Title
Task Demands and Segment Priming Effects in the Naming Task

Permalink
https://escholarship.org/uc/item/15t8h7j3

Journal
Journal of Experimental Psychology Learning Memory and Cognition, 45(5)

ISSN
0278-7393

Authors
Holbrook, Bryan B
Kawamoto, Alan H
Liu, Qiang

Publication Date
2019-05-01

DOI
10.1037/xlm0000631

Copyright Information

Task Demands and Segment Priming Effects in the Naming Task

Bryan B. Holbrook, Alan H. Kawamoto, and Qiang Liu

University of California Santa Cruz

Author Note

Bryan B. Holbrook, Department of Psychology, University of California Santa Cruz; Alan H. Kawamoto, Department of Psychology, University of California Santa Cruz; Qiang Liu, Department of Psychology, University of California Santa Cruz.

Qiang Liu is now at Electronic Research Laboratory, Volkswagen Group of America, Inc.

Correspondence concerning this article should be addressed to Bryan B. Holbrook, Department of Psychology, University of California Santa Cruz, 1156 High St., Santa Cruz, CA, 95064. Contact: bholbroo@ucsc.edu

Abstract

A central issue in the study of speech production is whether phonological encoding occurs sequentially or in parallel. Some of the strongest evidence for sequential phonological encoding comes from the number of segments primed effect – response latencies decrease when increasing the number of primed segments from 0 to 1 to 2 (e.g., Meyer, 1991). While it is often assumed that all participants adopt the same response criterion in the naming task, standard instructions can lead to the strategic adoption of different response criteria (such as an initial segment-based criterion or a syllable-based criterion). Furthermore, the number of segments primed effect might be driven by the manner of the initial segment such as the acoustic characteristics of plosives. In this study, participants named monosyllabic words varying in initial segment plosivity in a 0, 1, or 2 segments primed naming task, and were instructed in ways to induce either a segment or syllable criterion. Data were analyzed by acoustic latency, articulatory latency, and initial segment duration, as distinguishing between a segment and syllable criterion and sequential and parallel encoding requires more than just a single point in the time-course of articulation. Shorter acoustic latencies when priming 2 segments over 1 were contingent on the manner of the initial segment and the adoption of a segment criterion, clarifying the nature of the number of segments primed effect. Moreover, the similar acoustic latencies found across priming conditions when a syllable criterion was adopted support parallel phonological encoding.

Keywords: speech production, incremental articulation, phonological encoding, task demands, response criterion, parallel encoding

Task Demands and Segment Priming Effects in the Naming Task

The relation between motor planning and execution is an important issue in the study of action and behavior. For example, when speaking a word, are the segments (i.e., consonants and vowels) planned sequentially or in parallel, and do speakers begin to articulate the word as soon as the initial segment has been planned or must they wait until all segments have been planned? Making matters more complicated, initiating articulation based solely on the initial segment might not always occur, as speakers can vary in their *response criteria* – the amount of information designated to be planned before the speaker decides to initiate articulation. While much previous work has shown that speakers tend to use a syllable-based response criterion (*syllable criterion;* Levelt, Roelofs, & Meyer, 1999a), emerging evidence also suggests that speakers, under certain circumstances, will adopt an initial segment-based response criterion (*segment criterion;* Kawamoto, 1999). This distinction is crucial, as the predictions made by sequential (Levelt et al., 1999a) and parallel (Dell, 1986; Kawamoto, Kello, Jones, & Bame, 1998; Liu, Kawamoto, Payne, & Dorsey, 2017) accounts of phonological encoding vary depending on the adopted response criterion.

To evaluate the predictions made by both of these accounts, the time-course of speaking a single monosyllabic word will be assessed by presenting words with either no segment information known, the first segment known, or the first two segments known. We can then manipulate participants' response criteria to determine the basis for the shorter response latencies found as the number of primed initial segments in a target word increases from 0 to 2 (Meyer, 1991), a finding we will refer to as the *number of initial segments primed effect*.

In doing so, we consider two different models that can account for the number of initial segments primed effect: (1) the incremental encoding model (Levelt et al., 1999a; Roelofs, 1997), and (2) the incremental articulation model (Kawamoto et al., 1998; Liu, et al., 2017). While both models can account for the number of initial segments primed effect, they do so by making different assumptions about the nature of phonological encoding and the response criteria adopted by speakers at the articulatory stage. The incremental encoding model assumes sequential phonological encoding and a syllable criterion, and the incremental articulation model assumes parallel phonological encoding and a segment criterion. To clearly illustrate these different assumptions, we must walk through the main stages involved in speech production and explore how priming segments impacts this process.

Speech Production

When speaking a word spontaneously or cued by an object, the general process begins with conceptualization of the to-be-spoken word (Figure 1). Conceptualization then cues lemma retrieval, leading to word-form encoding processes and eventual articulation. When reading a word aloud, however, conceptualization is bypassed and phonological encoding begins based on the perception of the word-form. Considering that much of the evidence for the processes that we will be discussing comes from studies of picture naming (Roelofs, 2004), cued naming (Meyer, 1991), and word naming (Kawamoto, Liu, Lee, & Grebe, 2014), our first goal is to summarize the similarities between these naming tasks.

Naming Tasks. All naming tasks involving a single word result in that word being uttered. Even though picture or cued naming (where the word itself is not shown) and word naming (where the word itself is shown) ultimately result in similar actions (e.g., Roelofs, 2004), researchers studying these processes have focused on the different aspects involved and have tended to pursue their investigations relatively independently of each other.

Picture and cued naming researchers have focused on how a word's lexeme (conceptual representation) is activated by a picture or a semantically related word that activates the corresponding morpheme, after which its constituent segments are retrieved and spelled out in parallel. By contrast, researchers studying reading aloud have focused on how written symbols map to a sequence of segments. For a language that uses an alphabet, this mapping consists primarily of mapping a letter or a letter combination to a segment or segments.

However, as seen in Figure 1, segmental spellout from non-reading tasks and word-form encoding from reading tasks converge at the level of phonological encoding. The activation of segments prior to phonological encoding also occurs in parallel for both tasks, at least for stimuli that are legal words.[1] Once phonological encoding begins, the final stages of uttering a word follow the same general framework regardless of the specific naming task performed.

Phonological Encoding. After a picture or the spelling of a word is presented and all the segments of that word are activated in parallel as depicted in Figure 1, the segments are phonologically encoded. This process involves mapping segments with their corresponding phonological representations, generating the phonetic plan (Levelt, 1989). Two different assumptions about how these segments are phonologically encoded have been proposed – sequentially from beginning to end (Meyer, 1991; Levelt, et al., 1999a) or in parallel (Dell, 1986; Kawamoto et al., 1998; Liu et al., 2017).

These two possibilities for phonological encoding are shown in the left and right panels of Figure 2 for a word with three segments labeled S1, S2, and S3. The time-course of phonological encoding is shown with time on the horizontal axis. The point labeled '0' is onset

---

[1] Such parallel activation during segmental spellout can be seen in the speech production model proposed by Levelt et al. (1999a). This also holds true for the two major classes of reading models: purely parallel single-route models (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996) as well dual-route models that have a parallel lexical route (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001).

of the full target word, and the point ‘P’ coincides with the presentation of the prime. The time-course of encoding of the target’s segments when none of the segments is phonologically primed is shown in the first set of time-courses (labeled on the left as ‘0’) for the sequential and parallel encoding assumptions. For the sequential encoding assumption, phonological encoding begins with the initial segment being encoded. After the initial segment is encoded, the second segment is encoded. This process continues in a sequential fashion until all the segments have been encoded. By contrast, for the parallel encoding scheme, all the segments are encoded at the same time.

When one or more initial segments are primed, those segments can be encoded even before the specific target has been identified if there is sufficient time to do so (e.g., WEAVER++; Roelofs, 1997). When one segment is primed, it is encoded as soon as it can be. When the first two segments are primed, the second segment is encoded either sequentially after the first segment or in parallel with the first segment depending on the phonological encoding assumption.

Phonetic Encoding. The Phonetic Encoding stage uses the representation from the preceding Phonological Encoding stage to generate the motor plans that are then transmitted to the following Articulatory stage, as seen in Figure 1. These motor plans are based on the minimal phonological information that is transmitted to the Articulatory stage. The motor plan output from the phonetic encoding stage is also referred to as the *minimal unit of articulation.*

The original conceptualization of the minimal unit of articulation was the syllable, established by Levelt and colleagues (Levelt, 1989; Levelt & Wheeldon, 1994; Levelt et al., 1999a), citing Fujimura and Lovins (1978) and Lindblom (1983). For the syllable as the minimal unit of articulation, all the segments of a syllable must be phonologically encoded before the corresponding motor program can be retrieved or assembled and sent to the Articulatory stage (Levelt, et al., 1999a). However, since then, evidence has been found supporting the segment as the minimal unit of articulation (Kawamoto, 1999; Kawamoto et al., 2014). Under this view, the motor plan for segments are retrieved and sent to the Articulatory stage as soon as they become available, regardless of complete phonetic encoding of the syllable.[2]

Articulation. The final stage in uttering a word is Articulation. Of the last stages in speech production, this stage has received the least attention. However, this is changing due in part to consideration of a minimal unit based on the segment and the importance of response criteria. Critically, articulation does not necessarily commence immediately once the minimal unit of articulation has become available – information can be buffered before articulation is initiated based on the response criterion adopted by the speaker. Thus, if the minimal unit of articulation is the syllable, then speakers must adopt a syllable (or larger unit) criterion. However, if the minimal unit of articulation is the segment (as we assume here; Kawamoto et al., 2015), speakers are free to adopt a segment or syllable (or larger unit) criterion.

The notion of a response criterion is closely related to the issue of *planning scope* – how far in advance a response is prepared before speech begins (Cholin, Dell, & Levelt, 2011). As the scope of the upcoming plan determines the time it takes to begin the utterance (Meyer, Roelofs, & Levelt, 2003; Sternberg, Monsell, Knoll, & Wright, 1978), a shorter planning scope leads to shorter response latencies and possibly longer response durations. Such situations can arise due to individual differences (Kawamoto et al., 2014; Lange & Laganaro, 2014; Shriefers & Teruel, 1999), the imposition of response deadlines (Ferreira & Swets, 2002; Kello & Plaut, 2000; Damian, 2003), or when participants are encouraged to begin production before processing has

---

[2] The description here is a specific instantiation of the more complete version described elsewhere (Kawamoto & Liu, 2007; Liu, Kawamoto, Payne, & Dorsey, 2017).

finished (e.g., Fink, Oppenheim, & Goldrick, 2018). The flexibility of planning scope also allows for variability and strategic control of response criteria.

The time-courses of articulation on a given trial in the naming task for the segment and syllable criteria when 0, 1, or 2 segments are primed for sequential and parallel phonological encoding are shown at the bottom of Figure 2. These time-courses of articulation have the same time-scale as the time-courses of phonological encoding above them. For each time-course, the short vertical segment before the response is initiated marks the point in time when the information for a given motor plan has been identified during the phonological encoding stage. After a short delay corresponding to the time required for the motor plan to be retrieved and subsequently transmitted to the articulators, articulation is initiated.

The time-course of articulation of the initial segment corresponds to the three sequentially ordered phases described by Stevens (1998): (1) moving the supraglottal articulators to form the appropriate constriction in the oral tract, (2) moving air through or behind the oral constriction, and (3) releasing the oral constriction as the supraglottal articulators move away from the current position toward the target position of the next segment. The beginning of each of these three phases defines three points along the time-course of articulation of the initial segment corresponding to initiating articulation (the tip of the white triangle), initiating production (the left edge of the first rectangle), and completing production (the right edge of the first rectangle). The length of the first rectangle thus corresponds to the production duration of the initial segment. The production of the second and third segments corresponds to the second and third rectangles. Each of the three rectangles are rendered in the same shade of grey as the corresponding segment depicted in the time-course of phonological encoding.

For the segment criterion, articulation is incremental – articulation is initiated as soon as the initial segment has been phonetically encoded, but its production is completed only when the following segment has been phonetically encoded. As seen in Figure 2, this leads to latency and duration effects as the number of primed segments increases from 0 to 2. When the initial segment has been primed, articulation can be initiated even before the specific target for that trial has been presented if the delay between identification of the prime and target is long enough. If the next segment has not been phonetically encoded by the time production of the first segment would normally have been completed, the first segment is elongated to maintain continuous speech. This process continues until all segments have been produced. Note that the time-course of articulation is identical for sequential and parallel phonological encoding if the time to encode the second primed segment is roughly comparable for both phonological encoding conditions.

For the syllable criterion, articulation is initiated at different times depending on when all the segments have been phonologically encoded. Once initiated, production proceeds without any elongations because the motor program encompasses the entire syllable. For sequential phonological encoding, articulatory latency is shorter as the number of primed segments increases because there are fewer unencoded segments that need to be phonologically encoded, but for parallel phonological encoding, there is no effect of the number of primed segments because phonological encoding of the remaining segments takes the same amount of time.

Verbal Utterance. The output of the Articulation stage is the verbal utterance. It is now recognized that a great deal of variability in the acoustic latency of a response is due to features of the initial segment (Sakuma, Fushimi, & Tatsumi, 1997; Kessler, Treiman, & Mullenix, 2002; Rastle, Croot, Harrington, & Coltheart, 2005). For most segments except for plosives (/p/, /t/, /k/, /b/, /d/, and /g/) and affricates (/tʃ / and /ʤ/), acoustic onset occurs very shortly after the beginning of production. However, for plosives and affricates, the beginning of the production is silent for a substantially longer period of time as intraoral pressure is built up. In fact, for plosives, acoustic energy is only generated near the end of the segment after the following segment has been encoded and intraoral pressure is released resulting in longer acoustic latencies for plosives than non-plosives under identical conditions (Halle, Hughes, & Radley, 1957). Thus, plosive segments must be paired with a subsequent segment to be fluently produced in a word. As an example, consider the words *nun* and *pun*. With knowledge of the first segment alone, the /n/ in *nun* could be acoustically produced on its own while maintaining fluency through elongation; however, the /p/ in *pun* could not as all acoustic energy is released ballistically. While priming an initial plosive allows for articulation – moving the articulators into the proper position and building up the required air pressure – it does not enable acoustic production unless the subsequent segment is also known.

This distinction in the acoustic onset of non-plosives and plosives is indicated by an ‘n’ and ‘p’ above the initial segment on the time-course of articulation in Figure 2, with a light vertical line below the ‘p’ marking the release of intraoral pressure. Thus, acoustic latency conflates production latency and initial segment duration for plosives. As discussed above, plosives cannot be acoustically produced until the subsequent segment has been encoded, which can be clearly seen in the difference between nasals and plosives in the one segment primed condition across segment criterion models where the initial segment is primed but the subsequent is not. Moreover, if acoustic onset is determined using voice-keys, the problem that arises for plosives extends to voiceless fricatives (/f/, /s/, /ʃ /, /θ/, /h/) because voice-keys often miss the low intensity acoustic energy of these segments and is only triggered by a following voiced segment (Rastle & Davis, 2002).

Since non-plosives can be acoustically elongated while plosives cannot, the differences in terms of initiation are manifested in segment duration. For non-plosives, this duration can be determined directly based on acoustic output; however, for plosives, duration must be determined indirectly by comparing the difference in acoustic latency for plosives and non-plosives. More accurately, duration differences of plosive segments due to some factor can be determined by considering the interaction of plosivity and that factor with acoustic latency as the dependent variable (Kawamoto et al., 1998). Non-plosive segments (e.g., /n/ or /s/) can be acoustically produced in isolation and elongated while waiting for subsequent segments to be encoded, leading to an audibly long segment duration. As the acoustic duration of plosive segments (the release of intraoral pressure after the following segment has been encoded) cannot be elongated in the same way, it is approximately the same under different conditions.

Current Study

Given all we have discussed, a simple test of the two models of interest (incremental encoding and incremental articulation) can be devised. A key distinction between the incremental encoding model and the incremental articulation model is how they predict articulation to occur when assuming differing response criteria. As planning scope (and therefore response criteria) can be manipulated, it should be possible to systematically bias participants towards adopting either a segment or syllable-based response criterion. This then allows for a systematic investigation of the predictions made by the two different models and therefore an evaluation of processing schema in general.

In addition to temporal constraints and individual differences affecting planning scope, strategic differences due to the typical instructions used in these experiments might also affect the adoption of a particular response criterion. Standard naming task instructions, like those for pairwise priming tasks, are to "Respond as quickly and as accurately as possible" (e.g., Kawamoto et al., 2014). In such instructions, it is somewhat ambiguous as to what the exact task demands are. First, the beginning of the *response* can be either acoustic onset or articulatory onset, making it necessary to record both acoustic and articulatory latency when determining response latency. Even more importantly, the ambiguity of what the response is also allows for different interpretations of *quickly,* and specifically how it relates to the interpretation of what the *response* is, specifically the initiation or production of speech. If the *response* is the beginning of articulation or acoustic production – the interpretation likely intended by experimenters using latency measurements – then participants should interpret *quickly* as "begin to respond early" (i.e., minimize interval between stimulus presentation and acoustic or articulatory onset). However, the *response* can also be thought of as the complete production of the entire word, not just its beginning. If the *response* is the full pronunciation of the entire word, then participants should interpret *quickly* as "pronounce the word in as short of a time as possible" (i.e., minimize duration of vocal pronunciation).

This results in two strategies in theoretical opposition to each other with respect to response criteria. The "begin to respond early" instruction encourages use of a segment criterion, and will henceforth be referred to as *segment bias*. However, the "pronounce the word in as short a time as possible" instruction encourages use of a word criterion to minimize hesitations and elongations during production. If all the words in the experiment are monosyllabic, then this is also a syllable criterion, and will henceforth be referred to as *syllable bias*. Because the segment criterion makes identical predictions for sequential and phonological encoding and the syllable criterion does not, being able to induce a syllable criterion is particularly important. If ambiguity concerning the task demands is present due to one line in the instructions, it should be possible to systematically bias participants towards a certain response criterion simply by clarifying the instructions one way or the other. While this still leaves open the issue of whether the response is acoustic or articulatory, this issue can be addressed by measuring both acoustic and articulatory latency and examining both when analyzing the data.

We set out to test encoding and response criteria assumptions by comparing various chronometric measures when 0, 1, or 2 initial segments varying in manner are primed for phonological CVC(C) targets in the different instruction bias conditions. Through manipulating participants' adopted response criteria, we can investigate clearly the conditions under which the number of initial segments primed effect arises.

Method

Participants

30 undergraduate students (8 male, 22 female) from the University of California Santa Cruz psychology department participated in the experiment for course credit. All participants were native English speakers and had normal or corrected to normal vision. We chose to run 30 participants based on the effect size found by Kawamoto et al. (2014) and a desired power level of approximately 0.8. This sample size also allowed us to counterbalance the six unique lists. This study was approved by the UCSC Office of Research Compliance Administration.

Materials

60 monosyllabic orthographic CVC(C) English words were used in the experiment, preceded by 16 practice trials comprised of unique fillers and targets. The words were formed such that there were no repeating vowel-coda sequences (e.g., pack and tack), and all words had a short, regular vowel as the second letter of the word. The target words began with an *m*, *n*, *p*, or *t*; nine of each balanced between bilabial, alveolar, nasal, and plosive initial segments. Bilabials

and alveolars were used as they require clear lip movement along one dimension (open vs. closed). Nasals and plosives were approximately matched on number of letters (nasals *M* = 3.53, plosives *M* = 3.72; *t*(33) = -.952, *p* = .348), phonemes (nasals *M* = 3.35, plosives *M* = 3.28, *t*(33) = .466, *p* = .644), and frequency (nasals *M* = 19811, plosives *M* = 26753; *t*(33) = -.404, *p* = .689) according to the Corpus of Contemporary American English (Davies, 2008). The other 24 words were fillers, constructed in the same fashion as the target words, but beginning with a letter other than *m*, *n*, *p*, or *t*. All words were acquired from the MRC psycholinguistic database (Coltheart, 1981). Three versions of the stimulus list were constructed, counterbalancing segments primed and words between participants (e.g., the word *path* was presented with 0 segments primed for one third of the participants, one segment primed for another third, and two segments primed for the remaining third). A full list of the target and filler words can be found in the Appendix.

Prior to the experiment, participants had dots painted directly above, below, and on each side of their lips using Rubie's brand blue paint; the paint was water soluble, non-toxic and easily removed.

Apparatus

The experiment was controlled by a Pentium computer (Gateway 2000, N. Sioux City, SD), paired with a 22" 720 x 480 pixel resolution Dell LCD monitor with a screen refresh rate of 65hz, using the Runword software (Kello & Kawamoto, 1998). The stimulus spanned approximately 1.1 to 2.5 degrees of visual angle from where the participants were seated. All words were presented in a fixed width font.

Video data were recorded using a SONY EXview HAD 480TVL CCD lipstick camera. To eliminate the effects of head movement on the absolute positioning of the lips (necessary for a valid articulatory measure), the lipstick camera was mounted six inches in front of a Wilson adjustable batting helmet, model A5240, on a 1/2-inch-wide L shaped aluminum track bolted to the left side of the helmet. The video and audio capture were controlled by a Dell OptiPlex GX260 desktop computer using Adobe Premiere Pro 1.5 in conjunction with an ADS PYRO A/V Link IEEE1394a high-quality analog to digital video converter. All digitized video footage was captured at a frame rate of 29.97-frames/second with a 720 x 480 resolution and an audio setting of a dual channel stereo at a 48-kHz sampling rate. The verbal responses were recorded using an Ikam lavalier microphone attached next to the camera and routed through a small Xenyx 302USB audio mixer.

Procedure

The experimenter adjusted the camera to ensure that the dots were in the field of view. Participants then read instructions presented on the screen to themselves as the experimenter read them aloud. Half of the participants were instructed to “Pronounce the entire word as quickly as possible” (syllable bias condition) and the other half were instructed to “Begin your response as early as possible” (segment bias condition). Participants were informed that the target words were monosyllabic with a short regular vowel. They were also shown five example words from the practice set to clarify this.

At the onset of each trial, participants were shown a “Ready?” prompt, indicating to press any key when ready. This resulted in participants seeing either no letters of the upcoming word (0 segments primed), the first letter of the upcoming word (1 segment primed), or the first and second letter of the upcoming word (2 segments primed), followed by a fixed number of underscores (for a total of six characters). After a delay, the full word would appear. For the target words, the full word would appear 600 ms after the presentation of the prime. For the filler words, we implemented a variable delay (either 450 ms, 600 ms, or 750 ms) to prevent participants from anticipating the presentation of the full word. The word was then removed 1200 ms after presentation of the full stimulus. We chose to use a pairwise priming paradigm (not block or form-preparation) since this allows for a meaningful measure of articulatory onset as the moment of prime presentation is clearly defined.

Prime onset, stimulus onset, and stimulus offset were delineated through a series of tones produced by the computer and incorporated into one channel of the two-channel audio track. This allowed us to measure latency with respect to stimulus onset using auditory information only. Participants did not hear the tones, as they were recorded directly from the computer.

Participants were given 16 trials as a practice block before engaging the full experiment with 60 trials. Any questions participants asked were answered carefully without giving any additional information that was not stated in the instructions. After the experimental block, the participants were debriefed.

Results

Prior to analysis, all responses were screened for errors. Response errors accounted for 2.5% of the data, with 1.57% coming from the segment bias group and 0.93% coming from the syllable bias group. There was no significant difference in error rates between groups ($t(1078) = 0.974$, $p = .330$).[3]

The remaining data were then used to construct crossed random effects models (CREM) in SPSS as per Carson and Beeson (2013). All models included a random intercept and a fixed slope, and parameters were calculated based on a ML estimation (to allow for comparison of

---

[3] In keeping with the multilevel linear modelling (MLM) approach in the current study, the error analysis is fashioned as a random intercepts model that reflects effects from multiple data hierarchies (i.e., item, subject, etc.) to see if sophisticated MLM analyses are needed. The core rationale is that a non-significant random intercepts model would mean that variance component attribution in subsequent MLM analyses would likewise be non-significant, and hence, unnecessary. The current analysis yields similar results as the traditional by subjects ($t(28) = .856$, $p = .399$) and by items ($t(35) = 1.000$, $p = .324$) analyses.

empty and full models).[4] Models were built for determining effects on acoustic latency, initial segment duration, and articulatory latency. Means and standard deviations for all measures are found in Table 1, and a graphical summary of the results can be seen in Figure 3. A combined representation of the data over time can be found in Figure 4. Although Figure 4 displays the articulatory-acoustic interval (length of the grey bar), this variable was not directly analyzed as doing so is unnecessary for our purposes when analyzing both acoustic and articulatory latency separately.

While many researchers have previously utilized methods to encourage use of a segment criterion (e.g., response deadlines; Damian, 2003; Kello & Plaut, 2000), to our knowledge nobody has yet explicitly attempted to induce a syllable criterion. Given this, we needed to ensure that participants were completing the task in the syllable bias condition in the manner we instructed them to; specifically, we needed to ensure that word durations were consistent amongst priming conditions in this group. For words that ended with plosive segments, duration was marked at the end of the vowel to control for participants varying in their aspiration of such segments. As expected, word durations (in ms) were approximately the same amongst the 0-primed ($M = 289.49$, $SD = 159.44$), 1-primed ($M = 289.47$, $SD = 161.87$), and 2-primed ($M = 292.11$, $SD = 149.92$) conditions in the syllable bias group. There was no significant effect of the number of segments primed on whole word duration for the syllable bias condition ($F(2, 479) = .065$, $p = .937$) as a fixed effect in a CREM including both participants ($Z = 2.503$, $p = .012$) and items ($Z = 3.979$, $p < .001$) as random effects.

---

[4] Note that the potential interplay among different data hierarchies are taken into careful consideration in the crossed random effects models (CREM) used in the current study, whereas they are ignored in traditional analyses when the data are aggregated (i.e., by subjects or by items). In this respect, it can be said that data aggregation, though meant to control variance contamination across hierarchies, has the unintended consequence of artificially suppressing the denominator degrees of freedoms. Thus, the higher denominator degrees of freedom in the CREM analyses are appropriate and expected (see Carson & Beeson, 2013, for a detailed exposition).

Acoustic Latency

All acoustic latency measurements were acquired through digital analysis of the audio recording in Praat (Boersma, 2001) following the procedure outlined by Kawamoto et al. (2014). Onset was estimated automatically and adjusted manually based on the waveform and spectrogram when appropriate by trained research assistants. All research assistants were blind to bias condition and number of segments primed. The data were then examined to eliminate any latency results which were determined *a priori* to be an impossible valid response, specifically any negative acoustic latencies in the 0-primed condition for all initial segments or the 1-primed condition for plosives. This resulted in the loss of one trial from the segment bias group, bringing the total number of analyzed trials to 524 for the segment bias group and 529 for the syllable bias group.

The results of the acoustic latency analysis can provide insight into which encoding schema seen in Figure 2 is more likely being used. While both schemata make similar predictions in the segment bias condition, the sequential encoding schema predicts a significant main effect of the number of segments primed in both the syllable and segment bias conditions while the parallel encoding schema predicts no significant main effect of the number of segments primed in the syllable bias condition but a significant main effect in the segment bias condition.

A model that included the random effects of participant and item only was tested first, to determine the necessity of inclusion in the final model. The random effect of participants was significant ($Z = 3.737$, $p < .001$), and the random effect of items was not significant ($Z = .334$, $p = .738$). As such, the random effect of items was not included in the following models.

Fixed effects (instructions, number of segments primed, and manner) were next added in to the model. The main effect of the instructions was not significant ($F(1, 30) = 1.776$, $p = .193$), the main effect of the number of segments primed was significant (*F*(2, 1023) = 41.564, $p <$ .001), and the main effect of manner was significant (*F*(1, 1023) = 6.138, $p = .013$). The interaction between the instructions and the number of segments primed was significant (*F*(2, 1023) = 19.086, $p < .001$), the interaction between the instructions and manner was significant (*F*(1, 1023) = 8.780, $p = .003$), and the interaction between the number of segments primed and manner was not significant (*F*(2, 1023) = 1.273, $p = .280$). The three-way interaction between the instructions, number of segments primed, and manner was significant (*F*(2, 1023) = 3.543, $p = .029$). For the final model, non-significant effects were removed, as can be seen in Table 2.

To unpack the three-way interaction between instructions, manner, and number of segments primed, two tests of simple main-effects were carried out for the segment and syllable bias conditions, respectively. For the segment bias condition, the test of simple main effects showed significant main effects of the number of segments primed (*F*(2, 509) = 36.692, $p <$ .001) and manner (*F*(2, 509) = 9.226, $p =.003$), and a non-significant interaction between the number of segments primed and manner (*F*(2, 509) = 2.279, $p = .103$). Post hoc pairwise linear contrasts showed that acoustic latencies in the 0-primed condition were on average significantly slower than those of the 1-primed condition (*t*(509) = 8.09, $p < .001$), and those of the 2-primed condition (*t*(509) = 8.41, $p < .001$). Furthermore, the acoustic latencies of the 1-primed conditions were significantly slower than those of the 2-primed condition (*t*(509) = 2.803, $p =$ .005). To explore the difference between 1 and 2 segments primed in the segment bias condition, a post hoc test was conducted using the Bonferroni correction adjustment in SPSS mixed. This overall effect interacted with manner such that a significant difference was found among plosives (*M* = 49.154, *SE* = 13.719, $p = .001$) but not nasals (*M* = 19.282, *SE* = 13.597, $p = .469$).

For the syllable bias condition, the test of simple main effects showed a significant main effect of the number of segment primed (*F*(2,514) = 5.414, *p* = .005), a non-significant main effect of manner (*F*(1,514) = .291, *p* = .59), and a marginally significant interaction between the number of segment primed and manner (*F*(2, 514) = 2.813, *p* = .061). Post hoc pairwise linear contrasts showed that while the acoustic latencies of the 0-primed condition were significantly slower than the 1-primed and 2-primed conditions (*t*(514) = 2.581, *p* = .010, and *t*(514) = 3.065, *p* = .002, respectively), the contrast between the 1-primed and 2-primed were not significant (*t*(514) = .499, *p* = .618).

The final model included the random effect of participants and the fixed effects of number of segments primed, manner, the interaction between the instructions and the number of segments primed, the interaction between the instructions and manner, and the three-way interaction. This model was then compared with the empty model to give a measure of effect size, showing that the final model accounted for 50.9% of the variance above the empty model. The summary of the final model can be found in Table 2. As seen in Figures 3 and 4, the results of the acoustic latency analysis agree with the predicted results from the parallel encoding schema predictions in Figure 2. Specifically, no main effect of priming was found in the syllable bias condition while a main effect of priming was found in the segment bias condition, and the latency difference found between 1 and 2 segments primed in the segment bias condition was contingent on the manner of the initial segment.

Initial Segment Duration

All initial segment duration measurements were acquired through digital analysis of the audio recording in Praat (Boersma, 2001) following the procedure outlined by Kawamoto et al. (2014). Initial segment duration was defined and measured as the onset of acoustic energy to the onset of the subsequent vowel. Vowel onset was estimated automatically and adjusted manually based on the waveform and spectrogram when appropriate by trained research assistants. All research assistants were blind to bias condition and number of segments primed.

The results of the initial segment duration analysis can confirm that the segment bias group is indeed using a segment criterion and incremental articulation. If this is the case, then the segment bias group should have longer initial segment durations when at least the initial segment is primed amongst nasals (which can be acoustically elongated), and the syllable bias group should show no effect of priming on initial segment duration, as seen in Figure 2.

A model that included the random effects of participant and word only was tested first, to determine the necessity of inclusion in the final model. The random effect of participants was significant (Z = 3.489, $p < .001$) as well as the random effect of items (Z = 2.029, $p = .042$). As such, both random effects were included in the following model.

Fixed effects (instructions, number of segments primed, and manner) were then added in to the model. Adding in the fixed effects caused the random effect of items to no longer be significant (Z = 1.372, $p = .170$), and thus it was excluded from the final model. The main effect of the instructions was significant ($F(1, 30) = 9.390$, $p = .005$), the main effect of the number of segments primed was significant ($F(2, 1023) = 10.030$, $p < .001$), and the main effect of manner was significant ($F(1, 1023) = 22.600$, $p < .001$). The interaction between the instructions and the number of segments primed was significant ($F(2, 1023) = 10.256$, $p < .001$), the interaction between the instructions and manner was significant ($F(2, 1023) = 36.510$, $p < .001$), and the interaction between the number of segments primed and manner was significant ($F(2, 1023) = 3.760$, $p = .024$). The three-way interaction between the instructions, number of segments primed, and manner was significant ($F(2, 1023) = 3.573$, $p = .028$).

To unpack the three-way interaction between bias condition, manner, and number of segments primed, two tests of simple main-effects were again performed for the segment and syllable bias conditions, respectively. For the segment bias condition, the test of simple main effect showed significant main effects of the number of segment primed ($F(2, 509) = 11.29$, $p < .001$) and manner ($F(1, 509) = 32.259$, $p < .001$), as well as a significant interaction between the number of segments primed and manner ($F(2, 509) = 4.081$, $p = .017$). Post hoc linear contrasts showed that the initial segment durations in the 0-primed conditions were significantly shorter than those of the 1 and 2-primed conditions ($t(509) = 4.316$, $p < .001$, and $t(509) = 3.864$, $p < .001$, respectively), whereas the initial segment durations of the 1 and 2-primed conditions did not differ significantly from each other ($t(509) = .456$, $p = .649$). In the syllable bias condition, the test of simple main effects showed only a significant effect of manner ($F(1, 514) = 3.975$, $p = .047$), and no significant effect of either the number of segment primed ($F(2, 514) = .045$, $p = .956$) or the interaction between the number of segment primed and manner ($F(2, 514) = .043$, $p = .958$). Thus, no further post hoc analyses were necessary for the syllable bias condition.

The final model thus included the random effects of participants and the fixed effects of number of segments primed, manner, the interaction between the instructions and the number of segments primed, the interaction between the instructions and manner, the interaction between the number of segments primed and manner, and the three-way interaction. This model was then compared with the empty model to give a measure of effect size, showing that the final model accounted for 28.0% of the variance above the empty model. The summary of the final model can be found in Table 2. As seen in Figures 3 and 4, the results of the initial segment duration analysis confirm that the segment bias group was using a segment criterion and incrementally articulating the word, while the syllable bias group was using a syllable criterion (as also evidenced by the lack of word duration differences across priming conditions in the syllable bias group).

Articulatory Latency

Articulatory motion was determined through motion tracking the blue dots painted above and below participants' lips using the "Track Motion" feature in Adobe After Effects. This data (a series of x-y coordinates of each point for each frame of video data) was then incorporated into the audio track in Praat (Boersma, 2001) to allow for comparison to stimulus onset, providing a measurement of articulatory latency.

The data included in the articulatory latency analysis were determined through the relation between the place of the initial segment and the initial oral configuration of the participant. Bilabials (*p, m*) were included only if the participant had an open mouth position pre-stimulus presentation, and alveolars (*t, n*) were included only if the participant had a closed mouth position pre-stimulus presentation. This was done to get the most accurate measure of lip movement initiation across conditions (Kawamoto, Liu, Mura, & Sanchez, 2008). This resulted in a total of 531 trials (50.4% of the non-error set) being included. Initial lip position (open vs. closed) was determined through use of a reference frame where the participants' mouth was very slightly open – any separation less than this was deemed closed, and any separation more than or equal to this was deemed open. Articulatory latency was defined as the first change of at least 1.2 pixels[5] towards the correct target response, verified through visual inspection of the motion trajectory graph. This movement towards the target response is different for bilabials (lips begin to close from an open position) and alveolars (lips begin to open from a closed position), and these different gestures towards the different targets can be seen in Figure 5.

---

[5] This corresponds to a movement of about 0.25 to 0.5 mm, depending on the size of participants' lips, the precise size and placement of the dots, and the distance from the camera to the mouth which varied slightly based on the size of participants' heads.

The data were then examined to eliminate any latency results which were determined *a priori* to be an impossible valid response, specifically any negative articulatory latencies in the 0-primed condition for all initial segments. This resulted in the loss of two trials, one from the segment bias group and one from the syllable bias group. The final total number of analyzed trials was 255 for the segment bias group and 276 for the syllable bias group.

The results of the articulatory latency analysis can determine what participants are considering the response (segment bias) or pronunciation of the word (syllable bias) to be – acoustic or articulatory. If the response or pronunciation is determined to mean articulation, including silent articulation before acoustic production, then the pattern of results should be the same as in the acoustic latency analyses, as shown in Figure 2. This would then mean finding a significant effect of the number of segments primed in the segment bias condition but not the syllable bias condition. If, however, the response or pronunciation is determined to mean only acoustic production and not articulation, then it is possible that there will be a significant effect of priming on articulatory latency in both the segment bias condition and the syllable bias condition. This could occur as participants in the syllable bias condition would be free to use a segment criterion for the initiation of articulation prior to acoustic production as it is irrelevant to the perceived instructions and task demands.

A model that included the random effects of participant and item only was tested first, to determine the necessity of inclusion in the final model. The random effect of participants was significant (Z = 3.257, *p* = .001); however, including the random effect of items caused the model to fail to converge. As such, the random effect of items was not included in the following models.

Fixed effects (instructions, number of segments primed, and manner) were next added in to the model. The main effect of the instructions was not significant (*F*(1, 30) = 3.990, *p* = .055), the main effect of the number of segments primed was significant (*F*(2, 503) = 44.570, *p* < .001), and the main effect of manner was not significant (*F*(1, 503) = .005, *p* = .943). The interaction between the instructions and the number of segments primed was significant (*F*(2, 503) = 12.885, *p* < .001), the interaction between the instructions and manner was not significant (*F*(1, 503) = .653, *p* = .419), and the interaction between the number of segments primed and manner was not significant (*F*(2, 503) = .008, *p* = .992). The three-way interaction between the instructions, number of segments primed, and manner was not significant (*F*(2, 503) = .088, *p* = .916). For the final model, non-significant effects were removed, as can be seen in Table 2.

To unpack the interaction between the instructions and number of segments primed as well as investigate the effect of priming two segments over one, post hoc tests were conducted using the Bonferroni correction adjustment in SPSS mixed. A significant effect of the number of segments primed was found in both the segment bias condition (*F*(2, 503) = 49.990, *p* < .001) and the syllable bias condition (*F*(2, 503) = 5.632, *p* = .004). Although both conditions showed this overall main effect, they significantly differed in the magnitude of the priming benefit when at least the initial segment was primed – no significant difference was found between the segment bias and syllable bias conditions when 0 segments were primed (*M* = 29.966, *SE* = 40.160, *p* = .459), however a significant difference was found between the groups when 1 segment was primed (*M* = 127.612, *SE* = 39.428, *p* = .002) and 2 segments were primed (*M* = 109.357, *SE* = 39.803, *p* = .008). Lastly, there were no significant differences between 1 and 2 segments primed in both the segment bias (*M* = 11.087, *SE* = 23.848, *p* = 1.000) and syllable bias (*M* = 29.342, *SE* = 23.061, *p* = .611) conditions.

The final model included the random effect of participants and the fixed effects of number of segments primed and the interaction between the instructions and the number of segments primed. This model was then compared with the empty model to give a measure of effect size, showing that the final model accounted for 37.3% of the variance above the empty model. The summary of the final model can be found in Table 2. Since a significant effect of the number of segments primed was seen in both conditions as seen in Figures 3 and 4, participants likely interpreted the response or pronunciation of the word to be acoustic in nature, and not articulatory.

Additional Analyses

When a word is incrementally articulated, acoustic latency should be short and initial segment duration should be large with a strong negative correlation between the two variables. This indicates incremental articulation as the word is likely being articulated based on the prime before the remainder is encoded (short acoustic latency), leading to segment elongations (long initial segment duration). Separate correlations were conducted for the syllable bias and segment bias groups using every trial as an independent data point. The syllable bias group was found to have a small positive correlation, $r(527) = .173, p < .001$, while the segment bias group was found to have a much larger negative correlation, $r(522) = -.535, p < .001$.

Negative acoustic and articulatory latencies are a clear indicator that articulation is beginning before the presentation of the entire word. In the syllable bias group, there were no negative acoustic latencies. However, in the segment bias group, there were 18 valid responses given with negative acoustic latency. All valid trials with negative acoustic latency had either the initial segment primed (nasals only) or two segments primed (nasals and plosives). Negative articulatory latencies were even more common, with 15 occurrences in the syllable bias group and 53 occurrences in the segment bias group. Example trials that include negative acoustic and articulatory latencies can be seen in Figure 5.

Discussion

The results of the current study (Figure 4) lined up closely with predictions made by parallel phonological encoding accounts (Figure 2). Overall, three key results were found by the present study:

1. The pattern of results supports a parallel over sequential phonological encoding schema. This is primarily supported by the lack of a number of segments primed effect on acoustic latency in the syllable bias condition.
2. Participants can be biased towards adopting either a segment or syllable based response criterion based on the task instructions. Support for this comes from the duration effects seen in the segment bias group and the lack of any duration effects in the syllable bias group.
3. The number of segments primed effect can be accounted for by the incremental articulation model as the acoustic latency effect is due in part to the plosivity of the initial segment.

While these results are central to our main hypotheses, additional results from both conditions help to further explain the exact time-course of processing and responding occurring in both groups.

Syllable Bias

The results of the syllable bias group are clear. First, no significant effect of priming was observed in the acoustic latencies of nasals, although an effect was found for plosives. Second, as the number of initial segments primed increased, articulatory latencies decreased for both plosives and nasals. Third, there was no effect of segment priming on the initial segment duration of nasals and plosives, nor the duration of the whole word.

The articulatory latency results show that speakers interpret *response* as the acoustic response, rather than the articulatory response. To this end, the decrease in articulatory latency as the number of initial segments primed increases show that participants were actively preparing for the acoustic response, which includes moving the lips to the correct place and building up pressure for the plosives when either 1 or 2 segments were primed (as described by Kawamoto et al., 2008 in the delayed naming task). The ability for participants to build up pressure for the plosives based on the prime explains why an effect of priming was found for plosives on acoustic latency. Nonetheless, acoustic latency and initial segment duration show that participants did not initiate the acoustic response until the complete pronunciation became available.

Most importantly, lining up the results of the syllable bias group seen in Figure 4 with the predictions made in Figure 2 supports encoding occurring in parallel. The strongest indication can be seen in the almost identical acoustic latencies of the words with nasal initial segments, as the significant effect seen amongst plosives can be explained through articulatory processes alone.

Segment Bias

In contrast to the syllable bias group, a significant number of initial segments primed effect was observed in the acoustic latencies of both nasals and plosives, including a significant difference between the 1 and 2 segment(s) primed conditions driven by plosives. As discussed in the introduction, acoustically producing plosives fluently requires the availability of the next segment, whereas acoustically producing nasals does not. This difference between plosives and nasals can be seen most clearly in Figure 3.

Moreover, the initial segment duration was found to be affected by the number of initial segments primed with longer initial segment durations when the initial segment was known. Whereas initial segment duration did not vary for plosives (because they cannot be acoustically elongated), nasals showed significantly longer initial segment duration when at least the initial segment was primed, as had been previously reported (Kawamoto et al., 2014). However, no difference in initial segment duration was found between the 1 and 2 segment(s) primed condition. This lack of a difference is contrary to predictions made for the segment bias condition in Figure 2, which predicted initial segment duration to decrease when 2 segments were primed as speakers could begin articulating the second segment and avoid elongating the first. There are two possible reasons why this might not have occurred. First, participants were only instructed to "begin early", so incremental articulation was only strategically beneficial when producing the initial segment – there was no strategic benefit to elongating subsequent segments once the initial was produced. Another possibility is that participants found it easier and more natural to elongate the initial segment over the following vowel, which is louder, requiring more energy to produce and elongate. These two possibilities are not mutually exclusive, and it is likely that both factors contributed to the lack of this effect.

Finally, the significant number of initial segments primed effect observed in articulatory latencies was driven purely by the difference between the 0 and 1 segment primed conditions, with no difference observed between the 1 and 2 segment(s) primed conditions. As predicted, participants began the articulatory response based on the presence of the initial segment alone. Importantly, just as with the syllable bias group there was no effect of manner seen on articulatory latency.

In contrast to the syllable bias group, the presence of negative acoustic latencies in the segment bias group (Figure 5) indicates (and necessitates) incremental articulation and the use of a segment criterion. Trials with negative articulatory latencies were much more common than trials with negative acoustic latencies.

Accounting for the Number of Initial Segments Primed Effect

Levelt et al. (1999b) state that the incremental articulation model (Kawamoto et al., 1999) is unable to fully explain segment priming effects due to the number of initial segments primed effect (e.g., Meyer, 1991; Roelofs, 2002). This makes sense if the locus of the number of initial segments primed effect is phonological encoding as more segments primed means fewer segments remain to be encoded upon presentation of the entire word. Specifically, "according to the articulatory account, the speaker can initiate the utterance as soon as its first segment has been selected. Hence, the benefits of one or several shared word-initial segments should be equal" (Levelt et al., 1999b, p. 66).

By demonstrating parallel phonological encoding in the syllable bias group and showing that the number of initial segments primed and initial segment duration effects were exclusive to the segment bias group and driven by the manner of the initial segment, the current study showed that the locus of the number of initial segments primed effect is in articulation rather than encoding. When looking at articulatory latency, the benefits of priming 1 or 2 word-initial segments were in fact equal, as proposed by Levelt et al. (1999b, p. 66). Our results thus offer strong support for the incremental articulation model rather than the incremental encoding model.

At the same time, we should note that although we used the segment and syllable criteria as the most straightforward means of addressing the locus of the number of initial segments primed effect, it does not mean that those are the only criteria available. In fact, Liu and colleagues' (Liu et al., 2017) description of the expanded incremental articulation model explicitly argues for a response continuum between the segment, the syllable, or larger units. Accordingly, further inquiries are necessary to understand precisely how different speakers might adopt varying strategies, planning scopes, and response criteria to deal with the specific task demands of different situations.

Strategic Control over Response Initiation

Though the current study focuses squarely on response strategies that modulate the planning scope to the segment or word, it is important to note that planning scope modulation is by no means the only form of strategic control over response initiation. Another way is *cascaded processing* (Kello, Plaut, & MacWhinney, 2000; Kello, 2004; Goldrick & Blumstein, 2006; Smolensky, Goldrick, & Mathis, 2014). Cascaded processing allows partially activated representations from a preceding stage to be used in the next stage before processing is complete. Strategic control can thus be modeled through changing the rate of cascaded processing and the activation threshold parameters for when responses are initiated in the model put forth by Kello et al. (2000, Figure 7), for instance. Alternatively, multiple semantic, phonological, or phonetic candidates can be differentially activated in parallel (Goldrick & Blumstein, 2006; Smolensky et al., 2014, Fink et al., 2018), and strategic control could be modeled as altering the threshold of candidate activation required before a response is initiated. Though a detailed discussion of cascaded processing is outside the scope of the current study, its brief reference here is meant to highlight the fact that there may be different ways to model how a speaker may impart strategic control over the initiation and time course of a response. Accordingly, the precise juncture to which incremental articulation, cascaded processing, and other forms of strategic control may intersect warrants further investigation.

Final Remarks

Much of the empirical data collected in speech production and reading aloud research have been interpreted under overly simplistic views of what occurs during articulation. As researchers increase their consideration of production processes (Kawamoto et al., 1998; Kello et al., 2000; Goldrick, 2006), closer attention is needed to account for differential response criteria based on differential task demands. This shift in the explanatory focus from phonological encoding to articulation and strategic control has allowed us to provide an alternative account of the well-known number of initial segments primed effect in speech production. Accounting for differing response criteria helps to explain why earlier results appear to support the syllable as minimal unit of articulation (Meyer, 1991) and others support the segment (Kawamoto et al., 2014), providing some resolution to this ongoing debate. Rather than attributing systematic variance to vague individual differences – especially since the ambiguity of many naming instructions leave open to interpretation as how participants may or should respond – researchers can manipulate participants' strategies and response criteria to explain such variability in speech production.

References

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341-345. Retrieved from: http://www.fon.hum.uva.nl/paul/praat.html

Carson, R. J., & Beeson, C. M. L. (2013). Crossing language barriers: Using crossed random effects modelling in psycholinguistics research. *Tutorials in Quantitative Methods for Psychology, 9*, 25-41. doi:10.20982/tqmp.09.1.p025

Cholin, J., Dell, G. S., & Levelt, W. J. M. (2011). Planning and articulation in incremental word production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 109-122. doi:10.1037/a0021322

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A,* 497-505. doi:10.1080/14640748108400805

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual-route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256. doi:10.1037/0033-295X.108.1.204

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 416-431. doi: 10.1037/0278-7393.29.3.416

Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA): 520 million words, 1990-present.* Available online at https://corpus.byu.edu/coca/.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283. doi:10.1037/0033-295X.93.3.283

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language, 46*, 57-84. doi:10.1006/jmla.2001.2797

Fink, A., Oppenheim, G. M., & Goldrick, M. (2018). Interactions between lexical access and articulation. *Language, Cognition, and Neuroscience, 33*(1), 12-24. doi:10.1080/23273798.2017.1348529

Fujimura, O., & Lovins, J. B. (1978). Syllables as concatenative phonetic units. *In A. Bell & J. B. Hooper (Eds.), Syllables and segments. Amsterdam: North-Holland.*

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*, *21*(7-8), 817-855. doi:10.1080/01690960600824112

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649-683. doi:10.1080/01690960500181332

Halle, M., Hughes, G. W., & Radley, J.-P. A. (1957). Acoustic properties of stop consonants. *The Journal of the Acoustical Society of America, 29*(1), 107-116. doi:10.1121/1.1908634

Kawamoto, A. H. (1999). Incremental encoding and incremental articulation in speech production: Evidence based on response latency and initial segment duration. *Behavioral and Brain Sciences, 22*, 48-49. doi:10.1017/S0140525X99331770

Kawamoto, A. H., Kello, C. T., Jones, R., & Bame, K. (1998). Initial phoneme versus whole-word criteria to initiate production: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 862-885. doi:10.1037/0278-7393.24.4.862

Kawamoto, A. H., Liu, Q., & Kello, C. T. (2015). The segment as the minimal planning unit in speech production and reading aloud: Evidence and implications. *Frontiers in Psychology, 6,* 1-6. doi:10.3389/fpsyg.2015.01457

Kawamoto, A. H., Liu, Q., Lee, R. J., & Grebe, P. R. (2014). The segment as the minimal planning unit in speech production: Evidence based on absolute response latencies. *The Quarterly Journal of Experimental Psychology, 67*, 2340-2359. doi:10.1080/17470218.2014.927892

Kawamoto, A. H., Liu, Q., Mura, K., & Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language, 58*, 347-365. doi:10.1016/j.jml.2007.06.002

Kello, C. T. (2004). Control over the time course of cognition in the tempo-naming task. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 942-955. doi:10.1037/0096-1523.30.5.942

Kello, C. T., & Kawamoto, A. H. (1998). Runword: An IBM-PC software package for the collection and acoustic analysis of speeded naming response. *Behavior Research Methods, Instruments and Computers, 30*, 371-383. doi:10.3758/BF03200668

Kello, C. T., & Plaut, D. C. (2000). Strategic control in word reading: Evidence from speeded responding in the tempo-naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 719-250. doi:10.1016/S0749-596X(02)00512-0

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task-dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech production. *Journal of Experimental Psychology: General, 129*(3), 340-360. doi:10.1037/0096-3445.129.3.340

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language, 47*, 145-171. doi:10.1006/jmla.2001.2835

Lange, V. M., & Laganaro, M. (2014). Inter-subject variability modulates phonological advance planning in the production of adjective-noun phrases. *Frontiers in Psychology, 5,* 1-12. doi:10.3389/fpsyg.2014.00043

Levelt, W. J.M. (1989). Speaking. From intention to articulation. *MA: The MIT Press, Cambridge*.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999a). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-38. doi:10.1017/S0140525X99001776

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999b). Multiple perspectives on word production. *Behavioral and Brain Sciences, 22*, 61-69. doi:10.1017/S0140525X99451775

Levelt, W. J., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary?. *Cognition*, *50*(1), 239-269.

Lindblom, B. (1983) Economy of speech gestures. In P. F. MacNeilage (Eds.) *The Production of Speech.* Springer. doi:10.1007/978-1-4613-8202-7_10

Liu, Q., Kawamoto, A. H., Payne, K. K., & Dorsey, G. N. (2017). Anticipatory coarticulation and the minimal planning unit of speech. *Journal of Experimental Psychology: Human Perception and Performance.* Advance online publication. doi:10.1037/xhp0000443

Meyer, A. S. (1991). The time-course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30*, 69-89. doi:10.1016/0749-596X(91)90011-8

Meyer, A. S., Roelofs, A., & Levelt, W. J. M. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language, 48*, 131-147. doi:10.1016/S0749-596X(02)00509-0

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56-115. doi:10.1037/0033-295X.103.1.56

Rastle, K., Croot, K. P., Harrington, J. M., & Coltheart, M. (2005). Characterizing the motor execution stage of speech production: Consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance, 31*(5), 1083-1095. doi:10.1037/0096-1523.31.5.1083

Rastle, K., & Davis, M. H. (2002). On the complexities of measuring naming. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 307-314. doi:10.1037/0096-1523.28.2.307

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*, 249-284. doi:10.1016/S0010-0277(97)00027-9

Roelofs, A. (2002). Spoken language planning and the initiation of articulation. *The Quarterly Journal of Experimental Psychology, 55A*, 465-483. doi:10.1080/02724980143000488

Roelofs, A. (2004). Seriality of phonological encoding in naming objects and reading their names. *Memory & Cognition 32*, 212-222. doi:10.3758/BF03196853

Sakuma, N., Fushimi, T., & Tatsumi, I. (1997). Measurement of naming latency of kana characters and words based on speech analysis: Manner of articulation of a word-initial phoneme considerably affects naming latency. *Japanese Journal of Neuropsychology, 13*, 126-136.

Schriefers, H., & Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology, 11*(1), 17-50. doi:10.1080/713752301

Smolensky, P., Goldrick, M., & Mathis, D. (2014). Optimization and quantization in gradient symbol systems: a framework for integrating the continuous and the discrete in cognition. *Cognitive Science*, *38*(6), 1102-1138. doi: 10.1111/cogs.12047

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 118-152). New York: Academic Press.

Stevens, K. N. (1998). Acoustic Phonetics. *Cambridge: MIT*.

Table 1

Means and standard deviations in milliseconds (ms)

	Segments Primed		Acoustic Latency *M (SD)*	Initial Segment Duration *M (SD)*	Articulatory Latency *M (SD)*
Syllable Bias					
	0				
		Nasal	446 (78)	58 (28)	308 (62)
		Plosive	461 (94)	61 (16)	322 (51)
	1				
		Nasal	439 (96)	57 (23)	262 (138)
		Plosive	434 (93)	61 (16)	269 (151)
	2				
		Nasal	439 (114)	58 (29)	224 (206)
		Plosive	423 (89)	61 (16)	247 (173)
Segment Bias					
	0				
		Nasal	455 (81)	70 (32)	344 (77)
		Plosive	459 (80)	62 (17)	343 (77)
	1				
		Nasal	360 (168)	111 (105)	131 (269)
		Plosive	416 (132)	72 (26)	129 (234)
	2				
		Nasal	342 (189)	107 (90)	133 (253)
		Plosive	370 (185)	70 (31)	113 (266)

Table 2

Final models for acoustic latency, initial segment duration, and articulatory latency

	Acoustic Latency	Initial Segment Duration	Articulatory Latency
Random Effects			
Residual	**8083.96 (357.44)**	**1706.27 (75.44)**	**24651.95 (1557.36)**
Participants	**6882.45 (1836.79)**	**348.28 (102.40)**	**7635.67 (2332.98)**
Items			
Fixed Effects			
Intercept	**425.26 (23.47)**	**60.94 (6.58)**	**234.81 (28.13)**
Bias = Seg		9.32 (9.24)	
NP = 0	**33.39 (13.60)**	.27 (6.25)	**78.92 (23.79)**
NP = 1	8.43 (13.48)	-.07 (6.19)	29.34 (23.06)
NP = 2			
Manner	15.89 (13.56)	-2.60 (6.23)	
Bias = Seg * NP = 0	.28 (33.19)	-9.03 (8.82)	29.97 (40.16)
Bias = Seg * NP = 1	-16.70 (33.21)	1.77 (8.84)	**-127.61 (39.43)**
Bias = Seg * NP = 2	-57.43 (33.20)		**-109.36 (39.80)**
Bias = Seg * Manner	**-43.37 (19.23)**	**38.82 (8.83)**	
NP = 0 * Manner		-.96 (8.84)	
NP = 1 * Manner		-1.07 (8.77)	
NP = 2 * Manner			
Bias = Seg * NP = 0 * Manner	21.84 (19.20)	**-27.22 (12.48)**	
Bias = Seg * NP = 1 * Manner	-29.87 (19.32)	3.08 (12.48)	
Bias = Seg * NP = 2 * Manner			
Bias = Syl * NP = 0 * Manner	-28.91 (19.23)		
Bias = Syl * NP = 1 * Manner	-10.68 (19.09)		
Bias = Syl * NP = 2 * Manner			
Model Summary			
Deviance Statistic (-2LL)	12565.69	10887.76	6932.58
# of parameters	14	14	8

Note. Parameter estimates are based off the dependent variables in milliseconds (ms). NP stands for number primed. All estimates for manner are based off nasals. Standard errors for estimates in parentheses. Omitted parameters are left blank, and significant ($p < .05$) parameters are bolded.

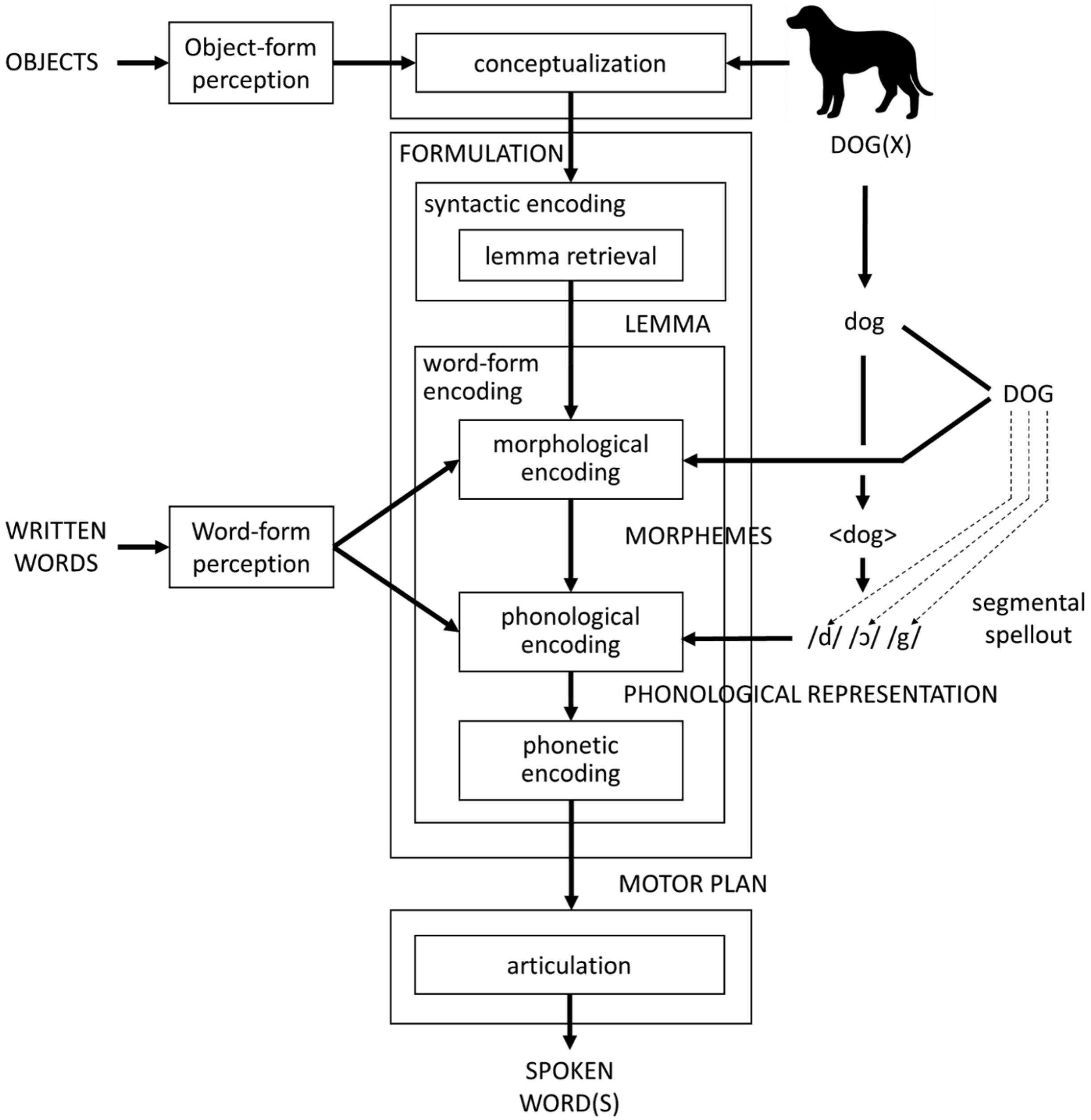

Figure 1. Illustration of the general speech production framework based on Levelt, Roelofs, & Meyer (1999a) and Roelofs (2004) including how different naming tasks are situated in this process.

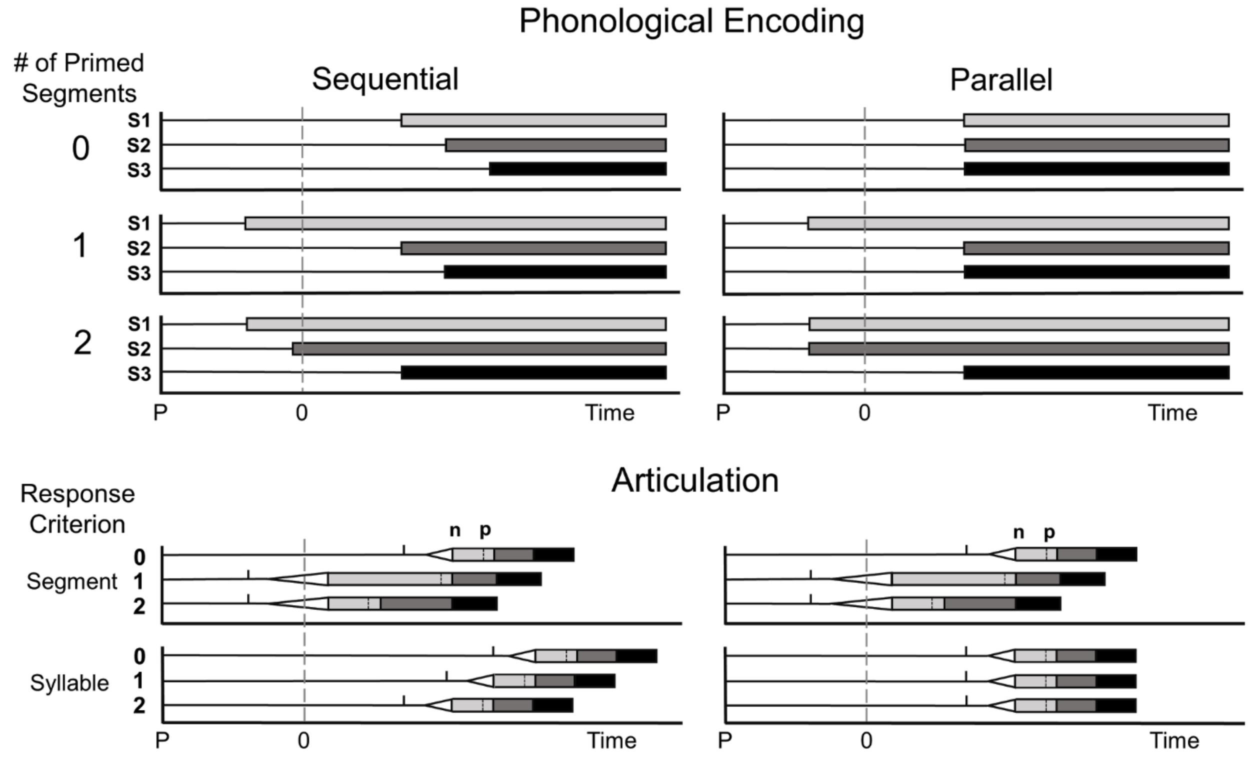

Figure 2. Predicted time-course of results from both sequential and parallel encoding accounts when adopting a segment or syllable response criterion in the present experiment. S1-S3 represent the phonological segments comprising the target word (here a CVC structure). P represents the onset of the prime (0, 1, or 2 segments), and 0 represents the onset of the full target word, 600 ms after the prime. In the Articulation process, the time-course of Articulation when 0, 1, or 2 segments are primed are shown with the same timing as the Phonological Encoding process above. The left edge of the white triangle represents articulatory onset, the left edge of the light grey box (S1) represents acoustic onset for nasals (solid line; *n*) and plosives (dotted line; *p*), and the filled bars represent the duration of the corresponding segments seen above.

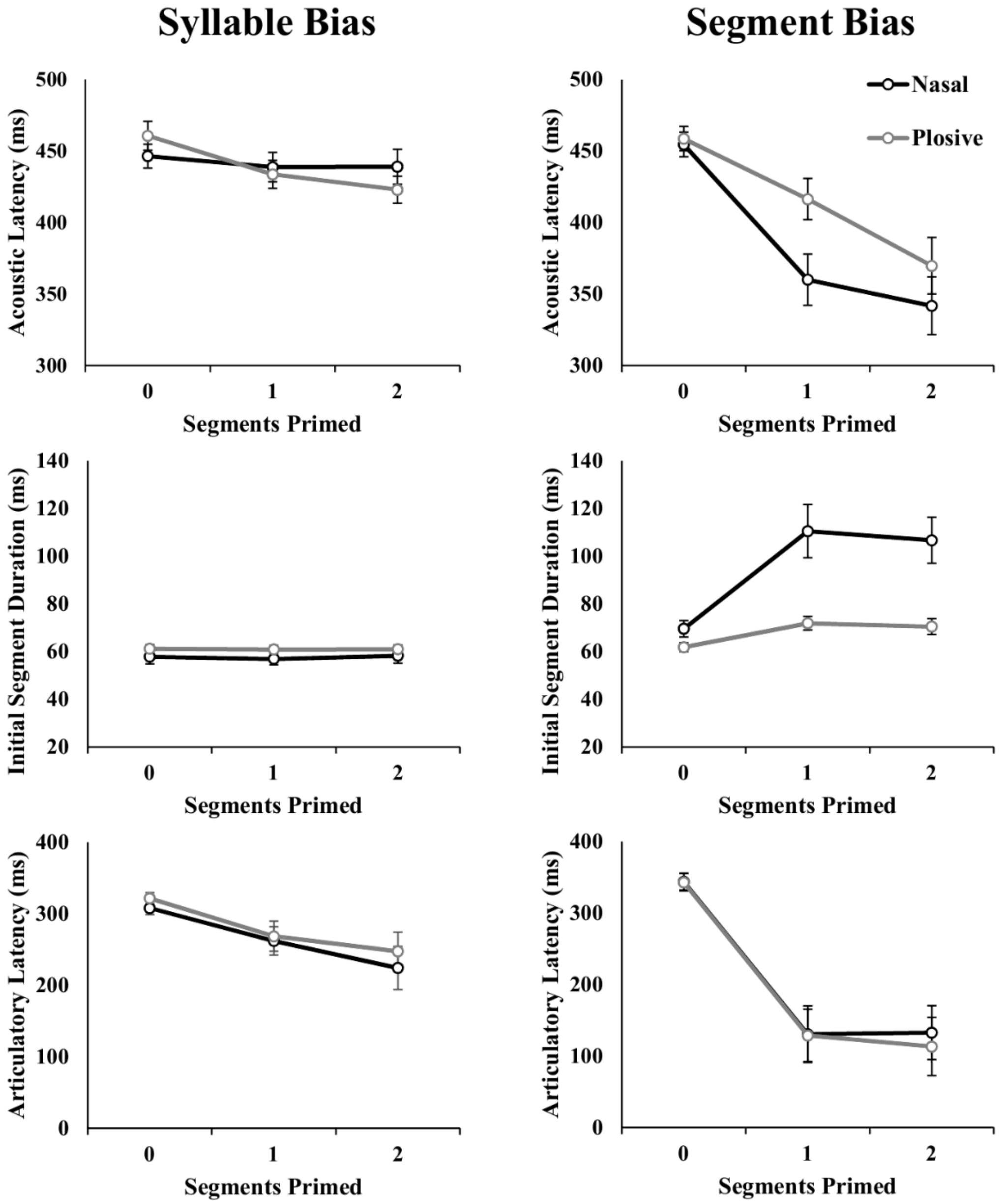

Figure 3. Acoustic latency, initial segment duration, and articulatory latency for both the syllable bias and segment bias groups across all priming and manner conditions. Error bars represent one standard error of the mean.

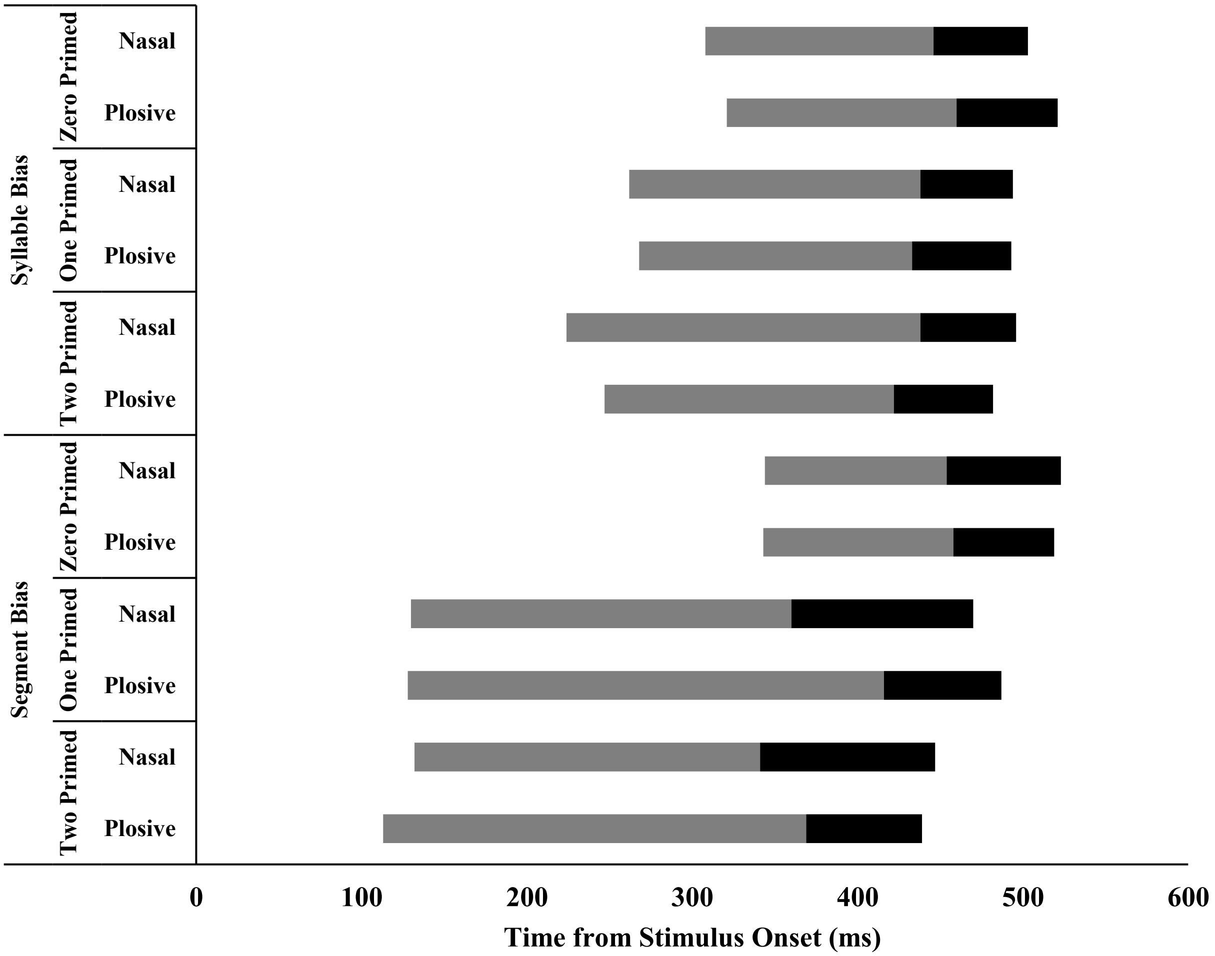

Figure 4. The time-course of articulation for the various conditions. The left edge of the grey bar marks articulatory latency, the left edge of the black bar marks acoustic latency, and the right edge of the black bar marks vowel onset. The length of the grey bar represents movement and wait time, and the length of the black bar denotes initial segment duration.

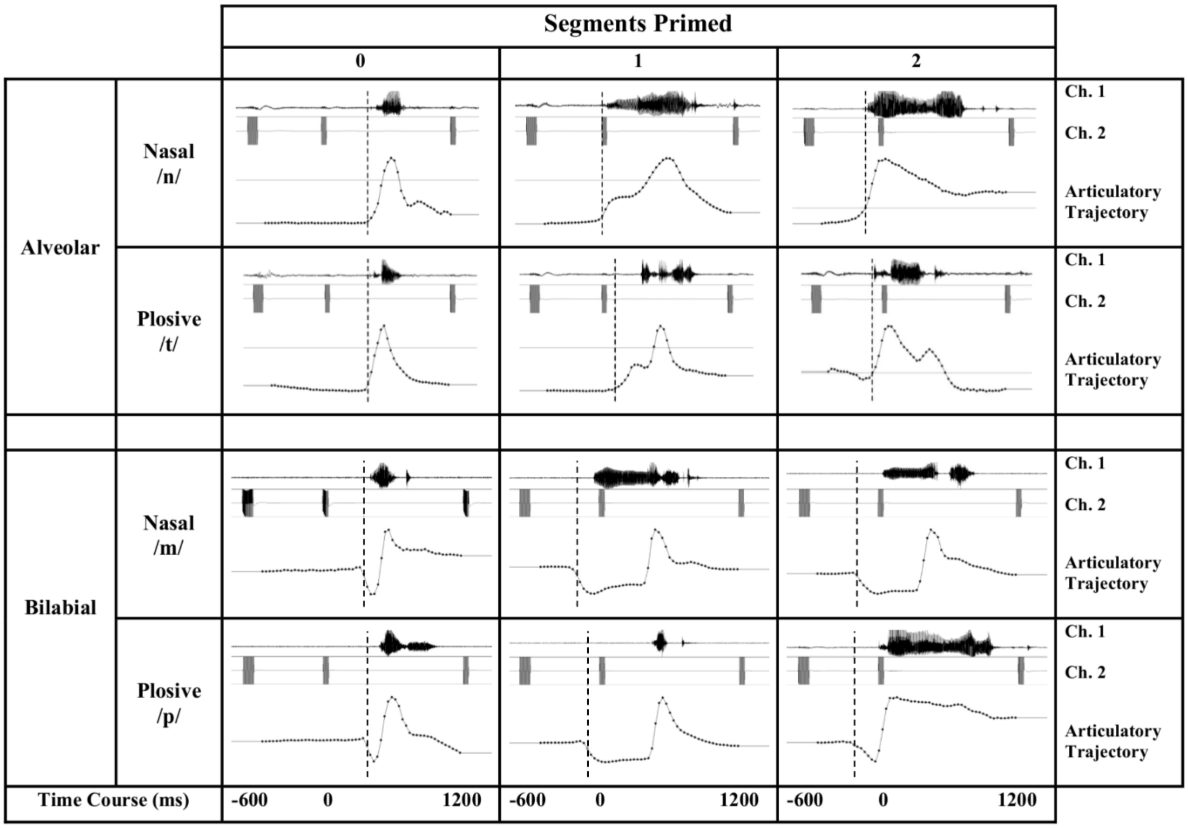

Figure 5. Praat analysis windows for selected alveolar and bilabial trials featuring negative latencies in the segment bias condition. Channel 1 displays the waveform of the participant speaking the word, channel 2 displays the timing beeps provided by RUNWORD (end of first beep = prime onset at -600 ms; end of second beep = stimulus onset at 0 ms; end of third beep = stimulus offset at 1200 ms), and articulatory trajectory is the position of the lips over time (higher = more separation). The dotted vertical line shows articulatory onset, the first moment of lip motion towards the target position (closed). Note the presence of negative articulatory latencies in all conditions when the initial segment is primed, and negative acoustic latencies when permitted by the interaction between the manner of the initial segment and the number of segments primed.

Appendix

Target Words:

mask, mast, match, men, met, mint, mod, musk, mutt

nag, nap, nick, nip, nix, nog, nub, null, nun

pan, pass, path, peck, pelt, pit, pop, pug, punt

tab, tack, tell, tend, test, tongs, toss, tot, tuck

Filler Words:

bad, big, bump, damp, ditch, dud, fans, fill, fish

hedge, hugs, judge, jut, kiss, lance, land, loft

red, rob, sat, sock, sunk, wags, wilt