Nonlinear effects of speech rate on articulatory timing in singletons and geminates

Sam Tilsen, Anne Hermes

Nonlinear effects of speech rate on articulatory timing in singletons and geminates

Sam Tilsen[1], Anne Hermes[2]

[1]*Department of Linguistics, Cornell University, Ithaca*
[2]*Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS/Sorbonne-Nouvelle, Paris*
tilsen@cornell.edu, anne.hermes@sorbonne-nouvelle.fr

Abstract

An understanding of the relations between speech rate and articulatory timing is critical to developing adequate models of articulatory control. In the case of geminate consonants, not much is known about how articulatory timing varies with speech rate, nor is it known whether the form of variation is similar between singletons and geminates. We investigated how gestural timing varies with speech rate in intervocalic /m/ and /mm/ from speakers of Tashlhiyt Berber, Japanese, and Italian. We found that while the timing of closure and release is nonlinearly constrained in singletons, such constraints do not apply to geminates. A secondary finding is that speech rate has complicated, speaker-specific effects on variability in timing. Together these patterns suggest that control of articulatory timing in singleton and geminate consonants may be accomplished by distinct mechanisms, particularly at slow rates of speech.

Keywords: speech rate, geminates, timing, speech production, selection-coordination, Articulatory Phonology, inter- and intragestural timing.

1. Speech rate and timing

This paper presents an investigation of how articulatory timing in singleton and geminate consonants varies as a function of speech rate. A sensible null hypothesis is that timing measures vary linearly with speech rate, but we suspect that the null hypothesis of linear rate effects may be incorrect. It is also possible that certain timing measures may be constant with respect to speech rate, or may vary nonlinearly. We analysed measures of articulatory timing and found substantial differences between rate-timing relations in singletons and geminates. Specifically, while (a) the interval between initiation of the constriction and the initiation of the vocalic movement was relatively independent of rate for both segment types (*c-v interval*), (b) the interval between constriction and release (*c-r interval*) was linearly related to speech rate for geminates, but nonlinearly related for singletons. This finding is important, because it puts constraints on models of speech production. An important aspect of our method is a technique for eliciting a wide range of speech rates, without relying on qualitative, categorical rate instructions, such as *speak fast* or *speak slow*. By eliciting continuous variation in rate, our method facilitates a more precise characterization of relations between rate and timing measures.

An understanding of how articulatory timing varies with speech rate is useful because it may help resolve between various theories and models of phonological representation. Standard varieties of phonological representation provide a number of possible options for conceptualizing the organization of intervocalic singletons and geminates in a segment sequence or within larger syllabic structure (Fig. 1). In some cases, phonological patterns—particularly quantity sensitivity, but also degemination—may provide arguments for some of these options on a language-by-language basis.

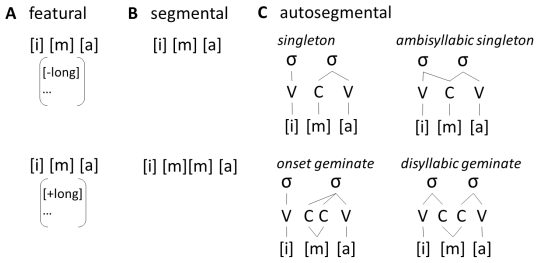

Figure 1: *Alternative representations of intervocalic singletons and geminates:* A: featural representations; B: segmental representations; C: autosegmental representations; Kubozono, 2017)

An alternative paradigm of phonological representation—Articulatory Phonology (cf. Browman & Goldstein, 1989, 2000)—raises a different set of questions regarding phonological representation. In the standard AP framework, word representations are instantiated by gestural scores, and timing relations between intervals of gestural activation are determined by phase-coupling of gestural planning oscillators; such coupling can obtain either in-phase or anti-phase modes. There are several open questions regarding the gestural composition of geminates and the pattern of coupling relations between gestures. Several alternative possibilities are shown in Fig. 2. One possibility is that geminates involve a single constriction gesture (monogestural representation); alternatively, two constriction gestures may be coordinated (digestural geminates). In addition, under the split-gesture hypothesis (Nam, 2007; Tilsen, 2017), the constriction and release phases of articulatory movements may be controlled by separate, dissociable constriction and relation gestures. Within each of these possibilities, there are numerous ways in which the pattern of coupling relations between gestures could obtain.

Furthermore, feedback mechanisms may be involved in the control of articulatory timing. This is the case in selection-coordination theory (Tilsen, 2016), which extends the AP framework to include feedback-based mechanisms of timing control. Specifically, two gestures may be competitively selected by use of external or internal feedback, rather than coordinatively controlled through phase-coupling. For example, if the release gesture of a consonant is competitively controlled relative to the constriction gesture, then there is a feedback threshold which determines when the constriction gesture is supressed, which in turn allows for the release gesture to be selected. The feedback threshold here relates to feedback regarding achievement of the constriction target. If the feedback threshold varies linearly with speech rate, a linear delay of the release initiation relative to closure initiation is predicted. In contrast, under a coordinative regime of control, where gestural initiations are triggered by phase-coupled

planning oscillators, the relative timing of gestural initiations can be constrained by bounds on the frequencies of gestural planning oscillators.

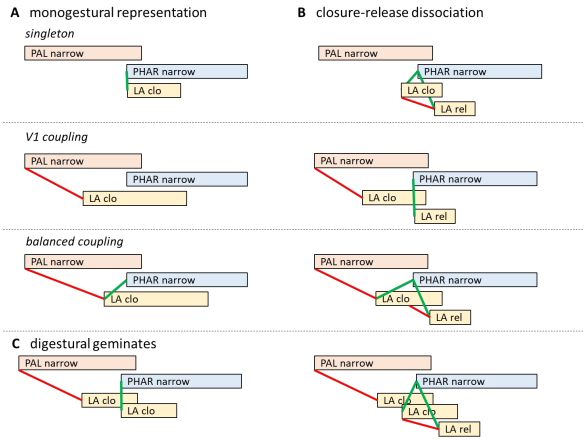

Figure 2: *Possible compositions and couplings for singleton and geminates*. A: monogestural representations with various coupling patterns; B: representations in which closure and release gestures are dissociated; C: digestural representations.

A first step toward using phonetic evidence to constrain articulatory control models is to obtain an adequate characterization of empirical patterns. In order for this characterization to be sufficiently general, it should address effects of speech rate on timing measures. This paper conducts an exploratory investigation of speech rate effects, with an immediate goal of placing constraints on the space of possible models. This will facilitate the longer-term project of inferring gestural organizations from phonetic data.

2. Methods

2.1. Participants and data collection

We recorded 3 speakers, one speaker of Tashlhiyt Berber (S1), one of Japanese (S2) and one of Italian (S3). Articulatory data were recorded with a 3-dimensional Electromagnetic Articulograph (Carstens Medizinelektronik; AG501). Sensors were located on the upper and lower lip, tongue tip, tongue blade, and tongue body. Sensors on the nasion and left/right mastoid processes were used for head movement correction, and sensor data were rotated so that the occlusal plane (estimated by a bite plate) was located horizontally. The sensor positions were sampled at 1250Hz, then downsampled to 250Hz and smoothed with a 40Hz low-pass filter and a 3-step floating mean. Time-synchronized acoustic data were recorded using a condenser microphone (AKG C420 headset) sampled at 48kHz. All data were converted to SSFF format using custom software (EMA2SSFF). Forced alignment of responses was conducted with Kaldi (Povey et al. 2011). Monophone Hidden Markov Models (HMM) were trained on 12 hand-labeled trials for each participant, with no imposed distinction between singleton and geminate phones.

2.2. Task and stimuli

The target words in all three languages were /ima/ and /imma/. Target words were produced in carrier phrases (see Table 1). Each speaker performed 32 blocks of 20 trials over two sessions (16 blocks per session), resulting in a total of 320 repetitions of each target word (640 trials). Blocks alternated between increasing and decreasing rate cues (see below). Target words (i.e. /ima/ vs. /imma/) were alternated every four blocks, and the first block always had the singleton target.

Table 1: *Carrier phrases in the experiments.*

Language	Carrier Phrase	Gloss
Tashlhiyt	Innajam ___ bahra.	He told you _ a lot.
Japanese	Kore wa ___ nano.	This is _.
Italian	Parli con ___ per favore.	Talk to _ please.

To elicit variation in speech rate, a visual analog cue for rate was employed. The visual cue was a red box that moved across the screen over a range of periods (in 20 steps from 750 to 3000 ms). Speakers were instructed to produce the phrases at the pace that reflected the speed of the moving box, after it moved off the screen. In every other block for a given target, the cue rate was either increased or decreased sequentially over the 20 steps continuum of target rates.

To diminish interspeaker differences in rate control strategies, we explicitly instructed speakers not to pause between words and instead to control their rate by producing the words of the phrase more slowly. As Fig. 3 shows, these instructions were generally successful: carrier phrase durations were relatively uniform and span a wide range (95% density intervals spanned 2.07 s, 1.74 s, and 2.09 s for S01, S02, and S03, respectively). Word and silent interval durations in Fig. 3 were obtained from the forced alignment.

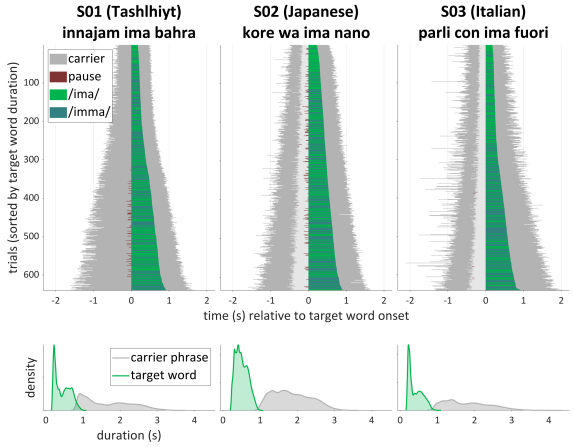

Figure 3: *Carrier phrase, pause and target word durations.* Top: word and silence interval for all trials, sorted by target word duration and aligned to target word onset. Bottom: Gaussian kernel densities of target word and carrier phrase durations for each speaker.

In analyses of rate effects, we use target word duration as an independent variable. This variable is preferable over the cue duration (i.e., target rate) because it is an index of *effective* speech rate, i.e., the rate that speakers employed in a given trial. The duration of the target word is used as a rate measure rather than the carrier phrase duration because rate variation may be manifested differently in different words of the carrier phrase, and this manifestation may differ across languages or speakers; the duration of the target word can be viewed as a more local measure of speech rate than the carrier phrase duration. However, we note that analyses conducted with carrier phrase duration as an independent variable (not included here) do not differ qualitatively from analyses with target word duration as an independent variable.

2.3. Data processing and analysis

Articulatory kinematic landmarks were extracted using the following procedure. First, the velocity extrema associated with the bilabial constriction of [m], bilabial release of [m], formation of [i], and formation of [a] were identified for each subject and target form. This was accomplished by first applying an iterated trajectory-alignment procedure using $\dot{x}_i$, the velocity of the relevant kinematic time series for trial i (LA for bilabial closure and release gestures, first principal component of TB for the vocalic gestures; examples of these are shown in Fig. 4). In each iteration of the alignment procedure, the mean trajectory $\hat{x}$ is calculated over $\dot{x}_i$ for each time step, over a 1 s period of signal which has been Gaussian-windowed to diminish the contribution of values further from the centre. The cross-covariance function between $\dot{x}_i$ and $\hat{x}$ is then calculated for each trial, and each time series is shifted according to the lag with maximum cross-covariance. The mean trajectory $\hat{x}$ is recalculated and the procedure is repeated until no more time shifting is required. Subsequently, the relevant velocity extremum for each trial/gesture is identified as the extremum closest to the velocity extremum of the mean trajectory. Gestural movement initiation and target achievement events are then located relative to velocity extrema using a 20% velocity threshold (i.e., onsets/targets are points in time when speed first rises above/falls below the value of the extremum). In the example of Fig. 4, velocity extrema landmarks are shown as green squares and gestural initiations/target landmarks are shown as red circles.

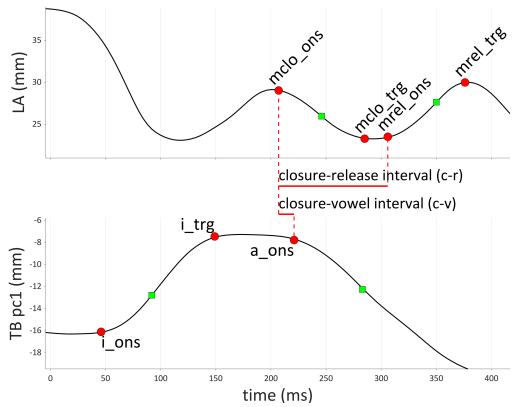

Figure 4: *Example of kinematic trajectories and articulatory landmark, i.e., initiations (ons) and targets (trg).* Top: lip aperture (LA) for bilabial closure (mclo) and release (mrel). Bottom: first principal component of tongue body position (TB pc1).

The variables analyzed in Figs. 5 and 6 are (i) the closure-release interval (abbreviated as *c-r interval*), which is the period of time from the initiation of the bilabial closure to the initiation of the release of that closure (Fig. 4: mclo_ons-mrel_ons); and (ii) the closure-vowel interval (abbreviated as *c-v interval*), which is the period of time from the initiation of the bilabial closure to the initiation of the vocalic gesture (Fig. 4: mclo_ons-a_ons) These intervals are represented by the horizontal red lines in Fig. 4. For both intervals, linear and nonlinear regression models (of the form $y = \beta_0 - \beta_1 e^{\beta_2 x}$) were fit to the data. Estimates of the coefficient of variation—the ratio of standard deviation to mean $\left(\frac{\sigma}{\mu}\right)$—as a function of target word duration in Fig. 7 were obtained by calculating the standard deviation σ and mean μ for a moving window of 100 observations; confidence intervals were obtained for each window with 1000 bootstrap samples.

3. Results

The main findings are: (a) the intergestural *c-v interval* is relatively constant, i.e., independent of rate, for both singletons and geminates; (b) the intragestural *c-r interval* increases linearly with rate in geminates but increases nonlinearly in singletons; and (c) coefficients of variation for the *c-r interval* were not constant and were non-monotonic for two of the three participants. Findings (a) and (b) together show that the initiations of consonantal closure and release gestures are approximately symmetrically displaced from the vocalic gestural initiation for singletons (i.e., a c-center effect), but not for geminates.

Linear and exponential fits of the *c-r* and *c-v intervals* as a function of target word duration (i.e., speech rate) are shown in the top and bottom rows of Fig. 5. Linear fits are depicted with dashed grey lines, exponential fits are shown with solid color lines. A visual comparison of fits shows that the c-r interval (top row) scales approximately linearly with speech rate for geminates but exhibits a nonlinear relation for singletons. This is supported by Akaike information criterion (AIC) differences ($\Delta AIC = AIC_{nonlin} - AIC_{lin}$). Notice that the rate effect in singletons attenuates at slower rates, suggesting that there is a constraint on this interval. Also, notice that for S01 and S03 the *c-r interval* distributions overlap substantially at fast-rates.

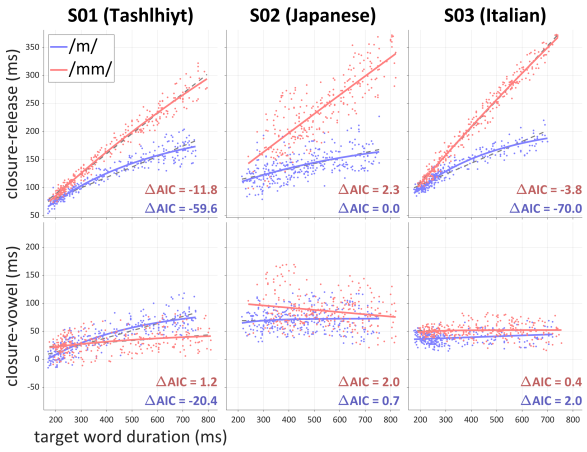

Figure 5: *Exponential and linear model fits of articulatory timing intervals.* (Blue=singleton, red=geminate). Top row: *c-r interval*, i.e., bilabial closure initiation to release interval as a function of word duration. Bottom row: *c-v interval*, i.e., bilabial closure initiation to vocalic gesture initiation as a function of word duration

In contrast, the *c-v interval* (Fig. 5: bottom) is relatively constant for both targets. An exception is S01 (Tashlhiyt) singletons, where it exhibits some degree of nonlinearity.

The patterns in Fig. 5 are illustrated in a different manner in Fig. 6, where the dashed line indicates the initiation of the vocalic gesture and speech rates are plotted on the vertical dimension. This figure reinforces the interpretation that the most extensive rate-related effect is a delay of the release of the bilabial closure relative to the initiation of the vocalic gesture in geminates. It also shows how the *c-r interval* in singletons expands with rate for S01 and S03.

The relations between rate and the coefficient of variation (ratio of standard deviation to mean) of the *c-r interval*, shown in Fig. 7, are highly nonlinear and speaker-specific. For S1, the coefficient of variation exhibits a peak at moderate speech rates for both /m/ and /mm/ targets. The same is observed for geminates of S3, but singletons the coefficient of variation

appears to increase linearly and reach a plateau. In contrast, for S2, the coefficient of variation appears to decrease exponentially with rate.

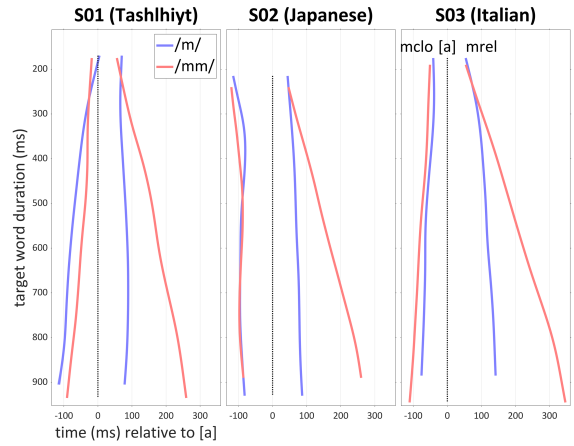

Figure 6: *Rate effect on relative timing.* Panels show spline fits of bilabial closure initiation (mclo) and release initiation (mrel), relative to initiation of vocalic gesture (vertical dotted line); fits are shown for singletons (blue) and geminates (red); vertical axis is duration of the target word.

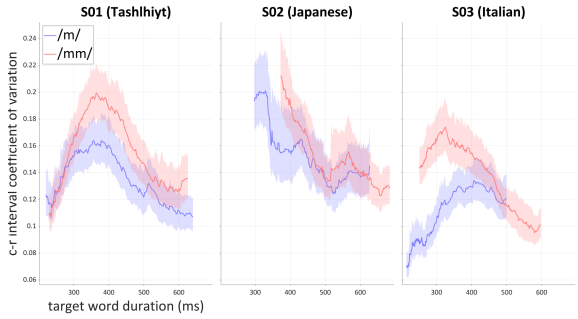

Figure 7: *Coefficient of variation of c-r interval as a function of target word duration.* Singleton (blue), geminate (red). Shaded intervals are 95% confidence intervals based on 1000 bootstrap samples.

4. Discussion and conclusion

The results suggest that (i) bilabial closure and release initiation are coordinatively controlled in singletons, but (ii) constriction release is not coordinatively controlled in geminates.

The *c-v interval* was relatively constant for both singletons and geminates; this indicates a precise, coordinative control of this interval. In contrast, the timing of the release gesture relative to either the initiation gesture (Fig. 5) or the initiation of the vocalic gesture (Fig. 6) appears to be constrained in singletons in a way that it is not in geminates. Specifically, the nonlinear relation between rate and closure-release timing in singletons can be interpreted as an attenuation effect, which is predicted by a model in which the timing of constriction and release initiation is triggered by anti-phase coupled planning oscillators. Under the reasonable assumption that the frequencies of the planning oscillators are sensitive to speech rate but limited to a specific range, the coupled oscillators model predicts this attenuation effect: as speech rate slows, oscillator frequency ω approaches a lower bound ω_{min} and thus the *c-r interval* $\delta_{c-r} = \frac{\phi}{2\pi\omega_{min}}$ approaches a maximum (here ϕ is the relative phase in radians of stabilized closure and release gestural planning oscillators; see Tilsen, 2017).

In contrast, for geminates the effect of slower rate is a linear delay of the release relative to closure initiation and vocalic gesture initiation. This pattern indicates that rather than being coordinatively controlled through phase-coupling, the timing of the release in geminates is governed by an alternative mechanism. A plausible model of this is feedback-based competitive control in the selection-coordination framework. This model holds that the bilabial constriction and release gestures are competitively selected: the release gesture cannot be selected until the constriction gesture is suppressed by feedback systems. The exact timing of the suppression is hypothesized to be governed by a threshold which applies to sensory feedback regarding achievement of the constriction target. Specifically, speakers are hypothesized to linearly increase the threshold (i.e., require more sensory feedback regarding constriction target achievement) in slower speech, and this results in a temporal delay between the initiation of the constriction gesture and its suppression, which in turn delays the selection of the release gesture. Note that the feedback may be a combination of external sensory feedback or internal predictive feedback.

Furthermore, analysis of the relation between speech rate and coefficient of variation reveals that the rate-dependence of variability is both target-dependent and speaker-dependent. The nonlinearity of the relation is not consistent with a model in which speech rate exerts a multiplicative effect on variance. The nonmonotonicity of this relation (for S1 and for S03 singleton targets) may suggest that timing control mechanisms used for both singletons and geminates are rate-dependent, although further analysis is necessary to investigate this possibility.

In conclusion, the findings are important because they indicate that timing of constriction gestures in intervocalic singletons and geminates cannot be governed by a monolithic control mechanism. Instead, an adequate model must generate a linear increase of the *c-r interval* for geminates and a non-linear attenuation of this interval for singletons. One such model is the competitive control model of selection-coordination theory, in which gestural activation interval durations can be controlled via sensory feedback thresholds.

5. Acknowledgements

We would like to thank Theo Klinker at the IfL-Phonetics laboratory at the University of Cologne for his support with the data collection and processing.

6. References

Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6(2), 201–251.

Browman, C., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de La Communication Parlée*, 5, 25–34.

Kubozono, H. (Ed.). (2017). *The phonetics and phonology of geminate consonants* (Vol. 2). Oxford University Press.

Nam, H. 2007. Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and moraic structure. In J. Cole & J. I. Hualde (Eds.), *Laboratory phonology 9 (phonology and phonetics)* (2007) (pp. 483-506). Berlin, New York: Walter de Gruyter.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., ... & Silovsky, J. (2011). The Kaldi speech recognition toolkit. IEEE Signal Processing Society.

Tilsen, S. 2016. Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics*, Vol. 55, 53-77.

Tilsen, S. 2017. Exertive modulation of speech and articulatory phasing. *Journal of Phonetics*.