Characterizing Spoken Responses in Masked Onset Priming of Reading Aloud using Articulography

Michael Proctor
Macquarie University

16 University Ave
Macquarie University
NSW 2109 Australia
+61 2 9850 4137

michael.proctor@mq.edu.au

Max Coltheart
Macquarie University

Louise Ratko
Macquarie University

Tünde Szalay
Macquarie University

Kenneth Forster
University of Arizona

Felicity Cox
Macquarie University

Abstract

A key method for studying articulatory planning at different levels of phonological organization is masked onset priming. In previous work using that paradigm the dependent variable has been *acoustic* response time (RT). We used electromagnetic articulography to measure *articulatory* RTs and the articulatory properties of speech gestures in non-word production in a masked onset priming experiment. Initiation of articulation preceded acoustic response onset by 199 ms, but the acoustic lag varied by up to 63 ms, depending on the phonological structure of the target. Onset priming affected articulatory response latency, but had no effect on gestural duration, inter-gestural coordination, or articulatory velocity. This is consistent with an account of the masked onset priming effect in which the computation from orthography of an abstract phonological representation of the target is initiated earlier in the primed than in the unprimed condition. We discuss the implications of these findings for models of speech production and the scope of articulatory planning and execution.

Levelt's (1989) general account of the cognitive architecture of the process of speech production remains one of the dominant models in the field, yet aspects of the account continue to be debated; in particular, the scope of the minimal planning unit for speech, and whether processing in the later stages of speech production occurs incrementally or in parallel. A key experimental paradigm for exploring these issues is reading aloud. Rastle, Harrington, Croot and Coltheart (2005) elaborated upon the Levelt model to make explicit how it applies to the reading-aloud task, defining Levelt's three stages thus:

i. An abstract phonological representation is created. In the case of reading aloud there are three routes from orthography to phonology (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001). The first involves conceptual knowledge: print accesses conceptual knowledge which in turn generates a phonological representation. This "lexical-semantic reading route" can only be used when the printed stimulus is a word. The second route involves dictionary lookup: the phonological representation corresponding to a printed word is accessed from a phonological lexicon. This "lexical non-semantic reading route" can also only be used when the printed stimulus is a word. The third route involves the application of grapheme-phoneme correspondence (GPC) rules to convert a printed letter string into a sequence of phonemes. This route can be used both for words and nonwords, and is the only route available to nonwords.

ii. An articulatory plan is computed from the abstract phonological representation. For monosyllabic non-words, this might be achieved through two possible mechanisms: the articulatory plan could be retrieved from a mental syllabary, or the plan could be assembled on the fly.

iii. Motor execution of the plan. Articulation can begin as soon as a plan has been prepared, but there is considerable debate about scope of the minimal planning unit for speech: whether it is the syllable (Levelt et al. 1999; Meyer, Belke, Hacker & Mortensen, 2007; Sulpizio & Burani, 2015; Sulpizio, Spinelli, & Burani, 2015), or smaller (Cortese, 1998; Kawamoto, Kello, Jones, & Bame, 1998; Schiller, 1998; 2000). Rastle, Harrington, Coltheart and Palethorpe (2000) and Rastle et al. (2005) present evidence that an articulatory plan must be computed for an entire syllable before motor execution can begin, consistent with earlier findings suggesting that the domain of motor planning is the (stress-bearing) syllable (Sternberg, Monsell, Knoll & Wright, 1978; Monsell, 1986; Sternberg, Knoll, Monsell & Wright, 1988).

Kawamoto, Kello, Jones, and Bame (1998) propose an alternative account of reading-aloud, in which the response begins as soon as the articulation of the initial phoneme is computed. Evidence from single letter priming experiments (Kawamoto, Liu, Lee, & Grebe, 2014), and prolonged duration effects localized to the initial segment of spoken responses (Damian, 2003) are consistent with segmental-level articulatory planning. Schiller and Costa (2006) argue that articulatory planning occurs at a sub-syllabic level to account for findings from naming experiments in which the magnitude of priming effects increases with the number of shared segments, but not as a result of syllable priming per se. Yet it is difficult to account for prosodic planning, irregular GPC mappings, and duration of non-continuant segments without motor plans that apply over a larger scope than the segment (Sulpizio & Burani, 2015). Holbrook, Kawamoto, & Liu (2019) demonstrated that participants can be biased towards either syllable-level or sub-syllabic response behaviour, which may explain some of the differences between these various and seemingly contradictory findings.

A central issue in this debate is the nature of coarticulation, and the extent to which it is planned (Whalen, 1990; Roelofs, 1997; Hawkins & Nguyen, 2004) or an automatic process (e.g. Saltzman & Munhall 1989). Rastle, Harrington, Coltheart and Palethorpe (2000) showed that the initial phoneme of a target utterance is influenced by anticipatory coarticulation from the following vowel in the speeded reading-aloud task, which supports the idea that motor execution cannot begin until the computation of an articulatory plan has been entirely completed. This is also consistent with a large body of work demonstrating that some aspects of coarticulation are pervasive and intrinsic to speech production and perception (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967; Fowler & Saltzman, 1993; Goldstein & Fowler, 2003). However, Liu, Kawamoto, Payne & Dorsey (2018) showed that some aspects of anticipatory coarticulation may be gradient in scope and degree, depending on the availability of information about an utterance, and speakers' approach to the task. Furthermore, in some circumstances, speech movements can be dissociated from higher-level linguistic units: individual speech movements can be initiated before other movements involved in producing the same segment or syllable (Tilsen, Spincemaille, et al. 2016), or halted before completion (Ladefoged, Silverstein & Papçun, 1973; Tilsen & Goldstein, 2012). Speakers vary in how and how much they demonstrate preparatory articulation when information is available which facilitates anticipatory coarticulation (Tilsen et al., 2016; Liu et al., 2018), but these findings are inconsistent with models of speech production in which all aspects of speech timing and implementation are pre-organized entirely at the level of the segment or syllable.

Additional challenges to traditional models in which phonology and articulation are discretely staged come from studies of speech errors and related work showing that multiple competing phonological representations can influence speech output. Goldstein, Pouplier, Chen, Saltzman, and Byrd (2007) demonstrate that speech errors do not typically involve complete substitution of one segment for another; rather, they are characterized by gestural intrusions of varying magnitude, and co-production of both target and error. These phenomena are difficult to reconcile with traditional models in which speech planning and execution are independently staged. Goldrick and Blumstein (2006) and McMillan and Corley (2010) have shown that induced errors leave articulatory and acoustic traces on target productions (which may go unnoticed by a listener), and account for these data with models in which multiple partially-activated phonological representations cascade into articulatory processes.

Further investigation is required to better understand exactly how articulation is planned at different levels of phonological organization, and how speech motor control may be influenced prior to and during execution. An important experimental method which has been used to gain insights into these issues is masked onset priming.

The masked onset priming effect (MOPE).

In the masked onset priming paradigm introduced by Forster and Davis (1991), the subject is presented on each trial with a row of hash marks for, typically, 500 ms, followed by a prime presented in lower case for about 50 ms, followed by a target item presented in upper case, to which the subject must respond (Figure 1). Because the prime is forward-masked (by hash marks) and backward-masked (by the target), subjects are typically unaware that a prime has been presented; nevertheless, the prime influences the response to the target.

The MOPE refers to the finding that RTs for reading aloud targets are shorter when prime and target share their initial letter/phoneme than when they share no letters/phonemes. This effect occurs when primes and targets are words (Bowers et al. 1998), and also when primes and targets are nonwords (Lukatela & Turvey, 1994; Berent & Perfetti, 1995; Kinoshita, 2000).

Using nonwords as primes and targets is the most direct way to study the MOPE because it avoids potentially confounding influences such as lexical frequencies of primes and targets and also makes it easier to use orthographically and phonemically simple letter strings as primes

and targets. In a study of this kind using CVC nonwords as primes and targets, Kinoshita (2000) reported that reading-aloud RTs to targets were faster when prime and target shared their initial letter/phoneme (as in the prime-target sequence *suf*–SIB) than when they shared no letters/phonemes (e.g. *mof*–SIB). There was no priming effect when the last letter was shared (*mub*–SIB and *mof*–SIB did not differ in RT). There was a non-significant 3 ms advantage when prime and target shared their first two letters/ phonemes (*sif*–SIB) compared to when they shared only their first letter/phoneme (*suf*–SIB).

These results were successfully simulated by Mousikou, Coltheart, Finkbeiner and Saunders (2010a) using the DRC computational model of reading aloud (Coltheart et al., 2001). Mousikou et al. (2010a) concluded that this successful simulation supported the hypothesis that the MOPE is due to a left-to-right process of translating letters to sounds – the "nonlexical reading route" – that is applied to the prime. Only the first letter of the prime can be reliably translated from orthography to phonology by this process because the backward masking effect of the target erases the letter representations of the prime before the serially-operating left-to-right orthography-to-phonology translation process can proceed beyond the first letter of the prime.

The study by Kinoshita (2000) which showed a significant first letter MOPE but only a trend toward a second letter advantage was repeated by Mousikou et al. (2010a) with more statistical power (more subjects, an extended set of stimuli, and a more sensitive method of measuring reading-aloud latency). When prime and target shared just their initial letter/phoneme, there was a MOPE of 33 ms (target RTs were 33 ms faster for *suf*–SIB than for *mof*–SIB). Having prime and target share their first two letters gave an additional 4 ms advantage: the MOPE for *sif*–SIB was 37 ms. This 4 ms advantage, though tiny, was significant (in both subject and item analyses). These results showed that in the MOPE paradigm, there can be a very small contribution to priming from the second letter of the prime. This finding suggests that on a very few occasions the left-to-right translation of letters to sounds that is being applied to the prime can extend to the second letter and begin translating it to phonology before all of the prime's letters are erased by the following target letters; the translation of the prime's letters to their phonology must cease at that point since the prime's letters are no longer available.

The priming of SIB by *sif* might be caused by their sharing their initial letters (an orthographic effect) or by sharing their initial phonemes (a phonological effect). These possibilities can be

disentangled. When they have been, evidence from Australian English (Mousikou, Coltheart and Saunders, 2010b) and Dutch (Schiller 2007) suggests that the MOPE is a phonological effect and not an orthographic effect (e.g. *kalt* primes CALC).

In virtually all previous work on the MOPE, the dependent variable has been "acoustic RT": the time between the onset of the target and the onset of acoustic energy emitted by the subject when reading aloud. Of the many limitations of this method (Kessler et al. 2002; Holbrook et al. 2019), most problematic is that the acoustic response lags initiation of articulation by a variable amount of time depending on the phonological structure of the utterance (Rastle & Davis, 2002). Davis et al. (2013) demonstrated how articulography can be used in MOPE experiments, and found that it provides "a more sensitive measure of prime effects than the acoustic (RT)". Davis et al. (2015) examined labial and tongue tip articulation using masked priming, but differences between articulatory and acoustic response times have not been systematically examined in this paradigm. Our goal here is to study the nature of vocal responses in a masked onset priming paradigm in new detail and with greater accuracy by systematically examining *articulatory* activity during reading aloud, and comparing the articulatory and acoustic responses.

Measuring articulation as subjects read aloud.

We measured articulatory responses directly by using electromagnetic articulography (EMA: Perkell et al., 1992). In this technique, sensors placed on key speech organs are tracked as they move in an electromagnetic field. We used sensors attached to the upper lip, lower lip, jaw, tongue dorsum, tongue blade and tongue tip (details below) to determine the configuration of the articulatory apparatus at any point in time during each trial of a reading-aloud experiment, including during the actual execution of the reading-aloud response. We were therefore able to determine the exact time after stimulus onset that the first target-related movement occurred ("articulatory RT"). This method allows investigation of how masked onset priming influences articulation of each part of the target – onset consonant, nuclear vowel, and coda consonant – by examining:
  – gestural duration in each part of the response
  – articulator velocities associated with each gesture in the response
  – total duration of the articulatory response

By providing finer details about the spoken response in the MOPE paradigm unavailable through acoustic analysis, these data can offer new insights into models of speech production and the mechanisms involved in planning and execution. One account of the MOPE, using the DRC computational model of reading (Mousikou et al. 2010a), asserts that the faster acoustic response in the primed condition results from a linear leftward shift of the entire spoken response. Mousikou et al. (2010a) propose that both facilitatory and inhibitory effects are involved: when prime and target do not match, the activation of the prime's first phoneme competes strongly with that of the target's first phoneme and delays the spoken response compared to the overlapped (matching) conditions (inhibitory effect). Additionally, when the target's first phoneme is pre-activated by its prime in the overlapped conditions, response times are reduced compared to the non-matching condition (facilitatory effect). However, because the DRC model requires that a full motor plan for a syllable must be complete before articulation can begin (Rastle et al. 2000), any response time advantage conferred through priming must affect the target as a whole, not its individual phonemes. This account therefore predicts that masked onset priming will not have any effects on any specific parameters of individual (segmental or sub-segmental) components of the articulatory response, nor on the duration of that response. If, on the other hand, articulation of segments within the target is differentially affected by priming, this would be more consistent with models which allow for articulation to be initiated before the motor plan for an entire syllable is complete (e.g. Kawamoto et al., 2019; Drake & Corley, 2015), and with models in which "phonetic processes flexibly and adaptively use whatever phonological information is available, very early in the time course of articulation" (Krause & Kawamoto, 2020: 121). Articulographic data allow us to examine the timecourse of articulation in spoken responses elicited using the MOPE paradigm, to test these different accounts with greater scrutiny than was previously possible using acoustic response time data alone.

EXPERIMENT

This was a replication of a study which measured only acoustic responses (Mousikou et al., 2010a, Experiment 1), with the sole difference that, in addition to acoustic measurements, we also measured articulatory responses. We elected to do this to determine whether the additional requirements of our study (the tracking of articulation) prevented our being able replicate Mousikou et al. (2010a)'s finding of a MOPE on acoustic responses.

Method

Materials

The experimental stimuli used in this study were those used in Mousikou et al. (2010a), reproduced in Appendix A. These items were 324 three-letter-long pronounceable nonwords with graphemic and phonological C_1VC_2 (consonant–vowel–consonant) structure, taken from the ARC Nonword Database (Rastle, Harrington and Coltheart, 2002). A total of 81 nonwords served as target items, and the remaining 243 served as their prime pairs. In particular, three groups of 81 prime–target pairs were formed with the targets remaining the same in all three groups. Three types of prime that matched on mean orthographic neighborhood size were used: primes that shared their first letter and phoneme with their targets (e.g., *suf*–SIB; one-letter overlap condition); primes that shared their first two letters and phonemes with their targets (e.g., *sif*–SIB; two-letter overlap condition); primes that shared no letters or phonemes with their targets in the same position[1] (e.g., *mof*–SIB; unrelated condition).

In addition to the three groups of 81 prime–target pairs that formed the experimental stimuli, 9 more pairs of primes and targets that matched the experimental stimuli on the same criteria were selected as practice items.

Design

Each condition (1L: one-letter overlap, 2L: two-letter overlap, 0L: no related letters) consisted of 81 prime–target pairs for a total of 243 pairs per subject in a fully counterbalanced design. This meant that every subject saw the 81 targets three times, each time in a different prime-type condition. A mixed design was used so that the three prime-type conditions were presented in a random order across the experiment. The 243 trials were divided into three blocks so that the same target would not appear more than once within the same block. A brief break was administered between the blocks. Three lists (A, B, and C) were constructed to counterbalance the order of block presentation.

Subjects

---

[1] Mousikou et al. (2010) noted: "Due to an oversight, two pairs in the unrelated condition shared the same phoneme in the same position—that is, dys–PIV and pym–VIC".

7 native monolingual speakers of Australian English (3M, 4F, mean age 22 years) participated in the experiment; all were undergraduate students at Macquarie University who reported having no history of speech or hearing problems. 2 participants were tested with List A, 2 with List B, and 3 with List C. All subjects provided informed consent before participation, and all procedures were conducted under the guidelines and with the approval of the Macquarie University Human Research Ethics Committee.

Procedure

Subjects were tested individually, seated approximately 150 cm in front of a Dell P2715Q 27" monitor, upon which the stimuli were presented. Subjects were instructed (verbally first and then by written instructions on the monitor) that a list of uppercase nonwords, preceded by a row of three of hash marks (###), would be presented on the screen one at a time and that their task was to read aloud the nonword in upper-case letters as quickly and as accurately as possible. The presence of a prime was not mentioned to the subjects. Stimuli were presented to each subject in a different random order, following a series of practice trials that matched the experimental stimuli on the same criteria. Instructions and stimuli were presented, and naming latencies were recorded to the nearest millisecond, using the DMDX display system (Forster & Forster, 2003) on an ASUS V230IC computer.

The trial design is shown in Figure 1. Each trial started with the presentation of a forward mask (###) that remained on the screen for 500 ms. The prime was then presented in lower-case letters for 50 ms,[2] followed by the target, which was presented in upper-case letters and acted as a backward mask to the prime. The target words appeared in white on a black background (Courier New, 12pt font) and remained on the screen for 2,000 ms or until the subject responded. Hashmarks, prime and target were presented in the same place on the screen. The inter-trial interval was 1,000 ms. Order of trial presentation within blocks and lists was randomized across subjects. Exit interviews were conducted at the end of each experiment to ask participants what they observed, and whether they noted anything unusual; none reported being aware of the masked prime.

---

[2] Mousikou et al. (2010a) used a prime duration of 55.6 ms, due to differences in monitor refresh rates.
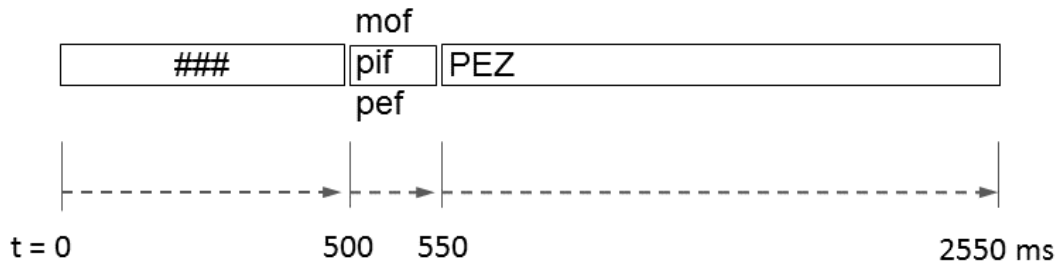
Figure 1: Structure of each trial in the MOPE experiment. Forward mask (500 ms), followed by prime (50 ms), followed by non-word target.

Measuring articulatory movements with EMA

Articulation was tracked using an NDI Wave system sampling 5DoF sensors at 100 Hz. Sensors were taped to the upper lip (UL) and lower lip (LL), and glued to the tongue tip (TT), tongue blade (TB), tongue dorsum (TD), and to the gum below the lower incisor to track jaw movement (JW). Sensors were aligned along the midsagittal plane of each participant's vocal tract. The most anterior lingual sensor was located on top of the tongue, approximately 10 mm posterior to the actual TT. The TD sensor was located approximately 40 mm posterior to the TT sensor, and the TB sensor was placed halfway between the other two lingual sensors, along the lingual midline (Figure 2). Three additional sensors were attached to the nasion (NA) and right and left mastoid processes (RM, LM), to track head movement.
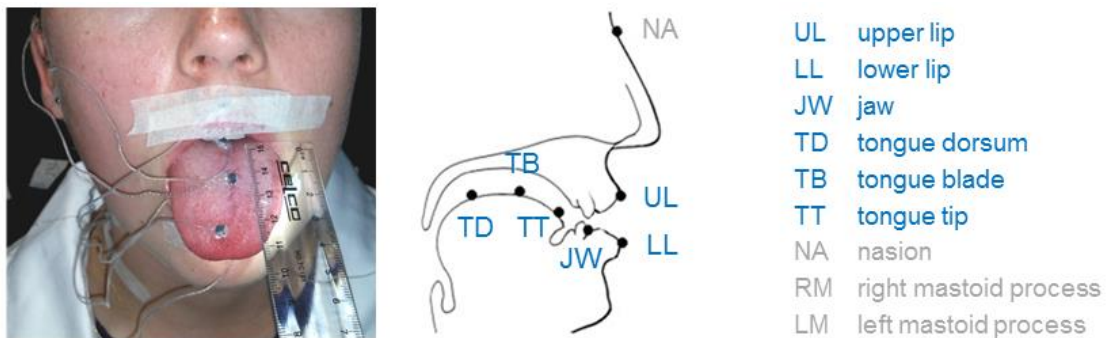

Figure 2: Configuration of EMA sensors used to track articulation.

Acoustic responses were recorded twice using two microphones located in front of the participant, 150 cm from their lips and offset 15° from their line of sight to the presentation monitor. A Røde NTG-1 was connected through a Focusrite OctoPre MkII preamplifier to the NDI Wave system, which recorded synchronized acoustic data simultaneously with the spatial data from the sensor coils. A second microphone (Røde NT1-A) was connected through a separate Focusrite OctoPre MkII preamplifier to the computer presenting the experimental

stimuli, so that the same responses could be captured by the DMDX system, and recorded as a series of WAV files synchronized with the beginning of each trial.

Two additional articulatory recordings were made in each experiment. A bite trial was recorded as the participant clenched their teeth onto a small plate, to determine location of the reference sensors with respect to the occlusal plane. Finally, a probe configured as a 6DoF NDI sensor was used to trace the midline of the roof of the mouth, to locate the participant's palate. All sensor position data were subsequently corrected for head movement with reference to the fixed reference sensors (NA-LM-RM), and rotated into a common coordinate system defined around an origin located on the midsagittal occlusal plane, immediately behind the upper incisors. Further details of the methods used for EMA acquisition and post-processing can be found in Tiede et al. (2010).

Measuring acoustic RT

Acoustic RTs were determined from the WAV files captured by DMDX, in which the start of each audio recording corresponds to the temporal origin of the associated trial (Figure 1: t = 0). In each recording, RMS energy and short-time average amplitude were calculated over the length of the signal, in 20 ms 75%-overlapped Hamming-windowed intervals. Working outwards from the amplitude peak, the first and last points in time were located at which the signal amplitude exceeded a subject-specific threshold, initially set at 10% of maximum. Working outwards from the point of peak RMS energy, the first and last points in time were also located at which signal energy exceeded another subject-specific threshold, initially set at 1% of maximum RMS energy. The acoustic limits of the vocal response were then automatically estimated as the most peripheral of the two estimates (energy- and amplitude-based). An expert annotator inspected the automatic estimates of the acoustic limits in each recording, superimposed on time-aligned waveforms and short-time spectrograms plotted up to 11 kHz, and manually corrected utterance limits where necessary (Figure 3).
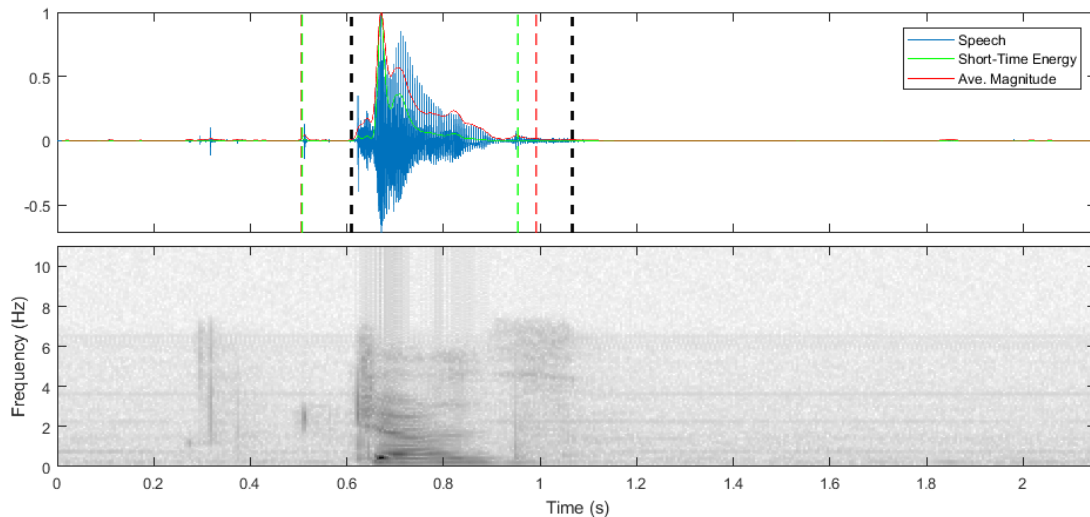
Figure 3: Estimation of acoustic RT and acoustic utterance duration using multi-pass semi-automatic detection of short-time acoustic energy and amplitude thresholds. Example utterance 'JOZ' [dʒɔz] by W01. Top panel: acoustic waveform and utterance limit estimates. Bottom panel: time-aligned speech spectrogram. Initial estimate incorrectly located at pre-utterance noise burst (505 ms; vertical dashed green line) manually corrected to beginning of acoustic release of onset affricate [dʒ] (614 ms; right-most vertical dashed black line). End of utterance noise manually corrected to end of fricative [z] using spectral information (1080 ms; right-most vertical dashed black line).

Measuring articulatory RT

Companion acoustic recordings of each trial – captured independently by DMDX and the articulograph – were aligned in post-processing by finding the lag associated with the maximal cross-correlation between the two signals. 2s intervals were selected from the Wave articulographic recordings, beginning at each point in time corresponding to the start of a trial, and loaded into Matlab for analysis. Sensor traces were low-pass filtered and conditioned using a DCT-based discretized smoothing spline (Garcia, 2010) operating over 2s analysis windows. An additional signal was derived for analysis of labial consonants: Lip Aperture (LA), the Euclidean distance between the UL and LL sensors (i.e. how far apart the lips are) at each point in time.

Articulographic data were analyzed using MVIEW (Tiede, 2010), a graphical interface which displays time-aligned articulatory and acoustic data and facilitates automatic analysis of speech gestures. For each target utterance, up to four gestures were identified: (g1) the primary articulation associated with the onset consonant, (g2) the primary articulation associated with the vowel, (g3) the primary articulation associated with the coda consonant, and (g0) any significant articulation before the onset consonant, such as lip opening or jaw lowering. Examples of these gestures are illustrated in Figure 4.
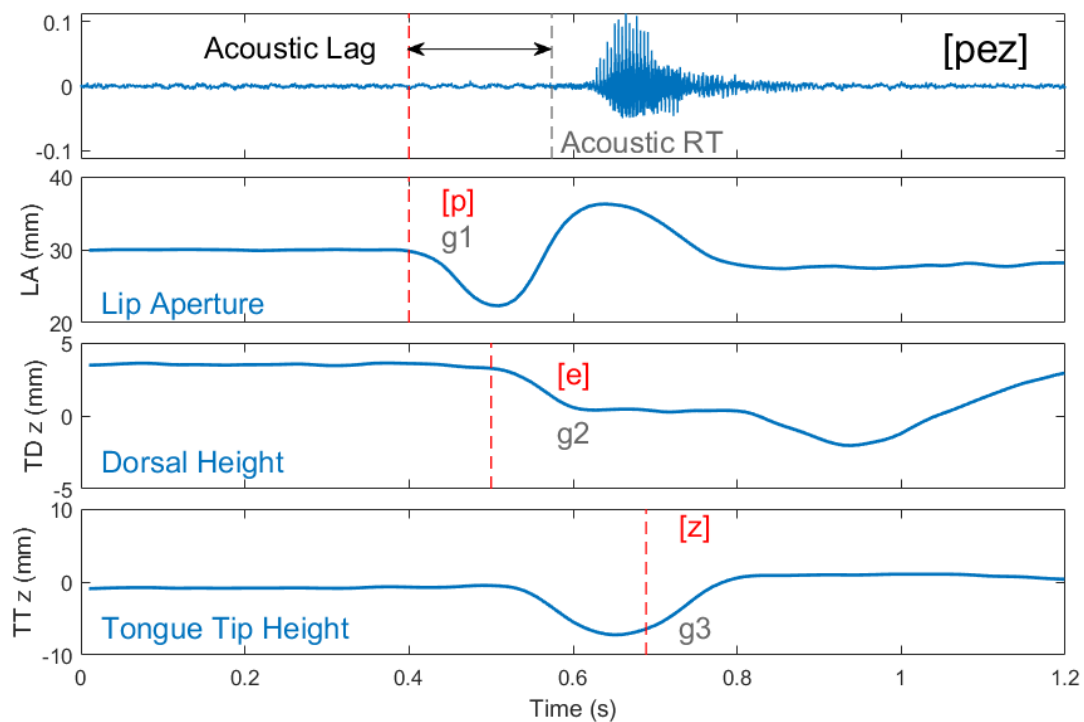
Figure 4: Gestural analysis of target utterance 'PEZ' [pez] by W02. Top panel: acoustic waveform. g1: labial closure gesture of onset consonant /p/ (0.40s); g2: tongue dorsum lowering gesture of vowel /e/ (0.50s); g3: tongue tip raising gesture of coda consonant /z/ (0.69s). No pre-onset gesture (g0) evident in this utterance. Acoustic signal begins at release of onset /p/ (end of g1) – grey dashed line (0.57s).

For each gesture in the target utterance, five temporal landmarks were identified from the associated sensor velocity (Figure 5). The first was GONS (Gestural ONSet): the point in time at which articulatory velocity first exceeds 20% of maximum closure velocity. This is a standard landmark used to locate the beginning of a speech gesture (Hoole et al., 1994). In the utterance of 'PEZ' in Figure 5, for example, the first articulatory movement associated with the target is labial closure to form the stop consonant [p]. LA velocity first exceeds 20% of maximum closure velocity 0.4 s into the trial, so the articulatory RT can be estimated at 400 ms.
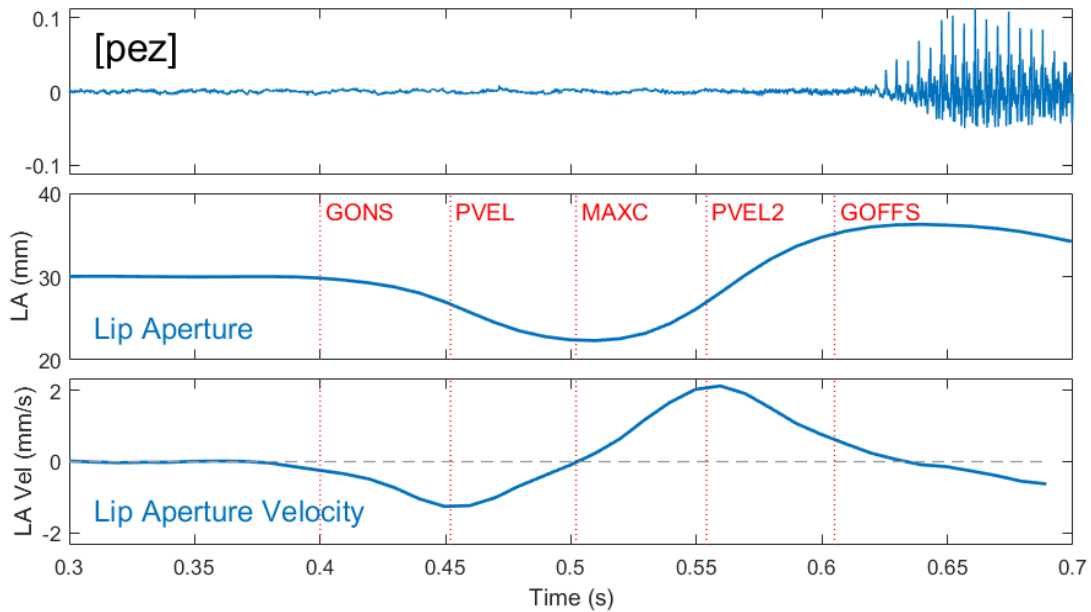
Figure 5: Temporal landmarks in initial gesture (g1) of utterance 'PEZ' [pez] by W02. Top panel: acoustic waveform; 2nd panel: lip aperture; Bottom panel: velocity of labial closure. Gestural landmarks: 1. GONS: gestural onset; 2. PVEL: max onset velocity; 3. MAXC: maximum constriction; 4. PVEL2: maximum offset velocity; 5. GOFFS: gestural offset.

Analyzing Effects on RTs

We dealt with issues concerning statistical sensitivity by using Bayesian analyses; specifically by calculating Bayes Factors (BFs). This requires that two hypotheses be specified, as BF is the ratio of the likelihood of one hypothesis to the likelihood of the other. Typically, one of these hypotheses is the null hypothesis; that was so in all of our BF calculations. Since our experiment was a replication of a previous study, the alternative hypothesis was informed by previous findings: the effects obtained by Mousikou et al. (2010a). For novel aspects of the study there were no directly comparable previous findings to inform an alternative; here the two hypotheses we compared were that the data can be described using a model including the effect being examined, versus a model without the effect of interest (Rouder et al, 2009; Kruschke & Liddell 2018; Vasishth et al. 2018). We used the conventional decision rule (Dienes and McLatchie, 2018) of accepting the null hypothesis if BF in favor of the null hypothesis is greater than 3, rejecting the null hypothesis if this BF is less than 1/3, and regarding our data as inconclusive if BF falls between these limits.

We consider this Bayesian approach to be superior to conventional hypothesis testing and power analysis for a number of reasons (see Dienes, 2014; Dienes & McLatchie, 2018; Kruschke & Liddell, 2018). Most importantly, with frequentist approaches, a non-significant

p-value is not informative about whether there is evidence for the null, evidence against the null, or no evidence for any conclusion between hypotheses. For example, a non-significant result can be consistent with substantial evidence in favor of the null, even when power is low; and a non-significant result can be consistent with no substantial evidence for or against the null, even when power is high (Dienes & McLatchie, 2018). In contrast, BF analysis can allow one to assert that the evidence favors the null hypothesis.

BFs for replication aspects of the study were calculated using an online implementation (Singh 2018; Baguley & Kaye 2010) of Dienes & McLatchie's (2017) BF calculator. Non-Bayesian Linear Mixed Effects modelling was conducted using the *lme4* (Bates et al. 2014, v.1.1-18-1) and *lmerTest* packages (Kuznetsova et al., 2017, v.3.0-1). All other statistics were calculated in R using the *BayesFactor* package (Morey & Rouder, 2018) with default rscale = 0.707. Where appropriate, we report *t*-values and significance levels for comparison with Bayesian analyses (Dienes & McLatchie, 2018). Bayesian *t*-tests on posterior distributions were calculated using Gaussian quadrature over JZS priors with r-scale=0.707 (Morey & Rouder, 2018), and are reported as BF_{10} (alternative/null).

Results

1,701 trials were presented in total (243 trials $\times$ 7 participants). In 17 trials the subject mispronounced the target, and in another 3 trials no response was recorded, leaving 1,681 valid responses for acoustic analysis. 17 trials (1.2 %) were excluded because the articulatory RT was faster than 100 ms (Luce 1986; Whelan 2008). In 181 trials no gesture associated with the onset consonant of the target could be reliably measured. 1,486 trials remained available for articulatory analysis of target onset consonant (C1), 1,574 trials for articulatory analysis of target vowel, and 1,326 trials for articulatory analysis of target coda consonant (C2).

RTs for individual participants and for the experimental population as a whole generally followed a log-normal distribution, so additional analyses were conducted using log-transformed and inverse RT data (-1000/RT) to ensure that statistical models were valid despite deviations from normality in the slower RTs (Baayen & Milin, 2010; Brysbaert & Stevens 2018). Raw RTs (ms) and models based on untransformed data are reported in all cases where no such discrepancy was found.

Acoustic RT: Target Onset

Spoken RTs, measured at acoustic onset of the target, were faster for primed trials than for unprimed utterances. Grand mean acoustic RT for the unprimed (0L) trials was 606 ms (s.d. 89 ms), 587 (97) ms for 1L–primed trials, and 583 (98) ms for 2L–primed trials (Figure 6). We analyzed these RT differences by subject and by item.
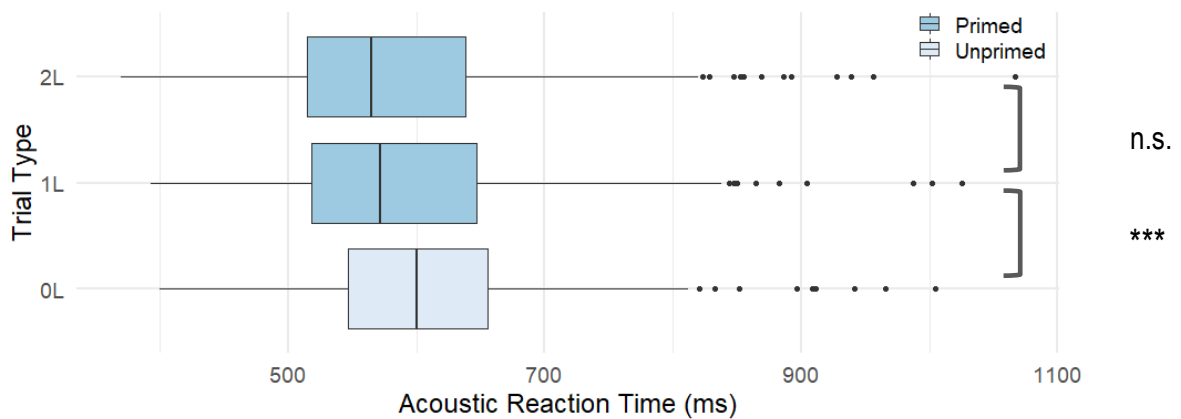

Figure 6: Grand mean acoustic RTs by trial type: 2L–primed targets (e.g. *sif*–SIZ) are produced after a shorter delay (583 ms) than 1L–primed (*sev*–SIZ; 587 ms) and unprimed (0L) utterances (e.g. *lep*–SIZ; 606 ms).

By-subject analyses

Averaged over individual subject means, the one-letter-overlap MOPE effect – the difference between the 0-letter overlap condition and the 1-letter overlap condition – was 20 ms (SEM 5.28, $t(6) = 3.79$, one-tailed $p < .005$). In this condition, Mousikou et al. (2010a) found a significant MOPE of 33 ms. Since our study was a replication of theirs, we use that value to represent the alternative hypothesis in our BF analysis. The BF (one-tailed) was 344 in favor of the alternative hypothesis over the null hypothesis, yielding very strong evidence that there was a one-letter-overlap MOPE in the acoustic RT data of our experiment.

Averaged over individual subject means, the second-letter-overlap MOPE effect – the difference between the 1L and 2L conditions – was 3 ms (SEM 1.51, $t(6) = 1.99$ one-tailed $p < .05$). In this condition, Mousikou et al. (2010a) found a significant MOPE of 4 ms, so we use that value to represent the alternative hypothesis in our BF analysis. The BF (one-tailed) was 3.84 in favor of the alternative hypothesis over the null hypothesis, yielding acceptable evidence that adding a second letter of overlap between prime and target augmented the MOPE in the acoustic RT data of our experiment, as was found by Mousikou et al. (2010a). For all

participants other than M02, acoustic responses to 1L– and 2L–primed trials were faster than unprimed responses by at least 13 ms (Figure 7).
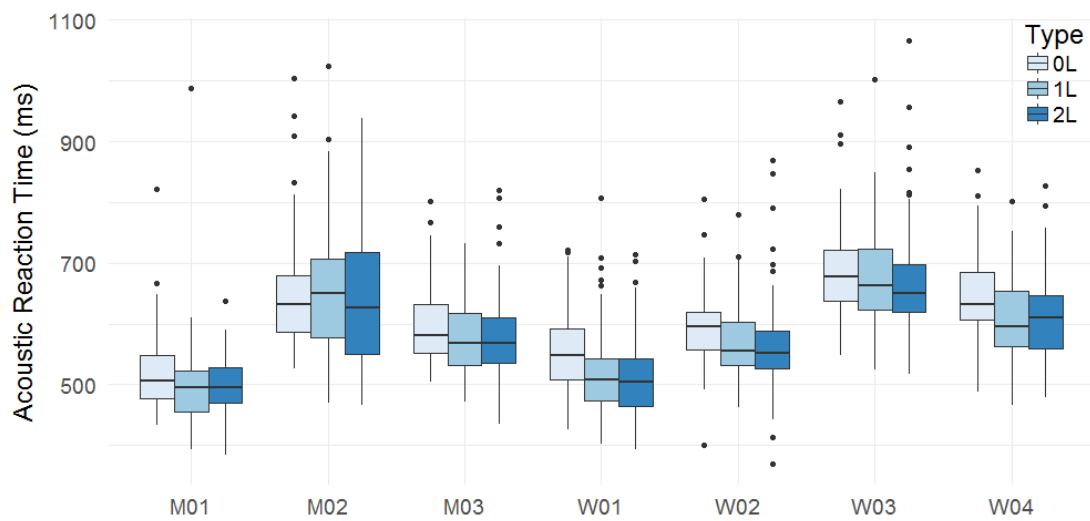

Figure 7: Acoustic RTs (ms) by subject (M01-M03, W01-W04) and trial type: (0L/1L/2L).

By-item analyses

Averaged over individual item means, the one-letter-overlap MOPE – the difference between 0L and 1L conditions – was 19 ms (SEM 4.22, $t(80) = 4.50$, one-tailed $p < .0005$). We again used the value obtained by Mousikou et al. (2010a) of 33 ms to represent the alternative hypothesis in our BF analysis. The BF (one-tailed) was 5424 in favor of the alternative hypothesis over the null hypothesis, yielding very strong evidence that there was a one-letter-overlap MOPE in the acoustic RT data.

Averaged over individual item means, the second-letter-overlap MOPE effect – the difference between 1L and 2L conditions – was 4 ms (SEM 4.44, $t(80) = 0.90$, one-tailed $p < .15$). We use the value Mousikou et al. (2010a) of 4 ms to represent the alternative hypothesis in our BF analysis. The BF (one-tailed) was 1.29 in favor of the alternative hypothesis over the null hypothesis, yielding inconclusive evidence concerning whether or not adding a second letter of overlap between prime and target augmented the MOPE in our acoustic RT data.

In sum, we replicated Mousikou et al. (2010a) in finding a very strong one-letter overlap MOPE, in both our subject and item analyses. There was much weaker evidence that adding a

second letter of overlap between prime and target augmented the MOPE in our acoustic RT data (a result obtained by Mousikou et al. (2010a)). This was not of great concern to us because the main aim of our study was to investigate how articulatory responses were affected by one-letter overlap priming.

Articulatory RT: Target Onset Consonant

The earliest robust articulatory landmark that can be compared across utterances is the onset (GONS) of the initial gesture (g1) associated with the first consonant in the target (Figure 4). As such, we take g1 GONS (Figure 5) to be the primary measure of articulatory RT, and focus our analyses on RTs measured at this point in time, before turning to the other gestural landmarks in the target response. Grand mean articulatory RT for unprimed (0L) trials, measured at g1 GONS, was 415 ms (s.d. 99 ms); mean articulatory RT for 1L–primed trials was 396 (107) ms. Mean articulatory response latency for 2L–primed trials was 394 (103) ms (Figure 8). We analyzed these RT differences by subject and by item.

Figure 8: Grand mean articulatory RTs (GONS of g1: first gesture in utterance) by trial type: 2L–primed (394 ms) and 1L–primed (396 ms) targets are produced after a shorter delay than unprimed (0L) utterances (415 ms).

By-subject analyses.

Averaged over individual subject means, the one-letter-overlap MOPE – the difference between the 0-letter overlap condition and the 1-letter overlap condition – was 22 ms (SEM 7.17). Given that no difference was found in acoustic RT lag between trial conditions, we can use previous findings on acoustic RT to inform our hypotheses about articulatory responses, which begin approximately 200 ms earlier, across the board. We therefore used the value of 33 ms found by Mousikou et al. (2010a) to represent the alternative hypothesis in a BF analysis.

The BF (one-tailed) was 37.9 in favor of the alternative hypothesis over the null hypothesis, so we conclude that there was a one-letter-overlap MOPE in our articulatory RT data.

Averaged over individual subject means, the second-letter-overlap MOPE – the difference between the 1-letter overlap condition and the 2-letter overlap condition – was 1 ms (SEM 6.04). We used the value of 4 ms found by Mousikou et al. (2010a) to represent the alternative hypothesis in a BF analysis. The BF (one-tailed) was 0.90 in favor of the alternative hypothesis over the null hypothesis, so these data are inconclusive as to whether a second letter of overlap between prime and target augmented the MOPE in our articulatory RT data.

For all participants other than M02, articulatory responses to 1L– and 2L–primed trials were faster than unprimed responses by at least 7 ms (Figure 9).

Figure 9: Articulatory RTs (measured at g1 GONS) by subject (M01-M03, W01-W04) and trial type: (0L/1L/2L).

By-item analyses

Averaged over individual item means, the one-letter-overlap MOPE – the difference between the 0-letter overlap condition and the 1-letter overlap condition – was 18 ms (SEM 5.11). We used the value of 33 ms found by Mousikou et al. (2010a) to represent the alternative hypothesis in a BF analysis. The BF (one-tailed) was 130.6 in favor of the alternative hypothesis over the null hypothesis, so we conclude that there was a one-letter-overlap MOPE in our articulatory RT data.

Averaged over individual item means, the second-letter-overlap MOPE – the difference between the 1-letter overlap condition and the 2-letter overlap condition – was 2 ms (SEM 5.22). We used the value of 4 ms found by Mousikou et al. (2010a) to represent the alternative hypothesis in a BF analysis. The BF (one-tailed) was 0.97 in favor of the alternative hypothesis over the null hypothesis, so our data are inconclusive as to whether a second letter of overlap between prime and target augmented the MOPE in our articulatory RT data.

In sum, we found a very strong one-letter overlap MOPE on articulatory RTs measured at g1 GONS. Our data provide no conclusive evidence that adding a second letter of overlap between prime and target augmented the MOPE in the articulatory RT data of our experiment.

Acoustic-Articulatory RT Lag

RTs measured at the beginning of the first gesture of the target (g1 GONS: Figure 5) were detectable, on average, 199 ms earlier than at acoustic onset (s.d. 87 ms). Grand mean acoustic RT lag – the difference between acoustic RT and g1 GONS (see Figure 4) – was 196 ms for unprimed trials (0L), 201 ms for 1L trials, and 199 ms for 2L trials. A Bayesian analysis examined the likelihood that acoustic lag was affected by Trial Type. For the null hypothesis, a linear model of acoustic lag was constructed with effects of Subject and Target alone. For the alternative, acoustic lag was modelled as a function of Trial Type with random slopes and intercepts for Subject and Target. Comparison of models yielded a BF < 0.0000001 (alternative/null), providing very strong evidence that priming does not affect the acoustic lag of the target response.

Acoustic RT lag varied with the phonological class of the onset consonant of the target. Grand mean acoustic RT for targets beginning with a stop /p b t d k g/ was 222 ms slower than the corresponding articulatory response; acoustic lag for sonorants /m n l ɹ/ was 176 ms; and for fricatives /f v s z/ was 163 ms (Figure 10). A Bayesian analysis examined the likelihood that acoustic lag was affected by target onset consonant class. For the null hypothesis, a linear model of acoustic lag was constructed with effects of Subject and Target alone. For the alternative, acoustic lag was modelled as a function of Consonant Class with random slopes and intercepts for Subject and Target. Comparison of models yielded a BF > 1,000,000 (alternative/null), providing very strong evidence that target onset consonant class affects the acoustic lag of the target response.
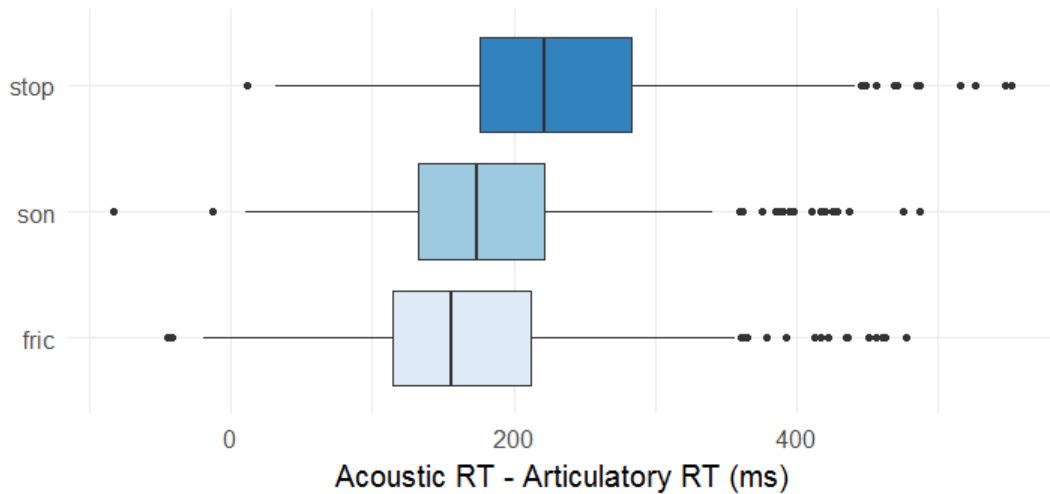
Figure 10: Mean acoustic RT lag (difference between acoustic and articulatory RTs) by type of onset consonant: targets beginning with stops are produced with a greater mean acoustic delay (222 ms) than sonorants (176 ms) and fricatives (163 ms).

Articulatory Response at Other Gestural Landmarks of the Target Onset Consonant

In addition to GONS, articulatory response latency was measured at each of the other four gestural landmarks in the Target onset consonant C1 (Figure 5). Grand mean RTs are summarized in Table 1. At each landmark, linear models were constructed of log RT as a function of Trial Type with random intercepts and random slopes for Subject and random intercepts for Target. The LMEs estimated MOPEs of at least –16 ms at all five landmarks on the consonant gesture. For the null hypothesis, a linear model of articulatory RT was constructed with effects of Subject and Target alone. BFs for model comparisons at each landmark are listed in Table 1 (BF Type). At all timepoints, BF > 3 (alternative/null), offering support for the hypothesis that priming affects the articulatory RT of the target response measured at each part of the gesture associated with the onset consonant.

Articulatory Landmark	0L	1L	1L MOPE	BF 1L-0L	2L	2L MOPE	BF 2L-0L	BF Prime
g1 GONS: gestural onset	413	390	-23	266	390	-23	68	2,640
g1 PVEL: max. onset velocity	483	461	-22	417	461	-23	380	9,197
g1 MAXC: max. constriction	594	576	-18	76	575	-20	65	1,020
g1 PVEL2: max. offset velocity	710	695	-14	4	691	-18	18	116
g1 GOFFS: gestural offset	764	749	-15	12	745	-20	22	263
Duration of Onset Consonant Gesture (GOFFS – GONS)	**351**	**359**			**355**			

Table 1: Grand mean articulatory RTs and MOPEs (ms) for each trial condition, measured at each gestural landmark on g1: target onset consonant C1. BF: likelihood that model including effect of each Trial Type can account for RT differences, compared to null model. BF Prime: likelihood that model including overall effect of Priming (1L or 2L) can account for RT differences at each landmark, compared to null model.

Additional analyses examined the effect of any type of priming at each landmark. The null hypothesis that articulatory RT was a function of Subject and Target alone was compared with the alternative that priming of either type and its interactions with Subject influenced RT. BFs estimated at all five landmarks (Table 1, rightmost column) exceeded 116 (alternative/null), providing very strong evidence that priming of either type (1L and 2L) reduced RT at all key points during the articulation of the onset consonant of the Target utterance.

Articulator Velocity

The velocity of the sensor attached to the primary articulator involved in consonant production (TT for /t, d, n, s, z, l/; TD for /k, g/, etc.) was measured at each gestural landmark in the Target onset C1. Grand mean articulator velocities for each trial condition are summarized in Table 2. Articulator velocity (z-transformed by subject) was modelled as a function of Trial Type with random intercepts and random slopes for Subject and Target. For the null hypothesis, a linear model of articulator velocity was constructed with effects of Subject and Target alone. BFs for model comparisons at each landmark are listed in Table 2, for velocity differences between the unprimed and 1L overlap conditions (BF 0L-1L) and between the unprimed and 2L overlap conditions (BF 0L-2L). The data are inconclusive about the effect of either type of priming on articulator velocity at gestural onset, and about the effect of 2L priming at peak velocity during gestural constriction formation (PVEL). At all other timepoints, BF < 1/3 (alternative/null), offering support for the null hypothesis that priming does not affect articulator velocity in the onset consonant of the response utterance.

Articulatory Landmark	v0L	v1L	v0L -v1L	BF 0L-1L	v2L	v0L -v2L	BF 0L-2L
g1 GONS: gestural onset	2.6	2.5	0.16	0.38	2.5	0.16	0.57
g1 PVEL: max. onset velocity	7.1	6.9	0.15	0.11	6.7	0.39	1.14
g1 PVEL2: max. offset velocity	13.4	13.2	0.20	0.10	13.1	0.36	0.13
g1 GOFFS: gestural offset	5.6	5.4	0.20	0.17	5.5	0.14	0.09

Table 2: Grand mean articulator velocities and velocity differences (cm/s) for each trial condition, measured at each gestural landmark on g1: target onset consonant C1. GONS is the point where articulator velocity first exceeds 20% of maximum closure velocity (PEL). GOFFS is the point where articulator velocity has fallen below 20% of maximum release velocity (PVEL2). BF: likelihood that linear model including effect of Trial Type can account for velocity differences at each landmark, compared to null model. MAXC landmark not shown because articulator velocities are zero at the point of maximum constriction in each gesture.

Articulatory RT: Target Vowel

Grand mean RT measured at articulatory onset of the target vowel (g2 GONS) was 622 ms for unprimed trials, 608 ms for 1L trials, and 609 ms for 2L trials. Individual subjects differed in mean articulatory RTs by up to 181 ms: mean latency for M01 was 518 (81) ms, compared to 699 (114) ms for M02. Grand mean RTs measured at each gestural landmark in the Target nucleus are summarized in Table 3.

Articulatory Landmark	0L	1L	1L MOPE	BF 1L-0L	2L	2L MOPE	BF 2L-0L	BF Prime
g2 GONS: gestural onset	621	607	-14	0.5	609	-12	0.05	0.8
g2 PVEL: max. onset velocity	685	672	-13	0.4	673	-12	0.03	0.6
g2 MAXC: max. constriction	771	756	-16	3.8	757	-14	0.28	11.4
g2 PVEL2: max. offset velocity	851	831	-20	4.0	834	-16	0.08	4.6
g2 GOFFS: gestural offset	929	910	-19	0.6	918	-11	0.01	0.3
Duration of Vowel Gesture (GOFFS – GONS)	**308**	**303**			**309**			

Table 3: Grand mean articulatory RTs and MOPEs (ms) for each trial condition, measured at each gestural landmark on g2: target vowel. BF: likelihood that model including effect of each Trial Type can account for RT differences, compared to null model. BF Prime: likelihood that model including overall effect of Priming (1L or 2L) can account for RT differences at each landmark, compared to null model.

At each gestural landmark in the vowel, linear models were constructed of RT as a function of Trial Type with random intercepts and random slopes for Subject and random intercepts for Target. The LMEs estimated MOPEs of at least –11 ms at all five landmarks on the vowel gesture, consistent with the differences in grand mean RTs. For the null hypothesis, a linear model of RT was constructed with effects of Subject and Target alone. BFs for model comparisons at each landmark are listed in Table 3 for RT differences between the unprimed and 1L overlap conditions (BF 0L-1L) and between the unprimed and 2L overlap conditions (BF 0L-2L).

At all timepoints, BF(2L-0L) < 1/3 (alternative/null), offering support for the null hypothesis that 2L priming does not affect RT measured on the target vowel gesture. At the beginning (g2 GONS), peak onset velocity (g2 PVEL) and end (g2 GOFFS) of the vowel gesture, BF(1L-0L) and BF Prime both fall between 1/3 and 3, indicating that these data are inconclusive concerning the effect of 1L priming on RT measured at these points in the target. However, at the articulatory target of the vowel (g2 MAXC), and at the point of maximum release velocity of the vowel gesture (g2 PVEL2), BF > 3, offering support for the alternative hypothesis that 1L priming affects target RT measured at these two key articulatory landmarks in the vowel.

Articulatory RT: Coda Consonant

Grand mean RTs measured at each gestural landmark in the Target coda consonant (C2) are summarized in Table 4.

Articulatory Landmark	0L	1L	1L MOPE	BF 1L-0L	2L	2L MOPE	BF 2L-0L	BF Prime
g3 GONS: gestural onset	772	756	-16	0.5	753	-20	2.6	7.6
g3 PVEL: max. onset velocity	826	808	-18	1.2	806	-20	1.6	9.9
g3 MAXC: max. constriction	962	940	-22	0.3	934	-28	0.6	2.3
g3 PVEL2: max. offset velocity	1,065	1,047	-18	0.0	1,040	-25	0.0	0.0
g3 GOFFS: gestural offset	1,156	1,139	-17	0.0	1,129	-27	0.0	0.0
Duration of Coda Consonant Gesture (GOFFS – GONS)	**383**	**383**			**376**			

Table 4: Grand mean articulatory RTs and MOPEs (ms) for each trial condition, measured at each gestural landmark on g3: target coda consonant C2. BF: likelihood that model including effect of each Trial Type can account for RT differences, compared to null model. BF Prime: likelihood that model including overall effect of Priming (1L or 2L) can account for RT differences at each landmark, compared to null model.

At each gestural landmark in the coda, linear models were constructed of RT as a function of Trial Type with random intercepts and random slopes for Subject and random intercepts for Target. The LMEs estimated MOPEs of at least –17 ms at all five landmarks on the coda consonant gesture, consistent with the differences in grand mean RTs. For the null hypothesis, a linear model of RT was constructed with effects of Subject and Target alone. BFs for model comparisons at each landmark are listed in Table 4 for RT differences between the unprimed and 1L overlap conditions (BF 0L-1L) and between the unprimed and 2L overlap conditions (BF 0L-2L).

At the beginning (g3 GONS) and at peak closure velocity (g3 PVEL) of the coda consonant, BFs for 1L and 2L conditions are inconclusive, but BF Prime > 3, offering evidence that overall the effect of either type of priming reduces RT measured at the beginning of the gesture associated with the coda consonant in the target utterance. These data are inconclusive about priming effects measured at the point in time when the coda consonant reaches its target (g3 MAXC). As the coda consonant is released (g3 PVEL2), and at the end of the utterance (g3 GOFFS), all BFs < 1/3 (alternative/null), offering support for the null hypothesis that neither type of priming affects RT measured on the final part of the target coda consonant gesture.

Response Duration

Duration of spoken responses was measured between the utterance limits illustrated in Figure 3. Acoustic durations averaged over individual item means did not differ between 0L and 1L trial conditions ($BF_{10} = 0.14$); nor between 0L and 2L trial conditions ($BF_{10} = 0.14$). These data provide very strong evidence that there was no effect of priming on the acoustic duration of responses in this experiment.

Articulatory duration was measured from the gestural onset of the onset consonant (g1 GONS) to the gestural offset of the coda consonant (g1 GOFFS) in each target. Mean articulatory duration was 740.9 (192) ms for unprimed (0L) trials, 746.7.2 (198) ms for 1L trials, and 740.2 (182) ms for 2L trials. Articulatory durations averaged over individual item means did not differ between 0L and 1L trial conditions ($BF_{10} = 0.12$); nor between 0L and 2L trial conditions ($BF_{10} = 0.15$). These data provide very strong evidence that there was no effect of priming on the articulatory duration of responses in this experiment.

Effects of priming on the articulatory response: summary

Table 5 summarizes the results of our statistical analyses for 1st-letter priming condition (1L) vs. no-overlap condition (0L).

Gestural landmark	Initial consonant		Vowel		Final consonant	
Gestural Onset (GONS)	Begins earlier	23 ms	Begins earlier	14 ms	Begins earlier	16 ms
Max Closure Velocity (PVEL)	Reached earlier	22 ms	Reached earlier	13 ms	Reached earlier	18 ms
Gestural Target (MAXC)	Reached earlier	18 ms	Reached earlier	16 ms	Reached earlier	22 ms
Max. Release Velocity (PVEL2)	Reached earlier	14 ms	Reached earlier	20 ms	Reached earlier	18 ms
Gestural Offset (GOFFS)	Ends earlier	15 ms	Ends earlier	19 ms	Ends earlier	17 ms
Duration (GONS - GOFFS)	No effect		No effect		No effect	
Velocity at GONS	Greater .16 cm/sec					
Velocity at PVEL	No effect					
Velocity at PVEL2	No effect					
Velocity at GOFFS	No effect					

Table 5: Effects of Priming on Gestural Timing and Velocity: 1st-letter priming condition vs. no-overlap condition. Light grey entries: BF analysis was inconclusive about whether or not there was a priming effect.

Table 6 summarizes the results of our statistical analyses for 1st- and 2nd-letter priming conditions (1L + 2L) vs. no-overlap condition (0L).

Gestural landmark	Initial consonant		Vowel		Final consonant	
Gestural Onset (GONS)	Begins earlier	23 ms	Begins earlier	12 ms	Begins earlier	20 ms
Max Closure Velocity (PVEL)	Reached earlier	23 ms	Reached earlier	12 ms	Reached earlier	20 ms
Gestural Target (MAXC)	Reached earlier	20 ms	Reached earlier	14 ms	Reached earlier	28 ms
Max. Release Velocity (PVEL2)	Reached earlier	18 ms	Reached earlier	16 ms	Reached earlier	25 ms
Gestural Offset (GOFFS)	Ends earlier	20 ms	Ends earlier	11 ms	Ends earlier	27 ms
Duration (GONS - GOFFS)	No effect		No effect		No effect	
Velocity at GONS	Greater .16 cm/sec					
Velocity at PVEL	Greater .39 cm/sec					
Velocity at PVEL2	No effect					
Velocity at GOFFS	No effect					

Table 6: Effects of Priming on Gestural Timing and Velocity: 1st- and 2nd letter priming condition vs. no-overlap condition. Light grey entries: BF analysis was inconclusive about whether or not there was a priming effect.

Two conclusions from Table 5 and Table 6 are clear:

(a) For any measure which reflects properties internal to the response (such as the duration of the initial consonant, the vowel, or the final consonant, or the velocity of any articulator movement, or the total duration of the articulatory response), there is no evidence that either type of priming had any effect. In these data therefore, priming does not significantly affect the organization or coordination of the articulatory response.

(b) For any articulatory measure of response, the gestural landmark always occurs earlier in the primed conditions than the corresponding landmark in the unprimed condition. This is so for landmarks in the onset consonant, vowel, and coda consonant. The number of milliseconds by which priming affects any part of the spoken response is very similar across all landmarks in all parts of the utterance, although in some cases (light gray cells) the evidence for a priming effect is inconclusive.

In sum, the effect of masked onset priming in these data is to shift the entire articulatory response forward in time, without having any effect on any properties internal to that response.

Discussion

These findings are consistent with the explanation of the MOPE proposed by Mousikou et al. (2010a) using the DRC computational model of reading. Under this account, the only effect of the MOPE is a wholesale temporal shifting of the entire articulatory response, with no effects of priming on any properties internal to that response, which is the pattern of results we observed in this study.

Given this, what are the implications of these findings for the three-stage model of speech production outlined at the beginning of this paper? Whereabouts in this sequence of stages might the MOPE arise?

The MOPE could arise at Stage 1, where an abstract phonological representation of the target is created via grapheme-phoneme correspondence (GPC) translation, if this process begins earlier in the primed conditions. The reason this may happen is that at the stage at which GPC translation of the target nonword is about to begin, its initial phoneme has already been partially activated by the nonword prime (on priming trials), conferring an advantage. In non-primed trials, on the other hand, some conflicting initial phoneme has been partially activated by the prime, conferring a disadvantage.

Suppose this were not the case, so that the timing of the creation of the abstract phonological representation of the target via GPC translation from print was unaffected by priming. In this case, could priming work by influencing Stage 2 instead? This would mean that the computation of an articulatory plan from the abstract phonological representation – despite beginning at the same time in primed and unprimed conditions – would be completed earlier in the primed condition. An assumption of the model is that Stage 2 does not deliver anything to Stage 3 until Stage 2 has completed its job. Why in this case would Stage 2 work faster in the primed than in the unprimed condition, even though Stage 1 did not? We see no mechanism by which this might occur in this model. Nor is there any mechanism by which, if priming affects neither Stage 1 nor stage 2, it could affect Stage 3. This is because of the ballistic nature of Stage 3: it requires delivery of a fully-specified articulatory plan from Stage 2, and executes that plan ballistically as soon as it is delivered (Rastle et al., 2000). In this framework, there is no provision for factors being able to affect the speed of this ballistic execution. We conclude that the three-stage model offers a natural explanation of the MOPE as arising at the first stage

of this model of speech production.

Alternative mechanisms for some of these stages of speech production are available in other models. Two key principles of the Levelt and DRC models are (i) the requirement that a complete articulatory plan for the syllable must be finalized before execution; and (ii) the assumption that articulation is ballistic. Both of these principles have been challenged by findings from other reading aloud studies and other work (see Krauss & Kawamoto, 2020, for an overview), which has important implications for the findings reported here.

If the minimal planning unit for speech is sub-syllabic (Kawamoto et al., 2014; Schiller & Costa, 2006), then if the initial segment of the target is primed, there is no requirement to wait at Stage 2 for additional articulatory information about the rest of the syllable before passing the segmental plan to Stage 3 for execution. In this case, in the 1L condition, we might expect to see earlier initiation of gestures associated with the target onset consonant with respect to the nuclear vowel gesture, similar to the duration differences reported by Damian (2003). Yet we observed no significant stretching of the initial gesture in these data: in 1L-primed targets, and in 1L trials, both the initial consonant and the following vowel reached their targets earlier, by approximately the same amount of time (14 to 20 ms). These data are inconsistent with a model in which the first segment of the target is articulated independently of the following vowel.

Another factor which should be considered when assessing the results of this study is the nature of the task. Consistent with the original study (Mousikou et al. 2010), target words were removed from the screen once a response was detected. This may have incentivized participants to delay their articulation until the whole word was known, biasing participants towards producing fully planned responses. As Sternberg et al. (1978), Whalen (1990), and Liu et al. (2018) have shown, individual speakers can differ in the way that they respond to similar production tasks, and different levels of speech planning may be evident in a spoken response depending on the specific strategy used and the type of information available to the speaker at different points in time. It is therefore possible that if participants in this study had been forced to respond earlier, differences may have been observed in the articulation of the initial segment depending on the type of priming.

Furthermore, even if responses were planned incrementally, it is possible that the nature of the task results in some responses that are indistinguishable from more globally planned utterances. In the original work using this paradigm, Forster & Davis (1991) argued that articulation may be delayed in the case of inconsistent primes because participants are required to suppress the original speech plan. Yet if this were the only mechanism differentiating the earlier responses to the matched primes, and if these utterances were planned incrementally, we might expect to see earlier articulation of the initial segment with respect to the rest of the word in the 1L condition trials, which we did not observe.

Turning to the assumption that articulation is ballistic, recent work has shown that aspects of speech production can be initiated and arrested independently of the speech plan associated with larger phonological structures such as the syllable. Evidence from delayed naming (Kawamoto, Liu, Mura, & Sanchez, 2008), and priming experiments (Fink, Oppenheim & Goldrick, 2018) reveals that when speakers have early access to information about part of an upcoming utterance, they may initiate articulation of some components, even while delaying the acoustic realization, which typically requires coordination of the anticipated component with other gestures. Tilsen and Goldstein (2012) and Tilsen et al. (2016) have demonstrated that speakers are able to control the timing of individual speech gestures independently of the segments they are associated with.

In this study, early information is available to the speaker about the initial consonant of the target in both 1L and 2L trials. In these cases, if the initial consonant is a nasal, liquid, or fricative /m, n, l, ɹ, f, v, s, z/, we might expect to see earlier articulatory and acoustic initiation of the target, resulting in longer articulatory and acoustic durations of the entire target through elongation of the initial consonant. If the initial consonant is a stop or affricate /p, b, t, d, dʒ, k, g/, we might expect to see earlier initiation of articulation, resulting in a greater articulatory duration and greater acoustic lag in the primed utterances. Yet we found no evidence for greater acoustic or articulatory duration of primed targets (Table 6), nor for greater acoustic-articulatory lag (p. 21) in primed targets compared to unprimed utterances. Further investigations are needed to examine evidence for anticipatory articulation of individual gestures in this paradigm, using in a more extensive and phonetically balanced dataset (see below).

Future Directions

These data raise many questions for further exploration. While it was important in this study to reuse the experimental materials and methods from Mousikou et al. (2010) for comparison of acoustic analyses, these materials are limited in some important ways, as the primes and targets were not designed as a phonetically-balanced set. For example, only two targets in this corpus begin with voiceless labial stops ('PEZ', 'PIV'), but ten targets are /z/-initial. 'J'-initial targets (there are eight in this experimental corpus), are typically realized with a post-alveolar affricate /dʒ/, which although also articulated with the front of the tongue, are not directly comparable with the other coronals in the corpus (/t-d-n-s-z-l/) because they have a more retracted place of articulation. Stimuli designed specifically for articulatory analysis should also exclude /h/-initial targets, as there is no way to directly measure glottal activity using EMA.

More fine-grained analyses of articulatory planning and coarticulation would require greater numbers of tokens with unambiguous orthographic mappings onto mono-gestural segments with more clearly differentiated goals of production. The nuclei of these targets, for example were realized as six different vowel qualities /ɪ-e-æ-ɐ-ɔ-ʊ/, not all of which contrast sufficiently in height, backness, or rounding to allow for robust analysis of potential coarticulatory influences on or from surrounding segments.

More research is required to examine potential influences of priming on articulator velocity. Davis et al. (2015) found that peak lower lip velocity in labial-initial targets and peak tongue tip velocity in coronal-initial targets were both significantly greater in control (unprimed) trials. These findings are consistent with an account in which an orthographically-overlapping prime assists the motor planning of the initial consonant of the target. In the 1L and 2L conditions, if the speaker has commenced articulation of the initial gesture of the prime, the articulators will continue to move towards the same constriction goal in the target. In the 0L trials, the speaker will have prepared or commenced an alternative set of gestural goals (that of the masked prime), and the articulators will have to accelerate to execute the new goals of production corresponding to the target, which does not share the same set of gestures. No significant differences in articulator velocities were found in this experiment, but the velocity data pattern in the same way across all landmarks as reported by Davis et al (2015). Articulator velocities are influenced by a variety of factors uncontrolled in either study, and additional data obtained from more constrained data sets will be required to explore these ideas in further detail.

Closer articulatory analysis would also be important to examine evidence for potential effects arising from competing phonological representations. In speech production models such as those proposed by Goldrick and Blumstein (2006), the output forms of segments are influenced by "cascading activation from the speech plan" (McMillan & Corley 2010: 246). If differences in the position and velocity profile of articulators were observed for primed vs. unprimed target forms, this would be consistent with cascaded models. Further analysis with larger, more phonetically balanced datasets will be required to examine the articulatory properties of primed utterances to assess the relevance of these models in the MOPE paradigm.

Open Practices Statement

The data and materials for all experiments are available at: http://mproctor.net/data/mope

References

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2): 12-28.

Baguley, T., & Kaye, W. (2010). Review of: Understanding psychology as a science: an introduction to scientific and statistical inference, by Z. Dienes. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 695–69

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H. & Grothendieck, G. (2014). Package 'lme4'. R foundation for statistical computing, Vienna, 12.

Berent, I., & Perfetti, C. A. (1995). A rose is a REEZ: The two-cycles model of phonological assembly in reading English. *Psychological Review*, 102, 146–184.

Bowers, J. S., Vigliocco, G., & Haan, R. (1998). Orthographic, phonological and articulatory contributions to masked letter and word priming. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1705–1719.

Browman, C. P., & Goldstein, L. M. (1989). Articulatory gestures as phonological units. *Phonology* 6, 201–251.

Browman, C. P., & Goldstein, L. M. (1992). Articulatory phonology: an overview. *Phonetica*, *49*(3–4), 155–180.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1(1).

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24, 209-244.

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*, 204-256.

Cortese, M. J. (1998). Revisiting serial position effects in reading. *Journal of Memory & Languange*, 39, 652–665.

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(3), 416.

Davis, C.; Proctor, M.; Shaw, J.; Derrick, D. & Kroos, C. (2013). Using electromagnetic articulography to examine masked onset priming. *Proc. 40th Australasian Experimental Psychology Conference*, 3-6 April, Adelaide, 16.

Davis, C.; Shaw, J.; Sherwood, S.; Proctor, M.; Kim, J. & Derrick, D. (2015). Examining speech production using masked priming. *Proc. Int'l Congress on Phonetic Sciences*, 10-14 Aug., Glasgow.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psycholology*, 5, 781.

Dienes, Z., & McLatchie, N. (2018). Four reasons to prefer Bayesian over significance testing. *Psychonomic Bulletin & Review*, 25, 207-218.

Drake, E., & Corley, M. (2015). Effects in production of word pre-activation during listening: Are listener-generated predictions specified at a speech-sound level? *Memory & cognition*, 43(1), 111-120.

Fink, A., Oppenheim, G. M., & Goldrick, M. (2018). Interactions between lexical access and articulation. *Language, cognition and neuroscience*, 33(1), 12-24.

Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language*, 30, 1–25.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers*, 35, 116–124.

Fowler, C. A. & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, 36, 171–195.

Garcia, D. (2010). Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics & Data Analysis*, 54(4), 1167-1178.

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.

Goldrick, M., & Chu, K. (2014). Gradient co-activation and speech error articulation: Comment on Pouplier and Goldstein (2010). *Language, Cognition and Neuroscience*, 29(4), 452-458.

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. Schiller & A. Meyer (eds.) *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159–207). Berlin: Mouton.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, 103(3), 386-412.

Hawkins S, & Nguyen N. (2004). Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. *Journal of Phonetics*, 32(2), 199–231.

Hertrich, I. & Ackermann, H. (1995). Coarticulation in Slow Speech: Durational and Spectral Analysis. *Language and Speech*, 38, 159-187.

Holbrook, B. B., Kawamoto, A. H., & Liu, Q. (2019). Task demands and segment priming effects in the naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(5), 807-821.

Hoole, P., Mooshammer, C. & Tillmann, H. G. (1994). Kinematic analysis of vowel production in German. *3rd Intl. Conf. on Spoken Language Processing (ICSLP 94)*, 53–56.

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

Kawamoto, A. H., Liu, Q., Mura, K., & Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language*, 58(2), 347–365.

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. *Quarterly Journal of Experimental Psychology*, 67(12), 2340-2359.

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language*, 47, 145–171.

Kinoshita, S. (2000). The left-to-right nature of the masked onset priming effect in naming. *Psychonomic Bulletin & Review, 7*, 133–141.

Krause, P. A., & Kawamoto, A. H. (2019). Anticipatory mechanisms influence articulation in the form preparation task. *Journal of Experimental Psychology: Human Perception and Performance*, 45(3), 319.

Krause, P. A., & Kawamoto, A. H. (2020). On the timing and coordination of articulatory movements: *Historical perspectives and current theoretical challenges. Language and Linguistics Compass*, 14(6), e12373.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25(1), 178–206. Springer.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82(13).

Ladefoged, P., Silverstein, R., & Papçun, G. (1973). Letter: Interruptibility of speech. *Journal of the Acoustical Society of America*, 54(4)

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation.* Cambridge, MA: MIT Press

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review*, 74(6), 431.

Liu, Qiang, Kawamoto, Alan, Payne, Katherine & Dorsey, Genvieve. (2018). Anticipatory Coarticulation and the Minimal Planning Unit of Speech. *Journal of Experimental Psychology: Human Perception & Performance*, 44, 139-153.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford: Oxford University Press.

Lukatela, G., & Turvey, M. T. (1994). Visual lexical access is initially phonological: 2. Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General*, 123, 331–353.

McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, 117(3), 243-260.

Meyer, A. S., Belke, E., Häcker, C., & Mortensen, L. (2007). Use of word length information in utterance planning. *Journal of Memory and Language*, 57(2), 210-231.

Monsell, S. (1986). Programming of complex sequences: Evidence from the timing of rapid speech and other productions. In H. Heuer & C. Fromm (Eds.), *Generation and modulation of action patterns* (pp. 72-86). Heidelberg, Germany: Springer-Verlag

Morey, R. D., & Rouder, J. N. (2018). BayesFactor Package (version 0.9.12-4.2).

Mousikou, P., Coltheart, M., Finkbeiner, M. & Saunders, S. (2010a). Can the DRC computational model of reading offer a valid account of the masked onset priming effect? *Quarterly Journal of Experimental Psychology*, 63, 984-1003.

Mousikou, B., Coltheart, M., & Saunders, S. (2010b). Computational modelling of the masked onset priming effect in reading aloud. *European Journal of Cognitive Psychology*, 22, 725-763.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, 47(5), 1048-1058.

Palo, P., Schaeffler, S. & Scobbie, J. M. (2014). Pre-speech tongue movements recorded with ultrasound. *Proc. 10th Intl. Seminar on Speech Production (ISSP)*, 5–8 May 2014, Cologne, 300-303.

Perkell, J. S.; Cohen, M. H.; Svirsky, M. A.; Matthies, M. L.; Garabieta, I. & Jackson, M. T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *The Journal of the Acoustical Society of America*, 92, 3078-3096.

Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, 39, 859 – 862.

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

Rastle, K., Croot, K., Harrington, J. & 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, 1083-1095.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology*, 55A, 1339–1362.

Rastle, K., Harrington, J.M., Coltheart, M. and Palethorpe, S. (2000). Reading aloud begins when the computation of phonology finishes. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1178-1191.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, 64(3), 249-284.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2), 225–237.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological psychology*, 1(4), 333-382.

Schiller, N. O. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language*, 39, 484–507.

Schiller, N. O. (2000). Single word production in English: The role of subsyllabic units during phonological encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 512–528.

Schiller, N. O., & Costa, A. (2006). Activation of segments, not syllables, during phonological encoding in speech production. *The Mental Lexicon*, 1(2), 231-250.

Schiller, N. O. (2007). Phonology and orthography in reading aloud. *Psychonomic Bulletin & Review*, 14(3), 460-465.

Singh, A. (2018). Online Bayes Factor (Dienes) Calculator. Retrieved from https://medstats.github.io/bayesfactor.html

Sternberg, S., Knoll, R. L., Monsell, S., & Wright, C. E. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica*, 45, 175-197.

Sternberg, S., Monsell, S., Knoll, R., & Wright, C. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117-152). New York: Academic Press.

Sulpizio, S., & Burani, C. (2015). Article Commentary: Reading segments is not reading words: Comment on Kawamoto et al. (2014). *Quarterly Journal of Experimental Psychology*, 68(3), 619-621.

Sulpizio, S., Spinelli, G., & Burani, C. (2015). Stress affects articulatory planning in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 41(2), 453.

Tiede, M., Bundgaard-Nielsen, R., Kroos, C., Gibert, G., Attina, V., ... & Best, C. (2010). Speech articulator movements recorded from facing talkers using two electromagnetic articulometer systems simultaneously. *Proc. Meetings on Acoustics*, 11(1): 060007.

Tilsen, S., & Goldstein, L. (2012). Articulatory gestures are individually selected in production. *Journal of Phonetics*, 40(6), 764–779.

Tilsen, S., Spincemaille, P., Xu, B., Doerschuk, P., Luh, W. M., Feldman, E., & Wang, Y. (2016). Anticipatory posturing of the vocal tract reveals dissociation of speech movement plans from linguistic units. *PloS one*, 11(1).

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of phonetics*, 71, 147–161. Elsevier.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 735.

Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, 18(1), 3-35.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, 58(3), 475-482.

Wright, R. (2004). Factors of lexical competition in vowel articulation. In Local, J., Ogden, R., & Temple, R. (Eds.), *Papers in Laboratory Phonology VI* (pp. 26–50). Cambridge: Cambridge University Press.