ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Insights Into the Effect of General Attentional State, Coarticulation, and Primed Speech Rate in Phoneme Production Time

Montse Soberanes,[a] Carlos A. Pérez-Ramírez,[b] and M. Florencia Assaneo[a]

[a] Institute of Neurobiology, National Autonomous University of Mexico, Juriquilla, Santiago de Querétaro, México [b] Autonomous University of Querétaro, Santiago de Querétaro, México

ARTICLE INFO

Article History:
Received August 27, 2024
Revision received November 28, 2024
Accepted January 5, 2025

Editor-in-Chief: Cara E. Stepp
Editor: Ayoub Daliri

https://doi.org/10.1044/2025_JSLHR-24-00595

ABSTRACT

Purpose: This study aimed to identify how a set of predefined factors modulates phoneme articulation time within a speaker.
Method: We used a custom in-lab system that records lip muscle activity through electromyography signals, aligned with the produced speech, to measure phoneme articulation time. Twenty Spanish-speaking participants (12 females) were evaluated while producing sequences of a consonant–vowel syllable, with each sequence consisting of repeated articulations of either /pa/ or /pu/. Before starting the sequences, participants underwent a priming step with either a fast or slow speech rate. Additionally, the general attentional state level was assessed at the beginning, middle, and end of the protocol. To analyze the variability in the duration of /p/ and vowel articulation, we fitted individual linear mixed-models considering three factors: general attentional state level, priming rate, and coarticulation effects (for /p/, i.e., followed by /a/ or /u/) or phoneme identity (for vowels, i.e., being /a/ or /u/).
Results: We found that the level of general attentional state positively correlated with production time for both the consonant /p/ and the vowels. Additionally, /p/ production was influenced by the nature of the following vowel (i.e., coarticulation effects), while vowel production time was affected by the primed speech rate.
Conclusions: Phoneme duration appears to be influenced by both stable, speaker-specific characteristics (idiosyncratic traits) and internal, state-dependent factors related to the speaker's condition at the time of speech production. While some factors affect both consonants and vowels, others specifically modify only one of these types.
Supplemental Material: https://doi.org/10.23641/asha.28608428

Speech, one of the most distinctive and uniquely human abilities, consists of sound segments that combine to form the words we use to communicate. These sound segments, known as phonemes, are the fundamental building blocks of all languages. Phonemes vary across languages and require the coordination of different structures within the vocal tract—such as the larynx, velum, tongue, lips, and jaw—for their production (Stevens, 2000). In general, phonemes can be divided into two main classes: vowels and consonants. Consonants require a constriction somewhere in the vocal tract for their production, whereas vowels do not. These two classes have been described in terms of their distinct characteristics in both the acoustic and articulatory domains (Browman & Goldstein, 1989; Hualde, 2013; Stevens, 2000), which has significantly contributed to the general understanding of speech production.

Speech production has been characterized as a spatiotemporal phenomenon, as it can be viewed as a dynamic process defined by both spatial and temporal features (Browman & Goldstein, 1992). A substantial body of research has focused on describing the spatial aspects of speech production, addressing both general and highly

Correspondence to M. Florencia Assaneo: fassaneo@inb.unam.mx.
Disclosure: *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

specific features. For example, studies have shown that speech production is not governed by a fixed kinematic pattern for phonemes but rather is guided by simple motor goals (Abbs et al., 1984). At the same time, other research has examined more detailed aspects, such as the subtle differences in vocal tract configuration that distinguish the same phoneme across different dialects (Jacewicz et al., 2009; Wieling et al., 2016). These examples illustrate the diverse range of research in this field. Curiously, the temporal aspects of speech production have not been as thoroughly described. Nevertheless, it is known that phonemes exhibit characteristic temporal features. For example, voice onset time (i.e., the time between stop consonant release and voicing onset) has been shown to play a role in phoneme identification (Abramson & Whalen, 2017), and phoneme duration has been found to influence perception (Celdrán, 1993). Furthermore, phoneme duration exhibits significant variability, even within the same language. In Spanish, for instance, phonemes have been reported to range from 30 ms to more than 200 ms (Del Barrio Estévez & Torner, 1999; Fernández, 2007; Krohn, 2019; Marín Gálvez, 1995; Mendoza et al., 2003).

Several factors have been shown to modulate phoneme articulation length, including age (Walker et al., 1992), gender, accent, phoneme position within the word, and phonetic class (Van Heerden & Barnard, 2008). While these studies have described demographic and linguistic features that influence articulatory duration, the variability within subjects and across utterances remains underexplored. There are several key factors that may modulate phoneme articulation within a subject. For example, motor skills have been shown to be influenced by attentional state (Song, 2019). Since speech is a sequence of motor gestures, we might expect a participant's attentional state to affect their articulatory features and, consequently, the timing of articulation. Another important factor is coarticulation (Daniloff & Hammarberg, 1973). It is well established that the articulation of one speech segment influences the configuration of surrounding segments (Assaneo, Ramirez Butavand, et al., 2019; Parker, 1974; Whalen, 1990). Therefore, it seems reasonable to hypothesize that the duration of a phoneme will be modulated by the identity of its neighboring segments.

Last, it is evident that we can volitionally modify our speech rate. However, speakers have a natural speaking rate (Hirose & Kawanami, 2002), and altering this rate comes at a cost, namely, a reduction in the precision with which a phoneme is articulated (Acheson & Hagoort, 2014; Fossett et al., 2016). In this study, we focused on the environmental factors that can temporarily modify phoneme articulation time while a participant maintains their natural speech rate. Previous research suggests that the motor brain areas responsible for speech production exhibit an oscillatory nature (Poeppel & Assaneo, 2020). In this context, a speaker's natural rate may emerge from the internal natural frequency of these speech motor areas, which determines syllabic duration. Additionally, research on motor areas more broadly indicates that their internal frequency can be temporarily altered through synchronization protocols (Bose et al., 2019). Building on these observations, we aim to explore the extent to which forcing a speaker to match a given fast or slow syllabic rate may have long-lasting effects on their natural speech articulation timing, particularly in terms of the precision of phoneme articulation.

In this study, we investigated the variability in articulatory duration of consonants and vowels within participants and explored how a set of factors might influence this duration. Specifically, we examined how general attentional state (Dromey & Shim, 2008), coarticulation effects from neighboring phonemes (Daniloff & Hammarberg, 1973), and primed speech rate impact the production of consonant–vowel (CV) syllables containing the voiceless occlusive consonant (/p/), the rounded vowel /a/, and the unrounded vowel /u/. This protocol was motivated by an exploratory procedure that revealed significant variability in articulation time even when the same phoneme was produced by the same participant.

Method

Participants

Participants aged 18–40 years were recruited for this study through announcements on social media and posters on bulletin boards. All participants were native Spanish speakers and answered "no" to the following screening questions: (a) Have you been diagnosed with any hearing problems? (b) Do you have a diagnosis of any neurological or psychiatric conditions? The study involved two cohorts. A small exploratory cohort consisted of three participants (all women; M_{age} = 29.3 years, SD = ±8.5), who completed the exploratory procedure. A separate cohort of 29 participants (17 women; M_{age} = 25.5 years, SD = ±5) participated in the main procedure. Details of the exploratory and main procedures are provided below.

The study was approved by the Ethics Committee of the Institute of Neurobiology at the National Autonomous University of Mexico under Protocol 096. All participants provided informed consent and received financial compensation upon completion of the experiment.

Experimental Setup

We designed an experimental setup to simultaneously record the electrical signals generated by the lip muscles during speech production and the corresponding speech sounds. These recorded electrical and acoustic

signals were then combined to estimate the articulation time of the produced phonemes (see the Estimation of Phoneme Articulation Times section). In the following paragraphs, we describe how the components of the setup were connected and how it functions.

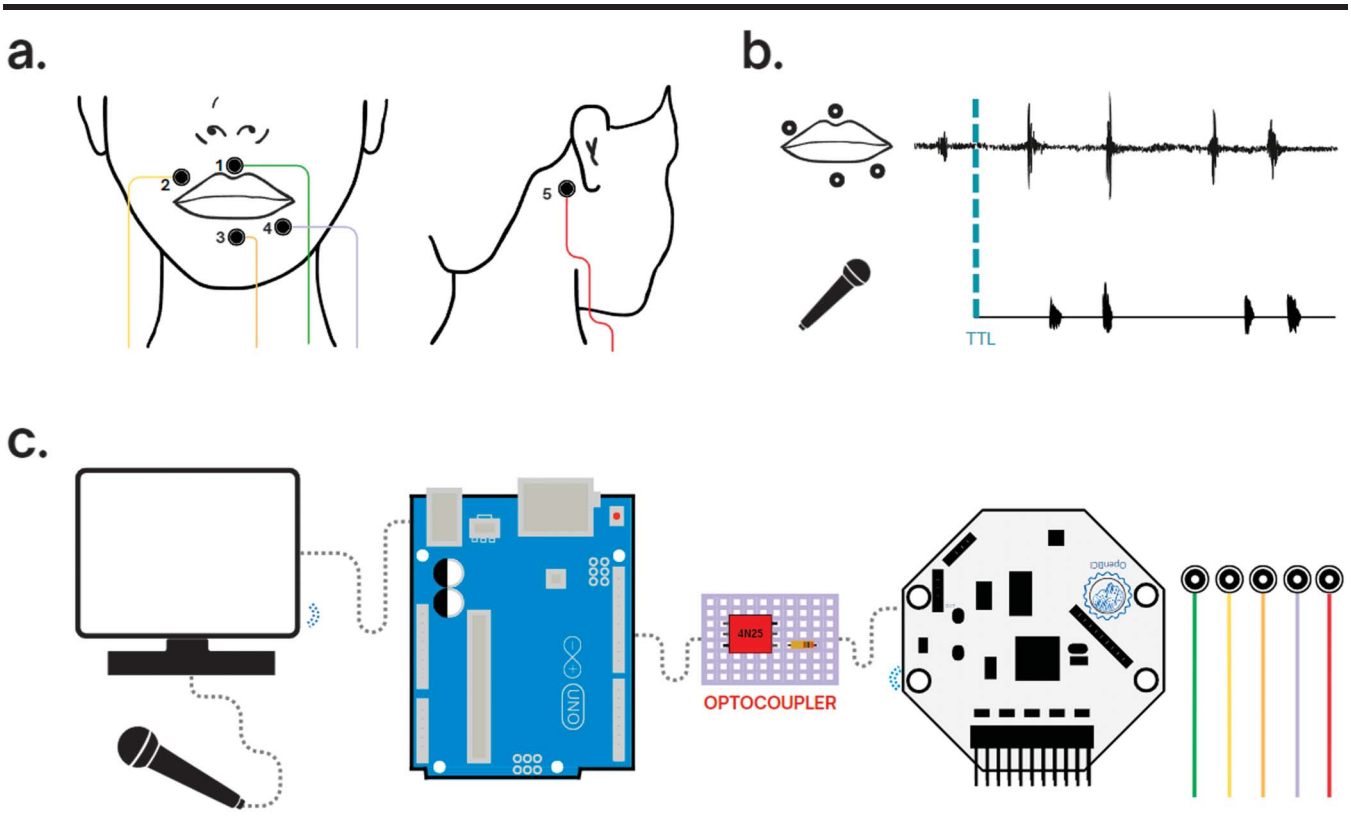
The entire study took place in a soundproof booth, which had a window that allowed the experimenter to see the participants. Participants were given ER-1 Insert earphones, seated 20 cm away from a microphone (Shure PGA48-XLR connected to the computer via a Focusrite Scarlett 2i2 Studio interface) to capture produced speech and in front of a computer (iMAC 3.2 GHz 6-Core Intel Core i7) for stimuli presentation and to register the audio signal, with a keyboard positioned at their dominant hand allowing response capture.

We used the neural interface Cyton Board by OpenBCI, an eight-channel biosignal recorder to register electromyographic (EMG) signal from the superior and inferior portions of the orbicularis oris muscle via a bipolar configuration: Two pairs of electrodes were placed on the upper and lower lips, respectively, with a reference electrode placed at the right mastoid (see Figure 1a).

OpenBCI Guided User Interface was used to record EMG signals from the upper and lower lip and the PsychPortAudio function from MATLAB PsychToolbox (Kleiner et al., 2007) was used to register the speech audio signals. To align the muscle activity recordings to the corresponding produced speech signals, we coupled the OpenBCI to the mic recording through an Arduino UNO. In further detail, our MATLAB script sent a TTL trigger to the Cyton through the Arduino board when the microphone was opened to start audio capture (see Figure 1b). Additionally, to match the voltage output/input between elements, we used a 4N25 optocoupler (see Figure 1c). This configuration ensured our muscle signal and our audio recording were properly aligned.

Participants were instructed not to apply any facial products on the day of the experiment, and men were asked to shave their beard and mustache to minimize signal noise. The area behind their right ear, right above the

Figure 1. Experimental setup. (a) Electrode placement areas. 1. Over the participants' philtrum. 2. Between the philtrum and the right oral commissure. 3. Under the lower lip, following the line of the philtrum. 4. Under the lower lip, between the philtrum and the left oral commissure. 5. Behind the ear, above the temporal bone. (b) Recordings. During electromyographic recording, a TTL trigger was sent to the Cyton board upon the beginning of audio capture to align the muscle and audio signals. (c) Hardware components. The microphone and the Arduino UNO board were directly connected to the computer, while the Cyton board was connected via a USB dongle. The Cyton and Arduino boards were connected by a small circuit that included a 4N25 optocoupler (marked in red).



temporal bone, and all skin areas surrounding their lips were cleaned using alcohol and Nuprep skin prep gel. Five gold cup electrodes were prepared with Tens 20 conductive paste before placing them on the participant's skin. All procedures were programmed and presented to participants in MATLAB (Version R2022b, The MathWorks, Inc.) using PsychToolbox.

Experimental Design

We conducted two different experiments, which we refer to as the exploratory procedure and the main procedure. Each procedure consisted of a sequence of blocks, with each block corresponding to a single task. In this section, we first describe each block in detail, and at the end, we explain how the blocks are organized to form each procedure.

Articulation block. Participants were presented with a sequence of tones that cued the production of a syllable. They were instructed to utter the syllable as soon as they heard each tone and to return to a resting position until they heard the next tone. The same CV syllable was produced during each block, while blocks with different CV syllables were used across the different experimental procedures. Before starting the task, the resting position was defined for the participants as lips closed and relaxed. The intertone interval was randomly assigned between 0.75 and 3.6 s. Participants' speech and muscle activity were recorded while they were connected to the experimental setup.

General Attentional State Assessment block: Participants' general attentional state was assessed using a modified version of the Eriksen flanker task (Eriksen & Eriksen, 1974). They completed 40 trials. On each trial, they were presented with a target arrow at the center of the screen, which pointed either to the left (⇦) or to the right (⇨), flanked by nontarget arrows that may or may not point in the same direction as the target. Participants were instructed to press a key on the keyboard corresponding to the direction of the target arrow as quickly as possible while ignoring the nontarget arrows. The intertrial interval (i.e., the time between the button press and the start of the next trial) was randomly set between 0.3 and 1 s. The number of correct trials was divided by the total number of trials (40) and multiplied by 100 to obtain an accuracy score, which was used as an indicator of the participants' general attentional state level.

Speech Rate Priming block: To prime participants' speech rate, they were instructed to continuously repeat the syllable /pe/ in synchrony with a tone that was repeated at a rhythmic pace for 50 s. The tone frequency was set to 1000 Hz, based on the observation by Mares et al. (2023) that participants are more likely to synchronize with this type of stimulus. Each Speech Rate Priming block could belong to one of two conditions: fast or slow. In the slow condition, the tone presentation rate was set to three tones per second (3 Hz), while in the fast condition, it was set to five tones per second (5 Hz).

To ensure that participants were synchronized to the intended rate during each block, the phase-locking value (PLV) between the envelope of the presented tone train and the produced speech signals was calculated using the following formula:

$$PLV = \frac{1}{T}\left|\sum_{t=1}^{T} e^{i(\theta_1(t)-\theta_2(t))}\right| \quad (1)$$

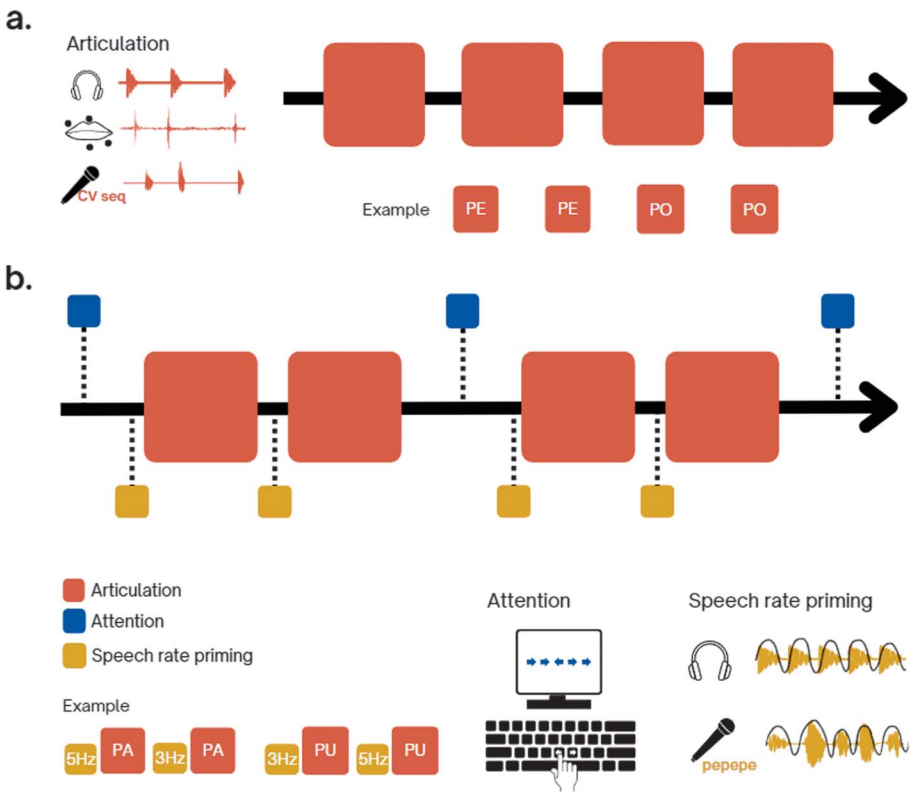
where t is the discretized time, T is the total number of time points, and θ_1 and θ_2 the phase of the envelope of the perceived and produced speech signals, respectively. The PLV was estimated in 5-s windows with a 2-s overlap (Assaneo, Ripollés, et al., 2019). The obtained PLVs were averaged within each block to yield a single synchronization score per block.

Exploratory procedure. An initial study was conducted with a small cohort to explore the variability in the phonemes articulation time. The procedure consisted of a sequence of four Articulation blocks. In two of the blocks, participants were instructed to pronounce the syllable /pe/, and in the other two blocks, they were instructed to pronounce the syllable /po/ (see Figure 2a). Each block included 60 cue tones, resulting in 60 expected utterances of the designated syllable, for a total of 120 utterances per CV sequence (/pe/ and /po/) per participant. The order of the /pe/ and /po/ blocks was randomized across participants. Participants repeated the same procedure in a second session, conducted 1 week later.

Main procedure. The main procedure aimed to assess the influence of three factors (general attentional state, intended speech speed, and coarticulation) on intraspeaker variability in phoneme articulation time. Participants completed four Articulation blocks, four Speech Rate Priming blocks, and three General Attentional State Assessment blocks.

In two of the Articulation blocks, participants were instructed to pronounce the syllable /pa/ and, in the other two, the syllable /pu/. The rounded vowel /u/ and the unrounded vowel /a/ were used to examine the coarticulation effect on /p/ articulation time. Each block contained 120 cue tones, corresponding to 120 utterances of the designated syllable, for a total of 240 utterances per CV sequence (/pa/ and /pu/) per participant. The order of the /pa/ and /pu/ blocks was randomized across participants. To assess the effect of speech speed, each Articulation block was preceded by a Speech Rate Priming block.

Figure 2. Exploratory and main procedure layouts. (a) Exploratory procedure. Participants completed four Articulation blocks, during which lip activity and sound production were recorded as they pronounced one of two consonant–vowel (CV) sequences (/pe/ or /po/) upon receiving an audio cue. The order of the instructed CV sequences was randomized for each participant; an example is provided. (b) Main procedure. Participants completed four Articulation blocks, producing one of two CV sequences (/pa/ or /pu/), each preceded by a Speech Rate Priming block. General Attentional State Assessment blocks were completed at the beginning, middle, and end of the experiment. The order of the /pa/ or /pu/ Articulation blocks and the slow or fast Speech Rate Priming block was randomized for each participant, with the condition that each CV sequence was preceded by both fast and slow priming speech rates; an example is provided.



Articulation blocks for each CV combination were primed once with a slow rate (3 Hz) and once with a fast rate (5 Hz). Finally, to measure the time evolution of the participant's attentional level, General Attentional State Assessment blocks were completed at the beginning, middle, and end of the protocol. The general overview of the main procedure is shown in Figure 2b. Articulation blocks preceded by a Speech Rate Priming block, where the participant's synchronization score did not reach 0.5, were excluded from subsequent analyses.

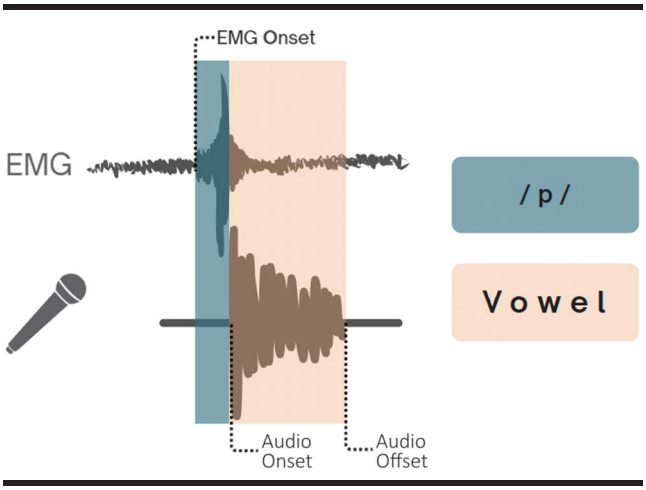
Estimation of Phoneme Articulation Times

The signals collected during each Articulation block, including EMG recordings and audio produced by the participant, were preprocessed for subsequent analysis. The continuous EMG signals, filtered between 10 and 45 Hz, were segmented into 750-ms epochs, with each cue tone marking the onset. Each epoch was z scored, with the first 100 ms defined as the baseline. The onset of lip muscle activity was determined as the first time point at which the absolute value of the EMG signal exceeded a threshold of 4 *SD*s above the baseline.

The audio signal was segmented using the same procedure as the EMG recordings, and the same thresholding method was applied to detect speech onset. The speech offset was defined as the last time point at which the signal exceeded this threshold.

The onsets of lip muscle activity, speech onset, and speech offset were used to calculate the articulation times for each phoneme. For the /p/ phoneme, articulation time was defined as the difference between the speech onset (the burst corresponding to the release of the occlusion, identified from the acoustic signal) and the onset of lip muscle activity (the start of the motor gesture, identified from the EMG). For vowels, the articulation time was determined by the estimated voiced duration obtained from the acoustic signal (see Figure 3). For each subject, durations for vowels and /p/ that exceeded 2 *SD*s from the mean of their respective distributions were considered outliers and were excluded from further analysis.

Figure 3. Strategy to estimate phonemes' articulation time. The onsets and offsets from the electromyography and audio signals were used to reconstruct articulation times of the produced consonant /p/ and vowels. EMG = electromyography.



Data Analysis: Exploratory Procedure

Articulation times for the consonant /p/ and vowels /e/ and /o/ were recorded in a data set that included the CV syllable (/pe/ or /po/), the participant, and the session corresponding to each duration. Unpaired and paired *t* tests, corrected for multiple comparisons using the Bonferroni method, were conducted to explore differences in durations across subjects, phonemes, and sessions.

Data Analysis: Main Procedure

To explore how different factors modulate consonant and vowel articulation time, two linear mixed-effects model analyses were conducted using the lmer (Bates et al., 2015) and buildmer (Voeten & Voeten, 2021) libraries in R —one for /p/ and another for the vowels (/a/ and /u/). In both models, a backward elimination procedure was performed, starting with a model that included priming speed, general attentional state, and vowel (incoming vowel for /p/ and phoneme identity for the vowels) as fixed factors. Priming speed (i.e., slow or fast) and general attentional state were fixed for all syllables within the same Articulation block. Priming speed was determined based on the rate of the Speech Rate Priming block preceding the corresponding Articulation block. General attentional state level was assigned according to the General Attentional State Assessment block closest in time to the Articulation block. Specifically, Articulation Block 1 used the attentional level assessed during the first General Attentional State Assessment block, Articulation Blocks 2 and 3 used the attentional level recorded at the midpoint of the protocol, and Articulation Block 4 used the attentional level measured at the end of the protocol.

Intercepts, but not slopes, were allowed to vary by participant. The models that best explained the durations were selected based on changes in the Bayesian information criterion. To confirm the significance of the predictor variables in the best models, we performed a deviation analysis using Wald-type III chi-square tests. Additionally, estimated marginal means and trends were computed using the emmeans R package (Lenth, 2024). All reported *p* values were Bonferroni-corrected for multiple comparisons.

Results

Exploratory Procedure

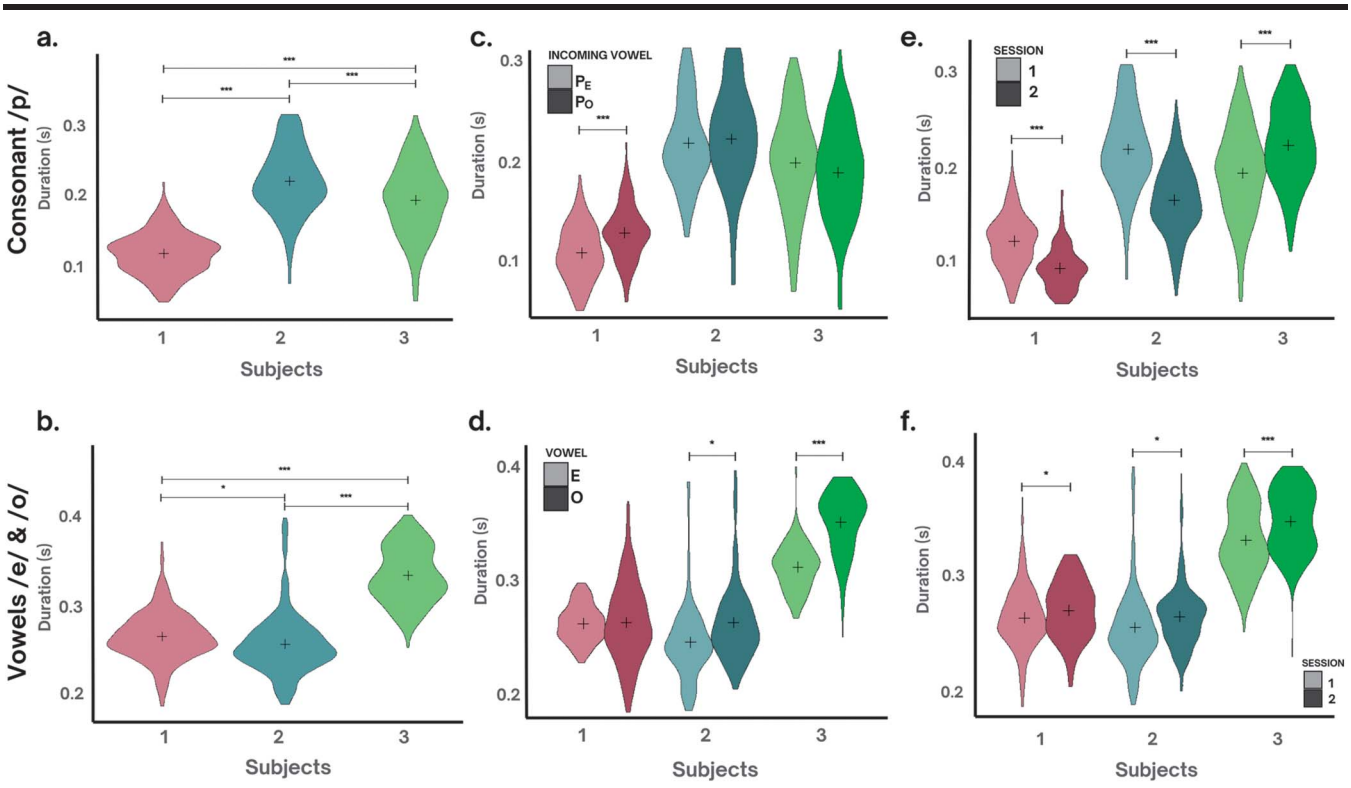
An initial exploration of consonant and vowel articulation time in a small cohort of three participants revealed significant differences both across subjects (see Figures 4a and 4b) and within subjects. Within subjects, during the same session, different vowels exhibited different durations, and the duration of /p/ was influenced by the following vowel (see Figures 4c and 4d, respectively). Articulation time for the same participant was also assessed in a subsequent session. We found that the same phonemes, produced in identical contexts by the same participant, showed significant differences across sessions (see Figures 4e and 4f). These results not only confirm the presence of idiosyncratic features in phoneme articulation but also highlight variability within individual speakers. Based on these findings, we developed our main procedure to investigate how a set of predefined factors influences articulation time within a speaker.

Main Procedure

After data collection and preprocessing, nine participants were excluded: four due to noisy EMG recordings and five because they did not achieve a synchronization level above 0.5 (see the Speech Rate Priming block for details). Thus, the analysis included data from a total of 20 participants (12 women, M_{age} = 25.25 years, SD = ±5.38).

For each Articulation block, the measured duration of the consonant /p/ and the vowels /a/ and /u/ (see Supplemental Material S1; in seconds: M_P = 0.18, SD = 0.04 and M_{VOW} = 0.3, SD = 0.07) were recorded, along with the corresponding general attentional state level, determined from the nearest General Attentional State Assessment block, and the primed rate (slow or fast), as set by the preceding Speech Rate Priming block. These data were then used to construct two models: one with the duration of the consonant /p/ and the other with the duration of

Figure 4. Exploratory procedure results. All panels show the distributions of the articulation times for the three evaluated participants. (a) All /p/ from the first session, independently of the following vowel. (b) All vowels from the first session, without distinguishing between /e/ and /o/ (n_{S1} = 229, n_{S2} = 192, n_{S3} = 233). (c) All /p/ from the first session differentiating between the following vowels. Light colors correspond to /p/ followed by /e/, while dark colors correspond to /p/ followed by /o/. (d) All vowels from the first session differentiating between /e/ and /o/. Light colors correspond to /e/, while dark colors correspond to /o/ (n_{S1pe} = 110, n_{S1po} = 119, n_{S2pe} = 98, n_{S2po} = 94, n_{S3pe} = 116, n_{S3po} = 117). (e) Comparison of all /p/ across the different sessions. (f) Comparison of all vowels across the different sessions each subject (n_{S1SS1} = 229, n_{S1SS2} = 147, n_{S2SS1} = 192, n_{S2SS2} = 218, n_{S3SS1} = 233, n_{S3SS2} = 215). In panels (e) and (f), light colors correspond to the first session, while dark colors correspond to the second session. Significance levels: *p < .05. ***p < .001.



the vowels as dependent variables. The independent variables included general attentional state level (Attention), priming rate (PRate), and vowel factors: incoming vowel (IVowel) for /p/ and phoneme identity (Vowel) for the vowels. Intercepts, but not slopes, were allowed to vary by participant.

After a backward elimination process, the model that best predicted the duration of the consonant /p/ included general attentional state level and the incoming vowel, but not the priming rate (see Table 1). Accordingly, we computed the estimated marginal means and trends for the factors included in the model. We found a positive linear relationship between /p/ duration and general attentional state level (trend = 2.65, p < .001; see Figure 5a), as well as shorter durations when the following vowel was /ɑ/ compared to when it was /u/ (*mean*/u/ = 192 ms, *mean*/ɑ/ = 176 ms, p < .001; see Figure 5b).

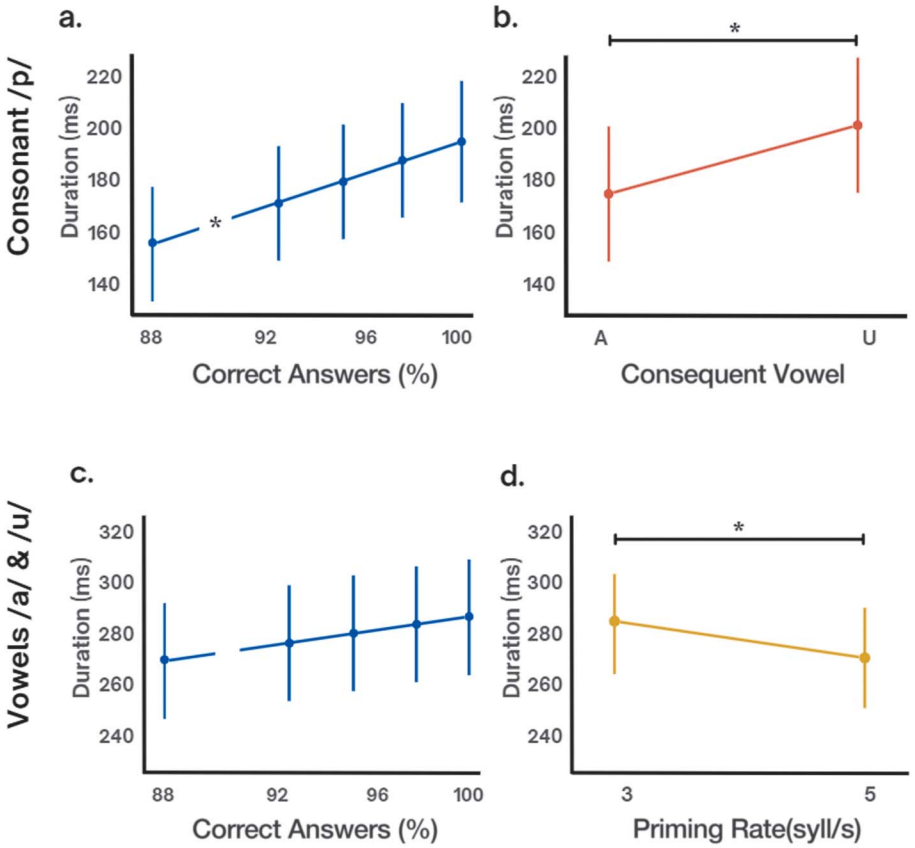
As with the consonant /p/, the model that best explained vowel durations (/ɑ/ and /u/) included general attentional state level (see Table 2). However, unlike the /p/ model, this one included priming rate but excluded phoneme identity. As with the /p/ analysis, post hoc tests showed that higher attentional levels were associated with longer phoneme durations (trend = 1.48, p = .0016; see Figure 5c). Furthermore, vowels produced after a faster speech rate priming (5 syll/s) were shorter than those

Table 1. Linear mixed-effects model results for the consonant /p/.

Consonant		
Best model	**Consonant ~ Attention + IVowel + (1 | Sub)**	
Analysis of deviance (Type III χ^2 test)		
	χ^2	*p*
Intercept	3.4544	.063
Attention	149.0618	< .001
Incoming vowel	36.5848	< .001

Note. Consonant = /p/ articulation time; IVowel = vowel following /p/ (/ɑ/ or /u/); Attention = general attentional state level on a scale from 1 to 100; Sub = participant ID.

Figure 5. Post hoc analyses. (a and b) Predicted /p/ duration as a function of general attentional state level and incoming vowel, respectively. (c and d) Predicted vowel duration (/ɑ/ and /u/) as a function of general attentional state level and priming rate, respectively. Dots represent model-predicted group means. Bars indicate 95% confidence intervals. *$p < .05$.



primed at a slower rate (3 syll/s; mean3 = 304 ms, mean5 = 299 ms, $p < .001$; see Figure 5d).

Discussion

This study examined intraspeaker variability in the articulation time of CV sequences. In an initial exploratory procedure, we observed that phoneme articulation time is not stable within individual speakers. Building on this observation, we conducted a follow-up experiment to investigate how general attentional state level, primed speech rate, and coarticulation influence phoneme articulation time within participants.

Table 2. Linear mixed-effects model results for the vowels (/ɑ/ and /u/).

Vowels		
Best model	**Vowels ~ Attention + PRate+ (1 | Sub)**	
Analysis of deviance (Type III χ^2 test)		
	χ^2	*p*
Intercept	10.8680	< .001
Attention	9.9768	< .001
Priming rate	17.0772	< .001

Note. Vowels = articulation time for /ɑ/ and /u/; Attention = general attentional state level on a scale from 1 to 100; Priming rate = slow (3 Hz) or fast (5 Hz); Sub = participant ID.

The exploratory procedure revealed variability in articulation time both between and within speakers. Significant differences were observed between participants in the duration of /p/, as well as the vowels (/e/ and /o/). These findings, based on three female Spanish-speaking participants, suggest the presence of idiosyncratic articulation features that go beyond gender and language—factors previously shown to influence phoneme articulation (Allen et al., 2003; Van Heerden & Barnard, 2008). This aligns with previous research showing temporal variability in speech segments due to speaker-specific articulator movements (Dellwo et al., 2015). Within participants, articulation time also varied. The duration of /p/ was influenced by the identity of the following vowel within the same session, and both /p/ and vowel durations differed significantly when produced under identical conditions on different

sessions, 7 days apart. These findings highlight the role of circumstantial factors, rather than idiosyncratic ones, in modulating articulation time. Based on these observations, we conducted a follow-up study (i.e., the main procedure) to examine how general attentional state, primed speech rate, and coarticulation affect phoneme articulation time at a participant's natural speaking rate.

The results from the main procedure show that, among the three factors analyzed, only general attentional state significantly affected both the consonant /p/ and the vowels (/a/ and /u/). Specifically, higher levels of attention were associated with longer production times for all phonemes. A plausible explanation for this finding is that increased attention leads to more pronounced motor gestures, thereby extending articulatory time. This aligns with previous research suggesting that greater attention demands can result in exaggerated motor movements (Dromey & Shim, 2008).

The articulation time of /p/ remains unaffected by the speech priming rate, but it is influenced by the coarticulation of the following vowel. Specifically, the duration of /p/ is longer when followed by /u/ than when followed by /a/. Previous research has shown that, due to coarticulation, stop consonants can adopt the features of adjacent vowels (Assaneo, Ramirez Butavand, et al., 2019; Keyser & Stevens, 2006; Martin & Bunnell, 1982). This may explain our findings: When /p/ is followed by the rounded vowel /u/, the consonant inherits this rounding feature, causing the lips not only to occlude but also to round. As a result, the motor gesture reaches a more distant point in the articulatory space, leading to a longer production time. This aligns with the observations from our exploratory procedure, where we noted that /p/ followed by the rounded vowel /o/ had longer durations than /p/ followed by the unrounded vowel /e/, at least in one participant.

Curiously, for the vowels, the phoneme's identity does not modify the duration (i.e., both /a/ and /u/ have equal durations). Three main articulatory features define vowels: jaw opening (closed, middle, or open), tongue position (back, central, or front), and lip configuration (rounded or unrounded; Katamba, 1989). The Spanish vowel /a/ is an open, central, unrounded vowel, while /u/ is closed, back, and rounded. As mentioned earlier, due to coarticulation, part of the vowel's motor gesture is already achieved during the production of the /p/. The lip roundness (or lack thereof) and tongue position may already be in place by the time the /p/ constriction is released. Therefore, the only remaining movement to achieve the vowel's configuration is the opening of the jaw. Although a larger jaw displacement is expected for /a/ than for /u/, our results show that this difference does not affect the vowel's duration.

In contrast, the factor that influences vowel articulation time is the primed speech rate. Vowel duration is longer when the Articulation block follows a slow Speech Rate Priming block compared to when it follows a fast Speech Rate Priming block. Our finding that speech rate priming affects vowels but not consonants aligns with existing theories. Specifically, it has been proposed that consonants are merely dynamic transitions between vowels (Browman & Goldstein, 1989). Thus, to modify speech rate, the strategy is to adjust the duration of vowels within syllables, while consonant durations remain relatively stable (Fujimura, 1981). Relating this to models proposing that speech syllabic rate emerges from brain regions functioning as neural oscillators (Assaneo & Poeppel, 2018; Poeppel & Assaneo, 2020), our results suggest that the natural frequency of these oscillators can be slightly adjusted through a brief auditory-motor synchronization task.

Additionally, from a methodological perspective, we have developed a low-cost system capable of accurately measuring lip muscle activity in precise synchrony with the produced speech signal. Considering the high cost of the technology typically required for articulography studies, this system could serve as a valuable tool for speech production researchers, particularly in underdeveloped countries.

Conclusions

Variability in phoneme articulation time has been previously reported, with several factors influencing it; however, intraspeaker variability has not been thoroughly explored. In this study, we utilized an in-lab developed system that combines EMG and audio recordings to measure phoneme articulation time. We investigated how coarticulation, speech rate priming, and general attentional state influence the duration of several Spanish phonemes (/p/, /a/, and /u/) within a speaker. Our results indicate that attention consistently affects all the phonemes studied, while coarticulation influences the consonant /p/, and speech rate priming affects the vowels.

Data Availability Statement

All data sets analyzed in the current study are available upon request from the corresponding author.

Acknowledgments

M.F.A. was supported by UNAM-DGAPA-PAPIIT IA200223. We thank Luis A. Tellez-Lima for his constructive advice.

References

Abbs, J. H., Gracco, V. L., & Cole, K. J. (1984). Control of multimovement coordination: Sensorimotor mechanisms in speech motor programming. *Journal of Motor Behavior, 16*(2), 195–232. https://doi.org/10.1080/00222895.1984.10735318

Abramson, A. S., & Whalen, D. H. (2017). Voice onset time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics, 63,* 75–86. https://doi.org/10.1016/j.wocn.2017.05.002

Acheson, D. J., & Hagoort, P. (2014). Twisting tongues to test for conflict-monitoring in speech production. *Frontiers in Human Neuroscience, 8,* Article 206. https://doi.org/10.3389/fnhum.2014.00206

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America, 113*(1), 544–552. https://doi.org/10.1121/1.1528172

Assaneo, M. F., & Poeppel, D. (2018). The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science Advances, 4*(2), Article eaao3842. https://doi.org/10.1126/sciadv.aao3842

Assaneo, M. F., Ramirez Butavand, D., Trevisan, M. A., & Mindlin, G. B. (2019). Discrete anatomical coordinates for speech production and synthesis. *Frontiers in Communication, 4,* Article 13. https://doi.org/10.3389/fcomm.2019.00013

Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience, 22*(4), 627–632. https://doi.org/10.1038/s41593-019-0353-z

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bose, A., Byrne, Á., & Rinzel, J. (2019). A neuromechanistic model for rhythmic beat generation. *PLOS Computational Biology, 15*(5), Article e1006450. https://doi.org/10.1371/journal.pcbi.1006450

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*(2), 201–251. https://doi.org/10.1017/S0952675700001019

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*(3–4), 155–180. https://doi.org/10.1159/000261913.

Celdrán, E. M. (1993). La percepción categorial de /b-p/ en español basada en las diferencias de duración [Categorial perception of /bp/ in Spanish based on duration differences]. *Journal of Experimental Phonetics*, *3,* 223–239.

Daniloff, R. G., & Hammarberg, R. E. (1973). On defining coarticulation. *Journal of Phonetics, 1*(3), 239–248. https://doi.org/10.1016/S0095-4470(19)31388-9

Del Barrio Estévez, L., & Torner Castells, S. (1999). La duración consonántica en castellano [Consonantic duration in Spanish]. *ELUA: Estudios de Lingüística, 1999*(13), 9–35. https://doi.org/10.14198/ELUA1999.13.01

Dellwo, V., Leemann, A., & Kolly, M.-J. (2015). Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors. *The Journal of the Acoustical Society of America, 137*(3), 1513–1528. https://doi.org/10.1121/1.4906837

Dromey, C., & Shim, E. (2008). The effects of divided attention on speech motor, verbal fluency, and manual task performance. *Journal of Speech, Language, and Hearing Research, 51*(5), 1171–1182. https://doi.org/10.1044/1092-4388(2008/06-0221)

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149. https://doi.org/10.3758/BF03203267

Fernández, G. A. S. (2007). *Duración relativa de los segmentos vocálicos del español de la Provincia de Ñuble, en hablantes de diferente nivel sociocultural y procedencia geográfica* [Relative duration of vowel segments in the Spanish spoken in the Province of Ñuble, among speakers of different sociocultural levels and geographical origins; Doctoral dissertation, Universidad de Concepción].

Fossett, T. R. D., McNeil, M. R., Pratt, S. R., Tompkins, C. A., & Shuster, L. I. (2016). The effect of speaking rate on serial-order sound-level errors in normal healthy controls and persons with aphasia. *Aphasiology, 30*(1), 74–95. https://doi.org/10.1080/02687038.2015.1063581

Fujimura, O. (1981). Temporal organization of articulatory movements as a multidimensional phrasal structure. *Phonetica, 38*(1-3), 66–83. https://doi.org/10.1159/000260015

Hirose, K., & Kawanami, H. (2002). Temporal rate change of dialogue speech in prosodic units as compared to read speech. *Speech Communication, 36*(1–2), 97–111. https://doi.org/10.1016/S0167-6393(01)00028-0

Hualde, J. I. (2013). *Los sonidos del español: Spanish language edition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511719943

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change, 21*(2), 233–256. https://doi.org/10.1017/S0954394509990093

Katamba, F. (1989). *An introduction to phonology* (Vol. 48). Longman.

Keyser, S. J., & Stevens, K. N. (2006). Enhancement and overlap in the speech chain. *Language, 82*(1), 33–63. https://doi.org/10.1353/lan.2006.0051

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*(14), 1–16.

Krohn, H. S. (2019). Duración vocálica en el español de la Gran Área Metropolitana de Costa Rica [Vocalic duration of the spoken Spanish at the metropolitan area of Costa Rica]. *Revista de Filología y Lingüística de la Universidad de Costa Rica, 45*(1), 215–224. https://doi.org/10.15517/rfl.v45i1.36736

Lenth, R. (2024). *emmeans: Estimated marginal means, aka least-squares means* (R package Version 1.10.3-090006) [Computer software]. https://rvlenth.github.io/emmeans/

Mares, C., Echavarría Solana, R., & Assaneo, M. F. (2023). Auditory-motor synchronization varies among individuals and is critically shaped by acoustic features. *Communications Biology, 6*(1), Article 658. https://doi.org/10.1038/s42003-023-04976-y

Marín Gálvez, R. (1995). La duración vocálica en español [Vocalic duration in Spanish]. *ELUA: Estudios de Lingüística, 10,* 213–226. https://doi.org/10.14198/ELUA1994-1995.10.11

Martin, J. G., & Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel–stop consonant–vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance, 8*(3), 473–488. https://doi.org/10.1037//0096-1523.8.3.473

Mendoza, E., Carballo, G., Cruz, A., Fresneda, M. D., Muñoz, J., & Marrero, V. (2003). Temporal variability in speech segments of Spanish: Context and speaker related differences. *Speech Communication, 40*(4), 431–447. https://doi.org/10.1016/S0167-6393(02)00086-9

Parker, F. (1974). The coarticulation of vowels and stop consonants. *Journal of Phonetics, 2*(3), 211–221. https://doi.org/10.1016/S0095-4470(19)31271-9

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience, 21*(6), 322–334. https://doi.org/10.1038/s41583-020-0304-4

Song, J.-H. (2019). The role of attention in motor control and learning. *Current Opinion in Psychology, 29,* 261–265. https://doi.org/10.1016/j.copsyc.2019.08.002

Stevens, K. N. (2000). *Acoustic phonetics* (Vol. 30). MIT Press. https://doi.org/10.7551/mitpress/1072.001.0001

Van Heerden, C. J., & Barnard, E. (2008). Speaker-specific variability of phoneme durations: Reviewed article. *South African Computer Journal, 2008*(40), 44–50. https://hdl.handle.net/10520/EJC28047

Voeten, C. C., & Voeten, M. C. C. (2021). *Package 'buildmer'* [Computer software]. https://cran.r-project.org/web/packages/buildmer/index.html

Walker, J. F., Archibald, L. M. D., Cherniak, S. R., & Fish, V. G. (1992). Articulation rate in 3- and 5-year-old children. *Journal of Speech and Hearing Research, 35*(1), 4–13. https://doi.org/10.1044/jshr.3501.04

Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics, 18*(1), 3–35. https://doi.org/10.1016/S0095-4470(19)30356-0

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S. N., & Baayen, R. H. (2016). Investigating dialectal differences using articulography. *Journal of Phonetics, 59,* 122–143. https://doi.org/10.1016/j.wocn.2016.09.004