articulatory phonology," in *Mind as Motion: Explorations in the Dynamics of Cognition*, eds R. F. Port and T. van Gelder (Cambridge, MA: The MIT Press), 175–193.
- Byrd, D., and Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *J. Phon.* 31, 149–180. doi: 10.1016/S0095-4470(02)00085-2
- De Ruiter, J. P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130
- Drake, E., and Corley, M. (2015). Articulatory imaging implicates prediction during spoken language comprehension. *Mem. Cogn.* 43, 1136–1147. doi: 10.3758/s13421-015-0530-6
- Folkins, J. W., and Abbs, J. H. (1975). Lip and jaw motor control during speech: responses to resistive loading of the jaw. *J. Speech Hear. Res.* 18, 207–220. doi: 10.1044/jshr.1801.207
- Heldner, M., and Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *J. Phon.* 38, 555–568. doi: 10.1016/j.wocn.2010.08.002
- Holbrook, B. B., Kawamoto, A. H., and Liu, Q. (2019). Task demands and segment priming effects in the naming task. *J. Exp. Psychol. Learn. Mem. Cogn.* 45, 807–821. doi: 10.1037/xlm0000631
- Jordan, M. I. (1986). *Serial Order: A Parallel Distributed Processing Approach (Tech. Rep. No. 8604)*. San Diego, CA: University of California, Institute for Cognitive Science.
- Kawamoto, A. H., Liu, Q., Lee, R. J., and Grebe, P. R. (2014). The segment as the minimal planning unit in speech production: evidence from absolute response latencies. *Q. J. Exp. Psychol.* 67, 2340–2359. doi: 10.1080/17470218.2014.927892
- Kelso, J. S., Tuller, B., Vatikiotis-Bateson, E., and Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 812–832. doi: 10.1037//0096-1523.10.6.812
- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychol.* 26, 22–63. doi: 10.1016/0001-6918(67)90005-4
- Kendrick, K., and Holler, J. (2017). Gaze direction signals response preference in conversation. *Res. Lang. Soc. Interact.* 50, 12–32. doi: 10.1080/08351813.2017.1262120
- Knudsen, B., Creemers, A., and Meyer, A. (2020). Forgotten little words: how backchannels and particles may facilitate speech planning in conversation? *Front. Psychol.* 11:593671. doi: 10.3389/fpsyg.2020.593671
- Koiso, H., Yasuo, H., Tutiya, S., Ichikawa, A., and Den, Y. (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. *Lang. Speech* 41, 295–321. doi: 10.1177/002383099804100404
- 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. doi: 10.1037/xhp0000610
- Krause, P. A., and Kawamoto, A. H. (2020a). 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. (2020b). On the timing and coordination of articulatory movements: historical perspectives and current theoretical challenges. *Lang. Linguist. Compass* 14:e12373. doi: 10.1111/lnc3.12373

- Krause, P. A., Kay, C. A., and Kawamoto, A. H. (2020). Automatic motion tracking of lips using digital video and OpenFace 2.0. *Lab. Phonol. J. Assoc. Lab. Phonol.* 11, 9, 1–16. doi: 10.5334/labphon.232
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38. doi: 10.1017/S0140525X99001776
- Levinson, S. C. (2016). Turn-taking in human communication—origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010
- Levinson, S. C., and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Front. Psychol.* 6:731. doi: 10.3389/fpsyg.2015.00731
- Liu, Q., Holbrook, B. H., Kawamoto, A. H., and Krause, P. A. (accepted). Verbal reaction times based on tracking lip movement. *Ment. Lex.*
- Liu, Q., and Kawamoto, A. H. (2010). “Simulating the elimination and enhancement of the plosivity effect in reading aloud,” in *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, eds S. Ohlsson and R. Catrambone (Austin, TX: Cognitive Science Society), 2284–2289.
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behav. Res. Methods* 49, 1494–1502. doi: 10.3758/s13428-016-0809-y
- Magyari, L., and de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. doi: 10.3389/fpsyg.2012.00376
- Meyer, A. S. (1990). The time course of phonological encoding in language production: the encoding of successive syllables of a word. *J. Mem. Lang.* 29, 525–545. doi: 10.1016/0749-596X(90)90050-A
- Meyer, A. S. (1991). The time course of phonological encoding in language production: phonological encoding inside a syllable. *J. Mem. Lang.* 30, 69–89. doi: 10.1016/0749-596X(91)90011-8
- Mooshammer, C., Rasskazova, O., Zöllner, A., and Fuchs, S. (2019). “Effect of breathing on reaction time in a simple naming experiment: Evidence from a pilot experiment,” in *Paper Presented at the International Seminar on the Foundations of Speech*, Sønderborg.
- Nelson, W. L. (1983). Physical principles of economies of skilled movements. *Biol. Cybern.* 46, 135–147. doi: 10.1007/bf00339982
- Nelson, W. L., Perkell, J. S., and Westbury, J. R. (1984). Mandible movements during increasingly rapid articulations of single syllables: preliminary observations. *J. Acoust. Soc. Am.* 75, 945–951. doi: 10.1121/1.390559
- Poppe, R., Truong, K., and Heylen, D. (2011). “Backchannels: Quantity, type, and timing matters,” in *Proceedings of the Intelligent Virtual Agents—11th International Conference, IVA 2011*, Reykjavik.
- Rastle, K., Croot, K. P., Harrington, J. M., and Coltheart, M. (2005). Characterizing the motor execution stage of speech production: consonantal effects on delayed naming latency and onset duration. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 1083–1095. doi: 10.1037/0096-1523.31.5.1083
- Rühlemann, C., and Gries, S. T. (2020). Speakers advance-project turn completion by slowing down: a multifactorial corpus analysis. *J. Phon.* 80:100976. doi: 10.1016/j.wocn.2020.100976
- Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simple systematic for the organization of turn taking in conversation. *Language* 50, 696–735. doi: 10.2307/412243
- Saltzman, E., and Byrd, D. (2000). Task-dynamics of gestural timing: phase windows and multifrequency rhythms. *Hum. Mov. Sci.* 19, 499–526. doi: 10.1016/S0167-9457(00)00030-0
- Saltzman, E. L., and Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecol. Psychol.* 1, 333–382. doi: 10.1207/s15326969eco0104_2
- Sternberg, S., Monsell, S., Knoll, R. L., and Wright, C. E. (1978). “The latency and duration of rapid movement: Comparisons of speech and typewriting,” in *Information Processing in Motor Control and Learning*, ed. G. E. Stelmach (New York, NY: Academic Press), 117–152.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci.* 106, 10587–10592. doi: 10.1073/pnas.0903616106
- Swets, B., Fuchs, S., Krivokapić, J., and Petrone, C. (2021). A cross-linguistic study of individual differences in speech planning. *Front. Psychol.* 12:655516. doi: 10.3389/fpsyg.2021.655516
- Tilsen, S. (2016). Selection and coordination: the articulatory basis for the emergence of phonological structure. *J. Phon.* 55, 53–77. doi: 10.1016/j.wocn.2015.11.005
- Tilsen, S. (2018). “Three mechanisms for modeling articulation: Selection, coordination, and intention,” in *Cornell Working Papers in Phonetics and Phonology, 2018* (Ithaca, NY: Cornell University).
- Tilsen, S. (2019). Motoric mechanisms for the emergence of non-local phonological patterns. *Front. Psychol.* 10:2143. doi: 10.3389/fpsyg.2019.02143
- Tilsen, S. (2020). Detecting anticipatory information in speech with signal chopping. *J. Phon.* 82, 1–27.
- Tilsen, S., Spincemaille, P., Xu, B., Doerschuk, P., Luh, W.-M., Feldman, E., et al. (2016). Anticipatory posturing of the vocal tract reveals dissociation of speech movement plans from linguistic units. *PLoS One* 11:e0146813. doi: 10.1016/j.wocn.2012.08.005
- Torreira, F., Bögels, S., and Levinson, S. C. (2015). Breathing for answering: the time course of response planning in conversation. *Front. Psychol.* 6:284. doi: 10.3389/fpsyg.2015.00284
- Turk, A., and Shattuck-Hufnagel, S. (2020). *Speech Timing: Implications for Theories of Phonology, Phonetics, and Speech Motor Control*. Oxford: Oxford University Press.
- Ward, N., and Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. *J. Pragmat.* 32, 1177–1207. doi: 10.1016/S0378-2166(99)00109-5
- Wheeldon, L., and Lahiri, A. (1997). Prosodic units in speech production. *J. Mem. Lang.* 37, 356–381. doi: 10.1006/jmla.1997.2517
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). “ELAN: a Professional Framework for Multimodality Research,” in *Proceedings of the LREC 2006, 5th International Conference on Language Resources and Evaluation*, Genoa.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Krause and Kawamoto. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.