Laboratory Phonology
Journal of the Association for Laboratory Phonology

Burroni, F., Bučar Shigemori, L. S., Kawahara, S., & Shaw, J. A. (2026). A gestural account of consonantal length contrasts: Articulatory evidence from Japanese bilabial, alveolar, palatal, and dorsal mimetic geminates. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 17(1), pp. 1–42. DOI: https://doi.org/10.16995/labphon.26525

Open Library of Humanities

A gestural account of consonantal length contrasts: Articulatory evidence from Japanese bilabial, alveolar, palatal, and dorsal mimetic geminates

Francesco Burroni*, Institute for Phonetics and Speech Processing, Ludwig-Maximilians-Universität München, Germany; Center of Excellence in Southeast Asian Linguistics, Chulalongkorn University, Thailand, francesco.burroni@phonetik.uni-muenchen.de

Lia Saki Bučar Shigemori, Institute for Phonetics and Speech Processing, Ludwig-Maximilians-Universität München, Germany

Shigeto Kawahara, The Institute of Cultural and Linguistic Studies, Keio University, Japan

Jason A. Shaw, Department of Linguistics, Yale University, USA

*Corresponding author.

We present articulatory data regarding how Tokyo Japanese speakers produce consonantal length contrasts. Six speakers produced 15 mimetic items across bilabial, alveolar, palatal, and dorsal places as singletons and geminates. Data were recorded using electromagnetic articulography. Analyses focused on intra-gestural timing, kinematics, and inter-gestural timing properties. Geminate production shows a modestly longer closing phase but a much longer plateau phase duration, with no substantial differences in release phases. Dynamic Time Warping analyses also localize temporal expansion to the plateau. Kinematically, geminates exhibit greater constriction amplitude, reduced peak velocity, and lower stiffness. Inter-gesturally, vowel–to–target and vowel–to–vowel lags increase, leading to increased acoustic duration of preceding vowels, while vowel–to–consonant onset lags remain stable. Linear Discriminant Analyses reveal that temporal cues outperform spatial ones in classification. Based on these findings, we propose possible phonological accounts, including a unified gestural account: a local prosodic temporal modulation (μ_T) prolongs activation, yielding plateau extension and emergent spatial strengthening, parsimoniously explaining gemination across articulators.

Laboratory Phonology: Journal of the Association for Laboratory Phonology is a peer-reviewed open access journal published by the Open Library of Humanities. © 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See https://creativecommons.org/licenses/by/4.0/. Articles published by the Open Library of Humanities may also contain other CC-licensed material, or material used or reproduced under an exception or limitation to copyright. Such material is not subject to the article's license. It is either marked with its own license, or marked that it has been reproduced with permission or under fair-use.
OPEN ACCESS

1. Introduction

1.1. General introduction

In a number of the world's languages, speakers contrast the production of two series of consonants, often termed singletons, or "short," and geminates, or "long" (Kubozono, 2017; Ladefoged & Maddieson, 1996; Ridouane, 2010). Labels like "short" and "long" showcase the importance that phonetic and phonological analyses have traditionally placed on acoustic durations. This emphasis may suggest, deceptively, that the production of singleton/geminate contrasts is a relatively straightforward matter of acoustic durations. However, beneath the surface of this apparent simplicity lie many layers of complexity.

First, it is not a given that speakers directly control acoustic durations in speech production: speech unfolds too rapidly to rely on acoustic outputs to guide the motor system (Guenther, 2016; Guenther & Hickok, 2015). Second, articulatory to acoustic mappings are non-linear, non-unique, and quantal (Atal et al., 1978; Keyser & Stevens, 2006; Neiberg et al., 2008; Stevens, 1989). Thus, ensuring appropriate acoustic outputs as durations increase is challenging—consider, for example, maintaining voicing during a voiced geminate consonant (Hayes & Steriade, 2004) or geminating a complex segment, like an affricate (Pycha, 2009). Third, how speakers regulate the duration of their vocal tract actions, which underlie acoustic durational differences, remains a controversial matter in the literature (cf. Fowler, 1980; Goldstein et al., 2009; Guenther, 2016; Guenther & Hickok, 2015; Kelso et al., 1986; Krause & Kawamoto, 2020; Saltzman & Munhall, 1989; Tilsen, 2022; Turk & Shattuck-Hufnagel, 2020a, 2020b). One class of models, intrinsic-timing models, assumes no direct control of timing and the emergence of timing from coupled oscillators (Goldstein et al., 2009), spatial control (Fowler, 1980; Kelso et al., 1986), and/or error signals (Guenther, 2016; Guenther & Hickok, 2015). In contrast, extrinsic-timing models assume that speakers directly plan movement durations (Turk & Shattuck-Hufnagel, 2020a, 2020b). Hybrid intermediate positions also exist (Saltzman & Munhall, 1989; Tilsen, 2022).

These three issues directly lead to the core problem: identical differences in acoustic duration may be brought about by different articulatory strategies, e.g., by multiple articulatory gestures, by prolonging or holding constrictions, by altering the relative timing of articulatory gestures, or by changing the speed of articulatory gestures. Acoustic realizations of duration are, thus, ambiguous with respect to the articulatory strategies that produce them.

Thus, a crucial question is the following: How do speakers control their vocal tract to produce singleton–geminate contrasts, which require fine-grained temporal control on very short, segmental time scales? (cf. Turk & Shattuck-Hufnagel, 2020a, 2020b, for studies addressing a similar issue concerning long vowels). This issue remains both understudied and largely unresolved. In this paper, we address this question and extend our previous work (Burroni et al.,

2025) on the topic by presenting new electromagnetic articulography (EMA) data on geminate production across multiple places of articulation from six speakers of Tokyo Japanese.

1.2. Geminate acoustics in Tokyo Japanese and other languages

Tokyo Japanese (TJ)—like other Japonic varieties (Shibatani, 1990)—is well-known for the presence of a robust singleton/geminate contrast, e.g., pairs like 先 [sɑki] 'point, edge' and 殺気 [sɑkːi][1] 'seethe; thirst for blood'. In addition, TJ speakers can productively create singleton/geminate minimal pairs to emphasize mimetic (also known as sound symbolic) words, e.g., ピカピカ [pikɑpikɑ] 'shiny' and ピッカピカ [pikːɑpikɑ] 'very shiny' (Kawahara, 2013). Mimetic gemination is particularly well-suited to study vocal tract control during geminate production because it offers an environment where singletons and geminates can be compared within the same lexical items, thus ensuring that all other phonological properties remain identical (e.g., syllable count, word length, segmental context, lexical frequency). Mimetic gemination is thus useful to control for additional factors, which are known to affect speech production in general and length contrasts in particular (Burroni et al., 2025; Shaw & Kawahara, 2019; Tomaschek et al., 2018).

The production of TJ consonantal length contrasts has been extensively studied in terms of its acoustic correlates, with a special focus on durational properties. This durational focus, common to TJ and other languages, is justified by the fact that, cross-linguistically, geminates are distinguished from singletons by longer constriction durations (Kawahara, 2015; Kubozono, 2017; Ridouane, 2010).

However, the production of geminate consonants in TJ (Kochetov, 2025; Löfqvist, 2005) also encompasses the presence of additional manifestations, beyond local differences in duration of the consonants themselves. For example, TJ geminates are accompanied by slightly longer preceding vowels and by other non-durational cues, such as larger intensity differences surrounding geminates, larger pitch-accent f0 drops across geminates, and vowels with lower F1 following geminates; see Kawahara (2015) for an overview. The TJ situation is not unique. While consonant duration is unambiguously the primary correlate of the singleton-geminate contrast cross-linguistically, the phonetic manifestation of the contrast is hardly reducible to duration alone. A growing body of evidence suggests that geminate consonants are also characterized by a range of non-temporal acoustic and articulatory properties, including differences in spectral properties, voice quality, and "articulatory strength," such as more extended linguo-palatal contacts (Abramson, 1998; Al-Tamimi & Khattab, 2015, 2018; Burroni et al., 2022, 2024; Dicanio, 2012; Kochetov, 2025; Kochetov & Kang, 2017;

[1] We follow the IPA convention to transcribe geminate consonants using the symbol /ː/. This choice additionally allows us to encompass both the *sokuon* /Q/ (Jap. /っ/) and *hatsuon* /N/ (Jap. /ん/) moraic phonemes and to avoid the bi-gestural assumption implicit in transcriptions like /CC/.

Local & Simpson, 1999; Löfqvist, 2005; Ridouane, 2010). These findings have led some researchers to propose that the contrast is better characterized as one of articulatory strength, whereby geminate consonants reflect *fortis* or *tense* articulations (Jessen, 2001; Kochetov & Kang, 2017; Kohler, 1984). Under this view, duration constitutes the primary cue while non-temporal correlates serve an enhancing function, a hierarchy supported by discriminant analyses showing duration as the dominant classifier with non-temporal cues contributing secondarily (Al-Tamimi & Khattab, 2018). Notably, the magnitude of non-temporal correlates varies cross-linguistically (Ridouane, 2010): in Lebanese Arabic, systematic non-temporal cues are present alongside durational differences (Al-Tamimi & Khattab, 2015, 2018). In other languages, like Itunyoso Trique, the picture is more complex, with some non-temporal patterns following directly from duration while others, like glottal spreading, appear to be independently specified (Dicanio, 2012). The exact relationship between durational and non-durational correlates thus remains an open theoretical question: feature-based accounts must stipulate which dimension is primary (Jessen, 2001; Kohler, 1984), whereas gestural models offer the possibility of a principled unification, whereby spatial strengthening could emerge from temporal modulations, rather than being independently specified. We explore this possibility in this work.

In sum, the presence of non-durational and non-local acoustic differences already suggests that the production of geminate consonants in TJ and in other languages is not simply a matter of acoustic durations (Abramson, 1998; Al-Tamimi & Khattab, 2015, 2018; Burroni et al., 2022; Dicanio, 2012; Local & Simpson, 1999; Löfqvist, 2006; Ridouane, 2003)—but rather different production mechanisms may be at play. In addition, as noted, how exactly longer acoustic durations come about in terms of articulatory activities in the vocal tract remains an open question in the literature, both with respect to their articulatory underpinnings and whether these are shared across languages (Burroni, 2022; Burroni et al., 2024, 2025; Dunn, 1993; Fujimoto et al., 2015; Kelso et al., 1986; Klatt, 1976; Kochetov & Kang, 2017; Löfqvist, 2005, 2006, 2007; Ridouane, 2003, 2010; Smith, 1992, 1995; Tilsen, 2022; Tilsen & Hermes, 2020; Turk & Shattuck-Hufnagel, 2020a).

1.3. Models of geminate production in TJ and beyond

The production of geminate consonants in both TJ and other languages has been examined in several studies (Burroni et al., 2024, 2025; Dunn, 1993; Fujimoto et al., 2015; Kochetov, 2025; Kochetov & Kang, 2017; Löfqvist, 2006, 2007; Morimoto, 2020; Morimoto et al., 2025; Ridouane, 2010; Smith, 1992, 1995; Tilsen & Hermes, 2020), yet how exactly geminate duration is controlled remains an open question. Two possibilities in particular have been considered: inter-gestural accounts and intra-gestural/kinematic accounts.

1.3.1. Inter-gestural accounts

Inter-gestural accounts hold that geminates may involve a different inter-gestural timing organization with surrounding vocalic gestures (Burroni, 2022; Nam, 2007b; Smith, 1992, 1995). These accounts imply that geminate closures may function as codas of preceding syllables, a fact that is reflected in their earlier initiation relative to a preceding vowel that is often shortened.

TJ geminates, however, stand out cross-linguistically by not showing this "compensatory" shortening effect in contrast to languages such as Kannada, Tamil, Telugu, Hausa, Italian, Icelandic, Norwegian, Hungarian, Shilha, Amharic, Galla, Dogri, Bengali, Sinhalese, and Rembarrnga, in which vowels preceding geminates are shorter than the vowels preceding singletons (Maddieson, 1984). Some works have suggested that the shortening may in fact have an articulatory basis related to the earlier initiation of geminate closures relative to the preceding vowel (e.g., Burroni et al., 2024; Celata et al., 2022).

The different behavior of TJ compared to other languages, especially Italian, has been attributed to general principles of consonant-vowel coordination in work by Smith (1992, 1995), who introduced an *inter-gestural coupling model* based on rhythmic classes.

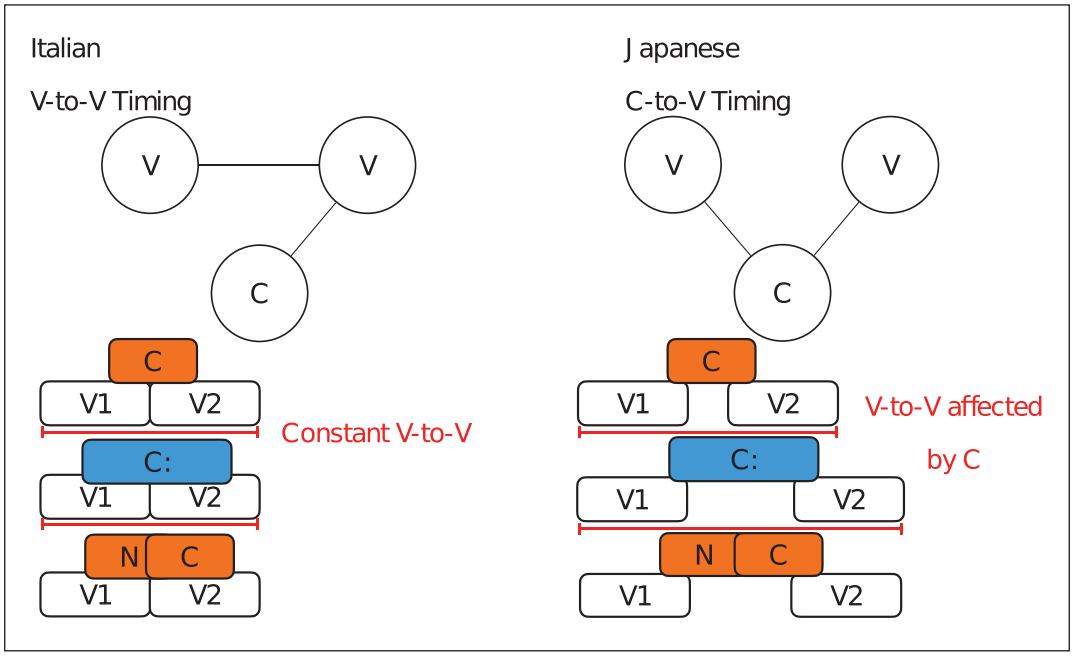

Figure 1: Left : Italian V-to-V Timing where consonants are superimposed on a constant V-to-V cycle. **Right**: Japanese C-to-V timing where consonants affect the duration of V-to-V intervals.

In this model, the durational differences observed between Italian and TJ are attributed to syllable- and mora-timing (see **Figure 1**, above). In TJ, a mora-timed language, in which

geminates affect vowel phasing, the duration of the entire word is also longer. In contrast, in languages like Italian, which is syllable-timed, a geminate intrudes in the constant V-to-V cycles and does not increase their overall duration. Thus, instead of lengthening the duration of the entire word, gemination results in greater overlap between vowel and consonant gestures.

Despite their intuitive appeal, there are limitations with inter-gestural coupling models, as well as empirical predictions that need additional verification. First, such models do not explain pre-geminate vowel lengthening, which occurs in Japanese. Second, these models are based purely on inter-gestural timing, missing the fact that geminates may also imply spatial differences (see Section 1.4) for the consonantal articulator and, perhaps, the vocalic articulator (Löfqvist, 2006; Morimoto et al., 2025). In contrast, spatial effects could, at least in part, be responsible for timing differences (Löfqvist, 2005). Third, an inter-gestural coupling model predicts that TJ should display a durational hierarchy that goes from zero intervening consonants (C)V∅V to one (C)VCV to an intervening geminate or cluster VC:V or VNCV, a behavior that, to our knowledge, has not been verified empirically.

Meanwhile, issues have arisen with respect to the inter-gestural coupling account of Italian as well. First, the fact that Italian V-to-V intervals are temporally constant has been questioned (Burroni et al., 2024, under review; Zmarich et al., 2011). Similarly, evidence that V-to-V intervals are not constant also comes from potentially reduced V-to-V coarticulation across intervening geminate consonant compared to singletons (Greca et al., 2025). Specifically, this reduced coarticulation hints at reduced overlap of vowels when a geminate intervenes, contra the predictions of inter-gestural accounts. Finally, Italian geminates also display clear kinematic differences when compared to singletons (Burroni et al., 2024, under review).

To sum up, the differences in geminate production between TJ and other languages, like Italian, may be the result of global differences in consonantal and vowel coordination and/or mora vs. syllable timing, but this question requires further empirical work. It is important to note, however, that inter-gestural models would be inadequate if it can be shown that geminate production consistently encompasses not only inter-gestural, but also kinematic and intra-gestural modifications, as some have reported for TJ and other languages. We now turn to this issue.

1.3.2. Intra-gestural/kinematic models

Smith (1992, 1995) already noted that realistic models of geminate duration could only be obtained by specifying longer gestural activation times for geminates.

Similarly, Löfqvist (2005) tested the hypothesis that the longer duration of geminates may be a consequence of different kinematic properties, specifically a more extreme or “virtual” target (Browman & Goldstein, 1990b; Löfqvist, 2005), perhaps accompanied by additional differences in stiffness (Browman & Goldstein, 1990b; Kelso et al., 1986). However, the conclusion he reached was negative: he acknowledged the possibility that geminate control may involve longer

activation times (Löfqvist, 2005). However, the mechanisms behind these longer activations remain unclear.

The discussion of these two articles brings us to a crucial issue. Phonological control of duration at the level of a segment is difficult to capture, even in very phonetic detail-oriented theories of phonology, like Articulatory Phonology (Browman & Goldstein, 1990a), or in theories of speech production, like the DIVA model (Guenther, 2016). Specifically, the kinematics of geminates, characterized by long plateaux phases (Burroni et al., 2024, 2025; Hagedorn et al., 2011; Smith, 1992, 1995), during which speakers sustain maximal constrictions, are particularly challenging to model. Neither Articulatory Phonology (Browman & Goldstein, 1990a) nor the DIVA model (Guenther, 2016) directly specify the control of durational characteristics of articulatory gestures or target regions that relate acoustics to articulation. Thus, direct control of maximal constriction phase, when present, must be taken as the byproduct of additional indirect mechanisms, such as modulations in target position, stiffness, or gestural activation. These changes in gestural parameters, which can model longer activations and differences in kinematic profiles, can also be brought about in a principled fashion.

An account of this type would entail that phonological length, like other suprasegmental phenomena, may involve prosodic modulations. If this is correct, prosodic modulation gestures (Burroni & Tilsen, 2022; Byrd & Krivokapić, 2021; Byrd & Saltzman, 1998, 2003; Saltzman et al., 2008), which can influence the time course of gestural activation (temporal modulation gestures, μ_T), or the gestural parameters themselves (spatial modulation gestures μ_S), could be the driving force behind differences between singletons and geminates, especially in a language like TJ where inter-gestural timing differences seem not to be observed, at least for bilabial and apical geminates (Burroni et al., 2025; Smith, 1992, 1995).

Before considering how geminate production may be controlled in TJ, it is necessary to recapitulate the main empirical findings. A key prediction of the prosodic modulation approach is that effects will be uniform across articulators (except if other independently motivated constraints are involved). This is because the relevant modifications in the AP framework are at a level of planning that is not specific to articulators. We thus need to first address whether geminate articulation displays similar characteristics across articulators.

1.4. The articulation of TJ geminate consonants

Unlike their acoustic manifestations, the articulatory mechanisms involved in TJ geminate production have not been investigated as thoroughly. Open questions remain regarding possible differences between singleton and geminate production in terms of both spatial and temporal properties (Burroni et al., 2025; Fujimoto et al., 2015; Ishii, 1999; Kawahara & Matsui, 2017; Kochetov, 2025; Kochetov & Kang, 2017; Löfqvist, 2006, 2007, 2017; Morimoto & Kitamura, 2019; Smith, 1992, 1995; Takada, 1985).

Turning to empirical findings, electropalatographic (EPG) work has shown that the production of TJ geminates is accompanied not only by longer acoustic durations, but also by more extended linguo-palatal contacts (Kawahara & Matsui, 2017; Kochetov, 2025; Kochetov & Kang, 2017). These findings are compatible with the notion that geminates may involve more extreme or even virtual targets (Burroni et al., 2024; Löfqvist, 2006). That is, targets that lie beyond the vocal tract limit and enforce longer, more extended articulatory contacts. Less consensus has been reached with respect to more fine-grained articulatory properties: intra-gestural (or durational) properties, kinematic (or spatial) properties, and inter-gestural timing properties.

With respect to intra-gestural or durational properties, an outstanding issue is how the longer acoustic duration of geminates are achieved kinematically. Some work has suggested that TJ bilabial and alveolar geminates are produced by maintaining maximal constriction phases, also known as gestural plateaux, for longer periods of time (Smith, 1992, 1995; Tilsen & Hermes, 2020), while others have suggested that lingual geminates are produced with more general slowdowns in articulatory movements to produce longer acoustic durations (Löfqvist, 2007). It is important to tease apart these two hypotheses before proposing models of geminate articulatory control.

With respect to kinematic or spatial parameters, whether geminate production involves other differences beyond more extended articulatory contact is also an unresolved issue in both TJ and other languages (Burroni et al., 2024; Kochetov, 2025; Türk et al., 2017). Geminate production may be expected to involve more extreme articulator positions, greater movement amplitudes, higher peak velocities (related to their greater amplitude), and lower stiffness, a measure related to how quickly articulatory gestures evolve to their targets independently of their amplitude and velocity (Browman & Goldstein, 1990a, 1990b; Byrd & Saltzman, 2003; Fuchs et al., 2011; Löfqvist, 2005; Saltzman & Munhall, 1989).

Yet no unequivocal conclusions have been reached on these matters in TJ. With respect to more extreme articulatory positions, these have been reported for bilabials (Löfqvist, 2005) and for some apical or laminal geminates, /t/ and /ɾ/ for two subjects, but not for other sounds, /s/, /n/, and /d/, and for the other six subjects by Morimoto (2020) in TJ. With respect to movement amplitude, a greater amplitude has been reported for both bilabial (Löfqvist, 2005) and lingual geminates (Löfqvist, 2007). Greater amplitudes were, however, in part driven by different onset positions. The higher peak velocity that may be expected in view of larger amplitudes, however, has not been observed for bilabials (Löfqvist, 2005), with some studies suggesting that average velocity may be slower for lingual geminates (Löfqvist, 2007) and perhaps for bilabial geminates as well (Smith, 1995). Stiffness has rarely been measured in TJ, but work on bilabial consonants suggests that geminates may display the lower stiffness typical of longer articulatory gestures (Browman & Goldstein, 1990b; Kelso et al., 1986). This is the case at least when stiffness is estimated using a regression line relating movement amplitude to peak velocity, as done in previous studies (Löfqvist, 2005).

Finally, with respect to inter-gestural timing properties, as we have anticipated, a well-known and typologically rare feature of TJ is the slightly longer durations of vowels in pre-geminate than in pre-singleton position. This pattern runs opposite to a typologically common pre-geminate shortening effect reported for languages like Italian and many others (Burroni et al., 2024; Fujimoto et al., 2015; Kawahara, 2015; Ladefoged & Maddieson, 1996; Morimoto, 2020; Morimoto et al., 2025; Takeyasu & Giriko, 2017). Early work by Smith (1992, 1995) suggested language-specific coordination preferences as a rationale for the lack of pre-geminate shortening in TJ. Smith's proposal is attractive, but subsequent work has called into question both the existence of inter-gestural timing differences between singletons and geminates (Löfqvist, 2017) and the relationship between pre-geminate vowel durations and the longer time to target for geminates (Fujimoto et al., 2015). Thus, more empirical work on inter-gestural timing properties of geminates is warranted.

To summarize, substantial uncertainties remain regarding intra-gestural, kinematic, and inter-gestural properties that differentiate the production of TJ singleton and geminate consonants. These considerations motivated us to re-examine all of these properties in a recent study.

1.5. Our previous work and the present study

In a previous study (Burroni et al., 2025), we examined the production of TJ mimetic geminates produced by seven speakers in terms of their intra-gestural, kinematic, and inter-gestural properties. Our main finding was that geminate production encompasses differences along all of these three dimensions, as summarized in **Table 1**.

Property	Measure	Observed Pattern
Intra-gestural	CLO Dur.	G > S
	PLAT Dur.	G > S
	REL Dur.	=
Kinematic	CD Max.	G > S
	Mov. Amp	G > S
	Peak Vel.	=
	Stiffness	G < S
Inter-gestural Timing	V_1 Dur.	G > S
	V_1-CLO Lag	=
	V_1-TARG Lag	G > S
	V_1-V_2 Lag	G > S

Table 1: Summary of previous findings by Burroni et al. (2025), where G stands for geminate and S stands for singleton.

With respect to the first dimension, intra-gestural or durational properties, we found that the production of geminate consonants involved slightly longer closing phases (CLO Dur.), much longer plateau phases (PLAT Dur.), and similar release phases (REL Dur.) compared to singleton production. Additionally, using Dynamic Time Warping (DTW) analyses, following Burroni and Tilsen (2022), we found that singletons can be mapped to geminates by locally stretching their plateau phases, suggesting the presence of sustained maximal constrictions. In addition, we observed robust correlations between geminate acoustic durations and articulatory plateau durations, suggesting that longer plateaux are the main articulatory mechanisms underlying longer acoustic durations.

With respect to the second dimension, kinematic or spatial properties, we found that geminates display more extreme vertical positions (CD Max.), greater movement amplitudes (Mov. Amp), no clear differences in peak velocities (Peak Vel.), and lower stiffnesses (Stiffness). All these properties are expected from slower gestures that may be regulated by more extreme targets.

With respect to the third dimension, inter-gestural timing properties, we found that vowels preceding geminates are acoustically longer than vowels preceding singletons. We found that pre-geminate lengthening strongly correlated with the longer times to target of geminate consonants relative to the preceding vowel (V_1-TARG Lag). We also found that the relative timing of preceding vowels and consonant initiations is comparable between both singletons and geminates (V_1-CLO Lag). Finally, we also observed longer acoustic lags between the preceding and following vowel (V_1-V_2) across geminates. All these findings suggest a kinematic basis for pre-geminate lengthening and that the implementation of length contrasts in TJ is accomplished by speakers via delaying the lingustic material directly following the target consonant. This is unlike some other languages—for example, Italian—where geminates are produced by starting the target consonant earlier relative to the preceding vowel (Burroni et al., 2024; Celata et al., 2022; Smith, 1995).

A major limitation of our previous study was that it was based only on consonants produced with the tongue tip and blade, i.e., alveolar and palatal consonants (/t/, /d/, /ɾ/, /z/, /ʑ/ /s/, /t͡s/, /t͡ɕ/). As noted, this is an important limitation if we wish to draw a more general conclusion about articulatory control in geminate production. Specifically, if the control of geminate production may reside in prosodic modulation, we should be able to observe consistent correlates across different places of articulation.

Accordingly, in the present study, we address this limitation and extend our work by considering whether our intra-gestural, inter-gestural, and kinematic findings generalize across a wider set of articulators and places of articulation. These consonants encompass new bilabial (/p/, /b/, /m/), alveolar (/n/), palatal (/ɲ/), and dorsal (/k/) mimetic geminates. These places of articulation cover basically all POAs in Japanese, with the exception of glottal consonants, which,

however, cannot be studied with EMA. In addition, based on our findings, we attempt to flesh out the relationships existing between intra-gestural, kinematic, and inter-gestural properties and how different phonological accounts may capture them, specifically focusing on TJ.

2. Methods

2.1. Participants

Six native Japanese speakers (two male, four female) participated in the experiment. All reported normal speech and hearing and daily use of the Tokyo dialect. Participants' experimental sessions reported in this study are new and collected specifically for this work (cf. Burroni et al., 2025, for a previous study with a different participant pool).

2.2. Materials

The speech material consisted of 15 existing Japanese mimetic words, three items for each of five consonantal types covering bilabial, alveolar, palatal, and dorsal places of articulation; see **Table 2**. Each target consonant was produced as a singleton or a geminate (for emphasis). All the target consonants were stops, either oral or nasal.

The choice of mimetic words over lexical singleton-geminate minimal pairs reflects a methodological preference for experimental control. Mimetic minimal pairs vary only the gemination of the target consonant within the same lexical item, thereby controlling for numerous factors, including syllable count, word length, segmental context, lexical statistics, and even meaning, factors known to modulate articulatory properties (Gahl & Baayen, 2024; Shaw & Kawahara, 2019; Tomaschek et al., 2018). Lexical minimal pairs, by contrast, involve different words with different meanings, potentially different frequencies, and semantic-syntactic distributions, all of which may introduce confounds that are difficult to disentangle from the contrast of interest. Mimetic gemination is also a fully productive phonological process in TJ (Kawahara, 2013), ensuring that participants can readily produce geminate renditions of the target words.

Target	W_1	W_2	W_3
bilabial stop	s**up(:)ɑ**supɑ	kʲ**ɑp(:)i**kʲɑpi	k**ib(:)ɑ**kibɑ
bilabial nasal	m**om(:)i**momi	g**ɑm(:)i**gɑ mi	ɕ**im(:)ɑ**ɕimɑ
alveolar nasal	k**un(:)e**kune	k**on(:)ɑ**gonɑ	h**en(:)ɑ**henɑ
palatal nasal	g**uɲ(:)ɑ**guɲɑ	g**oɲ(:)o**goɲo	h**eɲ(:)ɑ**heɲɑ
velar stop	p**ik(:)ɑ**pikɑ	t͡ɕ**ik(:)ɑ**t͡ɕikɑ	p**ɑk(:)i**pɑki

Table 2: List of the singleton set of the target words. The segments in bold contain the target consonant and the preceding and following vowel.

In each trial, participants produced singleton mimetic words in isolation, followed by their emphatic geminate variant also produced in isolation. Items were elicited in isolation to minimize total experiment duration and confounds that could arise from different prosodic realizations of a carrier phrase. The target consonants were never adjacent to word boundaries, thus also limiting the need for a carrier phrase. The order of lexical items was fully randomized. In total, 15 unique words × 2 realizations (singleton/geminate) × 10 repetitions × 6 speakers yielded a total of 1800 tokens for analysis.

2.3. Experimental procedures

Articulatory movements were captured using a Carstens AG501 system sampling at 1250 Hz. Acoustic data was simultaneously recorded at 25.6 kHz using a Sennheiser ME66 super-cardioid shotgun microphone positioned around 25 cm outside the EMA field. The sensor setup was as follows: five tongue sensors (three on the sagittal midline, two parasagittally). Additional sensors were placed on the lips, jaw, left/right mastoid processes, nasion, and maxilla. Participants sat in a sound-attenuated room; stimuli were displayed in Japanese orthography on a monitor 0.5 cm outside the EMA field. Stimulus presentation was controlled using MATLAB, enabling real-time monitoring for errors. Rare hesitations or mispronunciations were marked and reinserted among the remaining items.

Head movements were corrected computationally after data collection with reference to three sensors on the head, the sensor on the maxilla, and the two sensors on the biteplane using an orthogonal Procrustes transformation. The head corrected data were rotated so that the origin of transverse (inferior-to-superior) and coronal dimensions (medial-to-lateral) was based on the occlusion sensor at the front teeth, and the origin of the sagittal plane (anterior-to-posterior) was based on the upper incisor sensor.

2.4. Data processing and analyses

The acoustic signals were manually segmented at the word and segmental level in PRAAT (Boersma & Weenink, 2025) using standard criteria based on waveform and the spectrogram characteristics by the first and second authors and research assistants trained in phonetics. The segmentation was later double-checked by the first and second authors. The acoustic segmentation was used as a starting point for articulatory landmarking of target consonants. The only acoustic boundaries used in the analyses involve the acoustic duration of the target consonants and preceding and following vowels. Given the wide variety of vocalic contexts and the impossibility of consistently identifying vocalic landmarks from the kinematics alone, we adopt acoustic landmarking as events that reasonably correlate with kinematic events, especially target achievement, following previous and our own work, e.g., (Browman & Goldstein, 1988; Burroni et al., 2025).

All articulatory signals were smoothed using Kaiser passband filters with passband edges at [5 15] Hz (for the reference sensors), at [40 50] Hz (for the TT sensor), and at [20 30] Hz (for all other articulators). The damping in the stop band was 60 dB for all sensors. The sampling rate of EMA sensors was kept at 1250 Hz for all articulators.

The articulatory analyses reported in this paper focus on various places of articulation; bilabial consonants were measured based on the LA signal, the 3D Euclidean distance between the upper and lower lip sensor, alveolar consonants based on the tongue tip sensor, palatal consonants based on the tongue body sensor, and dorsal consonants based on the tongue forsum sensor. Consonants appear in a wide variety of vocalic contexts, and thus there is a strong influence of vowels on sagittal (anterior-to-posterior) tongue position. For this reason, our landmarking was based on the dimension that most reliably identifies consonantal articulation, namely the vertical movement (cf. Burroni et al., 2025; Shaw & Kawahara, 2018, among many others for a similar choice).

Consonantal gestures were automatically landmarked using a custom MATLAB routine *landmarkTraj()* (Burroni & Tilsen, 2025; Burroni et al., 2025) based on peak velocity thresholding. Acoustic boundaries were used to locate the consonant midpoint, which defined a symmetric window spanning twice the segment's acoustic duration. Within this window, velocity extrema were identified for upward closing and downward release movements in the relevant articulator's trajectory. Gestural onset and target were defined as the first and last timepoints surpassing 20% of peak velocity during closing phase, while release and offset were similarly defined for release velocity. Note that for dorsal consonants gestural parsing likely reflects combined consonantal and vocalic movements because the articulator is shared. We adopted the provision to parse the largest identifiable tongue dorsum constrictions, i.e., vertical movement, identically across singleton and geminate in a trial. Thus, our estimates may not perfectly isolate consonants in time, yet, (i) we were consistent across singletons and geminates, our distinction of interest, and (ii) could derive important parameters undeniably related to consonants, such as maximum articulator position (and amplitude) and other kinematic parameters.

All landmarks were visually inspected and adjusted for a small subset of tokens, ~50 tokens, as needed. Twenty-one tokens without clear gestural boundaries were excluded, along with their paired singleton or geminate, resulting in the exclusion of 42 tokens (~2% of the data). Excluding the paired token was necessary because some of our analyses require pairing singleton and geminate obtained from the same trial, as described below. After visual inspection of landmarks of all data points, no further outliers were excluded.

From the acoustic and articulatory signals, we derived a set of intra-gestural timing, kinematic, and inter-gestural timing measures. For intra-gestural timing, we investigated:

1. the closing phase duration (CLO Dur.), defined as the lag between gestural onset and target
2. the plateau duration (PLAT Dur.), defined as the duration of the lag between gestural target and release
3. the release phase duration (REL Dur.), defined as the lag between gestural release and offset

We also investigated the relationship between plateau duration and acoustic consonant duration using correlation following our previous work (Burroni et al., 2025).

Additionally, we also developed holistic analyses of the kinematic trajectories that rely on Dynamic Time Warping (DTW). Specifically, we took each singleton/geminate pair in a trial and used DTW to derive a pairwise warping function that identifies which portions of a singleton articulatory trajectory need to be stretched to "derive" a geminate. We inspected the shape of the warping functions and obtained localized average warps over normalized duration, following previous work (Burroni & Tilsen, 2022; Burroni et al., 2025).

For kinematic properties, we investigated:

1. maximal degree of constriction (CD Max), defined as either the maximum vertical articulator position or, for lip aperture only, the minimum value. This measure was z-scored within articulator sensor to allow for comparison across articulators that have different positions in the EMA field. For bilabials, z-scored CD Max was inverted, i.e., multiplied by –1, to ensure that higher z-scored values always indicate great articulatory contacts across all POAs
2. magnitude of peak velocity during closing phases (Peak Vel.), again z-scored to allow for comparison across articulators, which, again, have different velocities
3. magnitude of movement amplitude from onset to maximum constriction during closing phase (Mov. Amp)
4. kinematic stiffness (Stiffness), defined as the ratio of peak velocity to movement amplitude, providing a measure of movement speed normalized by amplitude

For inter-gestural timing properties, we investigated:

1. the duration of the vowel preceding the target consonants estimated from acoustics (V_1 Dur.)
2. the duration of the lag between the preceding vowel acoustic onset and the consonantal gesture articulatory onset (V_1-CLO Lag)

3. the duration of the lag between the preceding vowel acoustic onset and the consonantal gesture articulatory target (V_1-TARG Lag)
4. the duration of the trans–consonantal lag between the preceding and following vowels acoustic onsets (V_1-V_2 Lag)

We also investigated the relationship between consonantal and preceding vowel duration using correlations, as in previous work (Burroni et al., 2025; Fujimoto et al., 2015).

2.5. Statistical analyses

All intra-gestural, kinematic, and inter-gestural properties were analyzed using linear mixed-effects regression models in MATLAB's *fitlme()* function. The effect of gemination was assessed via log-likelihood ratio tests, comparing a baseline model with a term for place of articulation (POA, reference set to alveolars) to an alternative model including an additional fixed effect for gemination (reference coded as singleton). Models featured the same random effect structure, with by-subject and by-item random intercepts and by-speaker and by-item random slopes for gemination (Barr et al., 2013), the main term of interest, as well as for POA, when the models were able to support the inclusion. If this first comparison resulted in the simpler model being preferred, we also tested the effect of place of articulation against an intercept only model. Conversely, if the more complex model with effects of both place of articulation and gemination was preferred, we tested whether including an interaction between gemination and POA was warranted. We do not explicitly report all pairwise differences of POA not related to gemination, as these are beyond the scope of our paper. The interested reader can obtain them from our OSF materials, which contain models where the intercept level has been rotated across all POAs so that all possible pairwise estimates can be computed.

Correlations were calculated using MATLAB's *corr()* function and the correlation types set to Spearman rank correlation. We chose Spearman correlation because this method is less sensitive to outliers than Pearson correlation.

All data and scripts necessary to replicate our analyses and figures are publicly available online in an OSF repository.[2]

3. Results

3.1. Intra-gestural properties

For closing phase duration, **Figure 2** top right, we found a singleton intercept value of 70 ms (SE 11 ms) for alveolars, bilabials, and palatals, with no significant differences among these POA. Dorsals have longer closing phases by +39 ms (SE 18 ms). Geminates have longer

[2] https://osf.io/4tk8q/overview.

closing phase durations than singletons with an effect size estimated at +27 ms (SE 7 ms). No significant interaction between geminate and place of articulation was observed on closing phase duration.

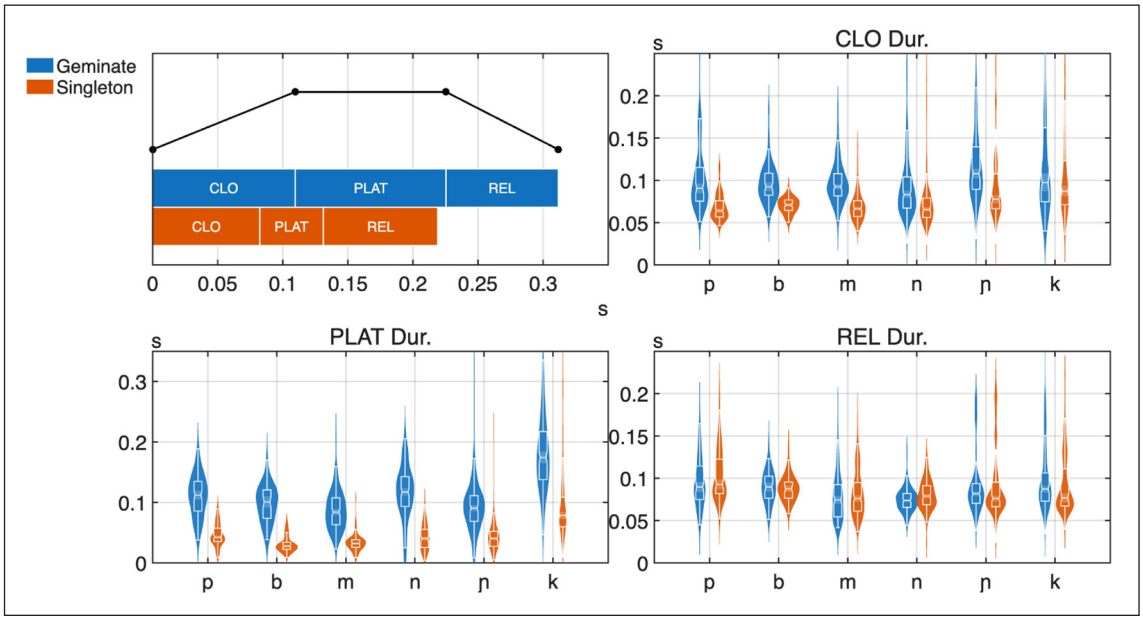

Figure 2: Top left: Schematic illustration of intra-gestural phases and their mean durations. **Top right**: Boxplot with superimposed Gaussian kernel density estimate (kde) of CLO duration. **Bottom left**: Boxplot of PLAT duration. **Bottom right**: Boxplot of REL duration.

For plateau duration, **Figure 2** bottom left, we found a singleton intercept value of 41 ms (SE 7 ms) for alveolars, bilabials, and palatals. Dorsals have longer plateau durations by +48 ms (SE 10 ms). Geminates have longer plateau durations than singletons, with an effect size estimated at +75 ms (SE 10 ms). We also found a significant interaction between gemination and POA: the effect of gemination on plateau duration is reduced for geminate palatals, with a difference estimated at –26 ms (SE 11 ms).

For release duration, **Figure 2** bottom right, estimated at 91 ms (SE 10 ms), we found no significant effects of either gemination or POA.

The fact that singletons and geminates differ primarily in terms of their plateau duration is also evident from DTW (Dynamic Time Warping) analyses. When each mimetic singleton is stretched to its geminate counterpart, a strong distortion of time is observed around the midpoint of the consonant, 0.3 to 0.6 of its proportional duration (**Figure 3F**). Each singleton sample between 31% and 63% of the singleton trajectory is repeated between 1.5 to almost 3 times to derive a geminate, indicating a stretching of the plateau region (**Figure 3G**).

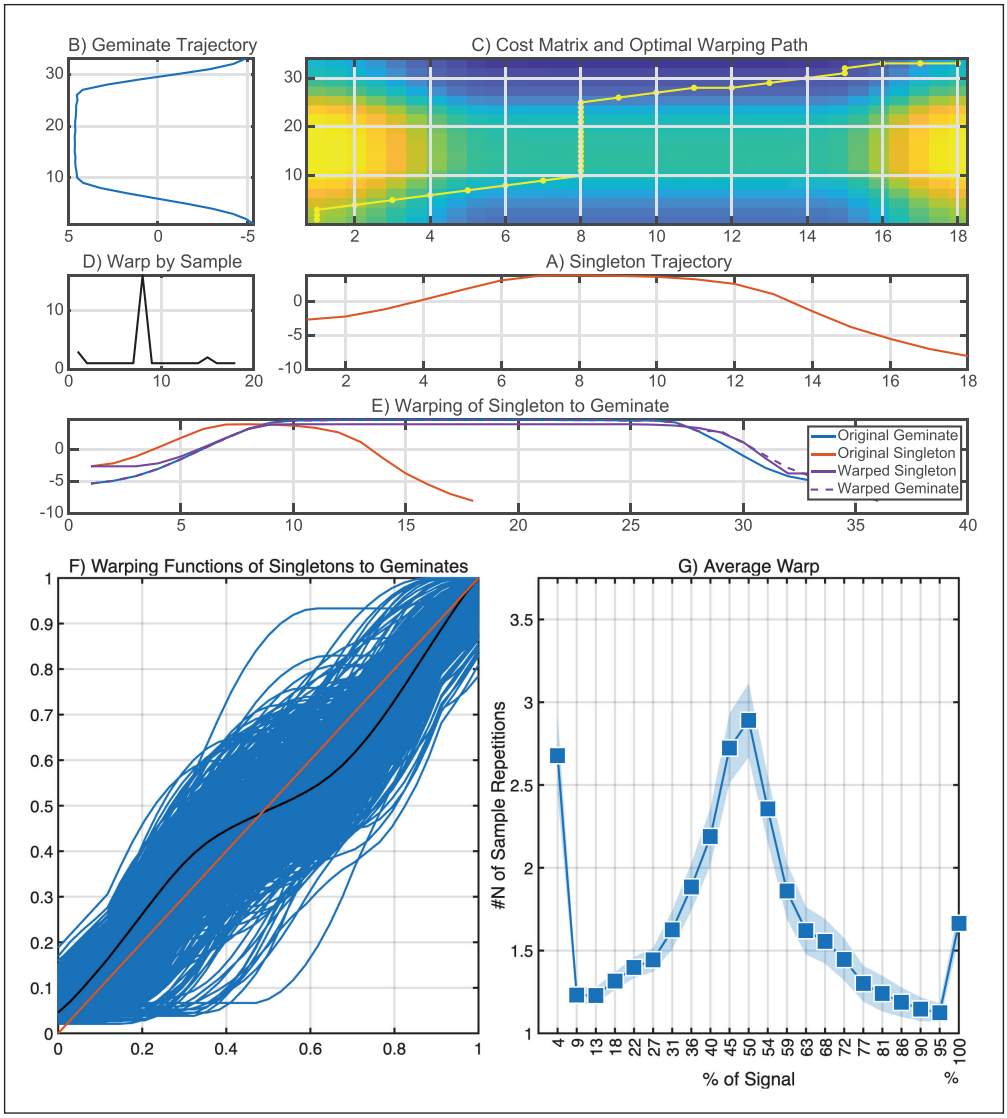

Figure 3: A: Example of articulatory vertical movement (Tongue Tip) for singleton /n/. **B**: Example of articulatory vertical movement (Tongue Tip) for geminate /n:/. **C**: Cost matrix and optimal warping path for singleton to geminate alignment. **D**: Warp function showing the number of repetitions each sample undergoes to stretch singleton to geminate. **E**: Alignment of singleton and geminate TTy trajectories. **F**: Warping function of singleton to geminates (blue lines), with average warping functions (solid black line) of singleton to geminate showing a strong distortion of linear time (orange line) in the plateau region (0.2-0.6). Time is normalized between 0 and 1. **G**: Average warp at different % of the trajectory.

The observed longer acoustic duration for geminates is closely related to the longer plateau, as indicated by a robust correlation between target consonant acoustic duration and plateau duration ($\rho = 0.75$, $p < 1 \cdot 10^{-318}$), **Figure 4**.

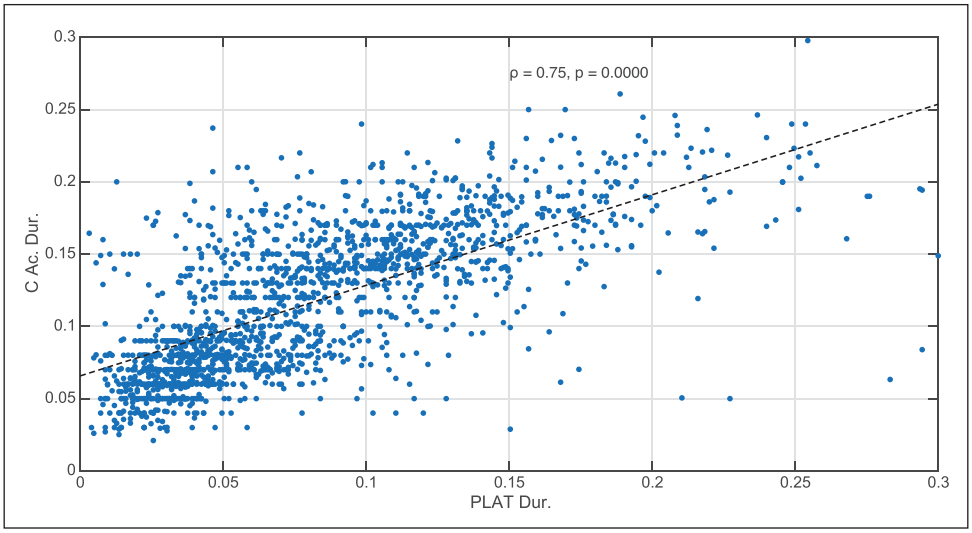

Figure 4: Spearman's Rank Correlation between articulatory plateau duration and acoustic consonant duration.

3.2. Kinematic properties

For maximum articulator position, **Figure 5** top left, a proxy for Constriction Degree, we found an intercept value of 0.76 z-scores (SE 0.05) for alveolar, bilabials, palatals, and dorsals. Geminates are produced with more constricted targets +0.15 z-scores (SE 0.56). No interaction between POA and gemination was observed.

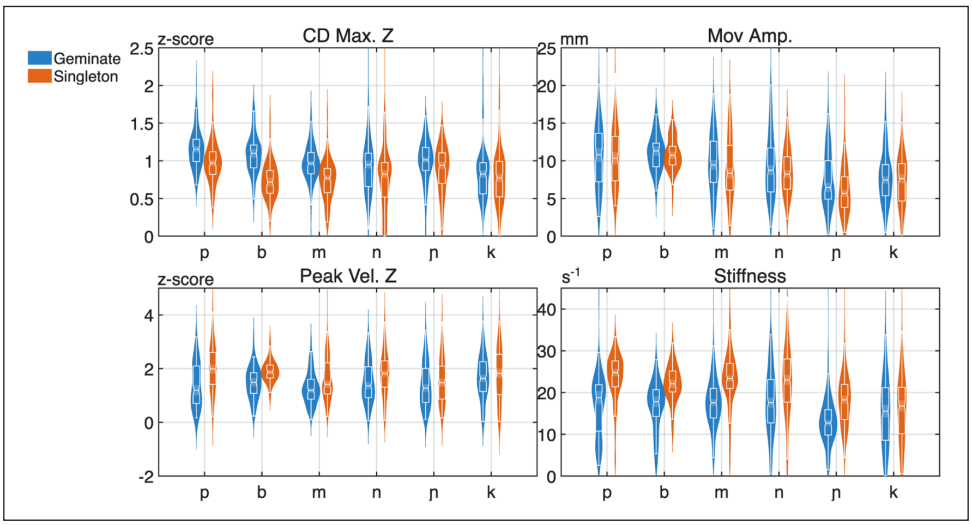

Figure 5: Boxplots with superimposed Gaussian kde of **Top left**: Maximum articulator values. **Top right**: Boxplot of movement amplitude values. **Bottom left**: Boxplot of CLO peak velocity values. **Bottom right**: Boxplot of stiffness values.

For movement amplitude, **Figure 5** top right, we found an intercept value of 8.5 mm (SE 1.5) for alveolar, bilabials, palatals, and dorsals. Geminates are produced with greater movement amplitude by +0.7 mm (SE 0.3 mm).

For movement peak velocity, **Figure 5** bottom left, we found an intercept value of 1.8 z-scores (SE 0.34) for alveolar, bilabials, palatals, and dorsals. Geminates are produced with slower movements with an effect size of –0.29 z-scores (SE 0.09 z-scores).

For movement stiffness, **Figure 5** bottom right, we found an intercept of 22.6 s^{-1} (SE 2 s^{-1}) for alveolars, bilabials and palatals, dorsals have lower stiffness, with an effect size of –6.5 s^{-1} (SE 2.6 s^{-1}). Geminates are produced with less stiff movements, with an effect size of –4.66 s^{-1} (SE 1.2 s^{-1}). Interactions between POA and gemination were also observed: the effect of gemination is basically absent for dorsals, with an effect size of +3.66 s^{-1} (SE 1.47 s^{-1}) counteracting the effect of gemination.

3.3. Inter-gestural timing properties

For V_1 Dur., **Figure 6** top left, we found an intercept of 49 ms (SE 10 ms) for alveolar, bilabials, palatals, and dorsals. Vowels preceding geminates are longer +33 ms (SE 4 ms) in the case of alveolars, but the effects are less strong for other POA contexts: the effect is reduced for bilabials with an effect size of –10 ms (SE 3 ms), for dorsal with an effect size of –19 ms (SE 4 ms) and, marginally, for palatal geminates with an effect size of –7 ms (SE 4 ms).

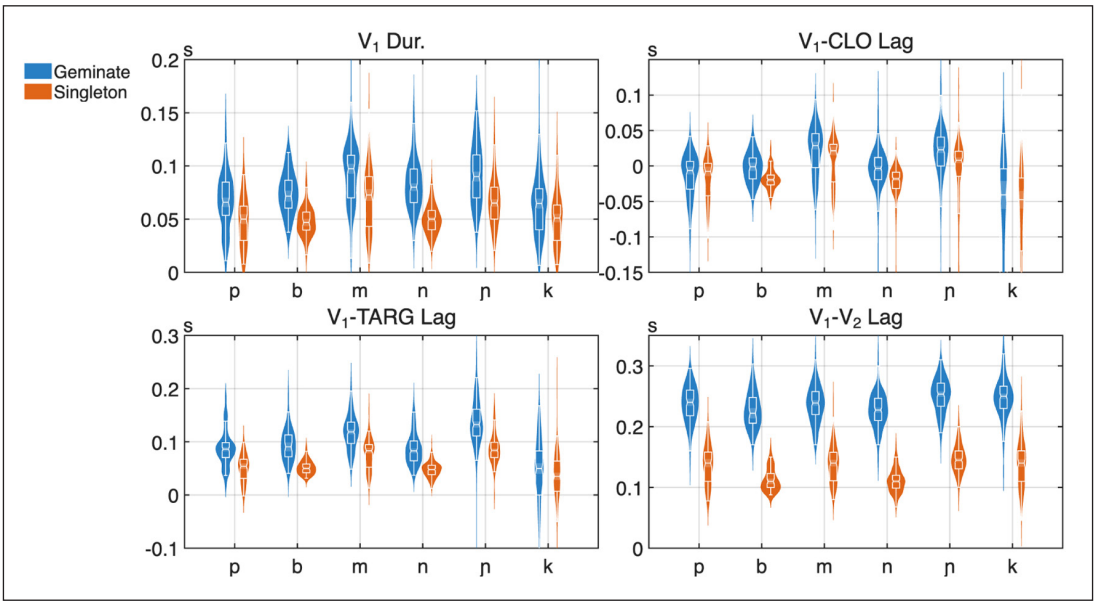

Figure 6: Boxplot with superimposed Gaussian kde of maximum inter-gestural timing values. **Top left**: V_1 Dur. **Top Right**: V_1-CLO lag. **Bottom left**: V_1-TARG lag. **Bottom right**: V_1-V_2 lag.

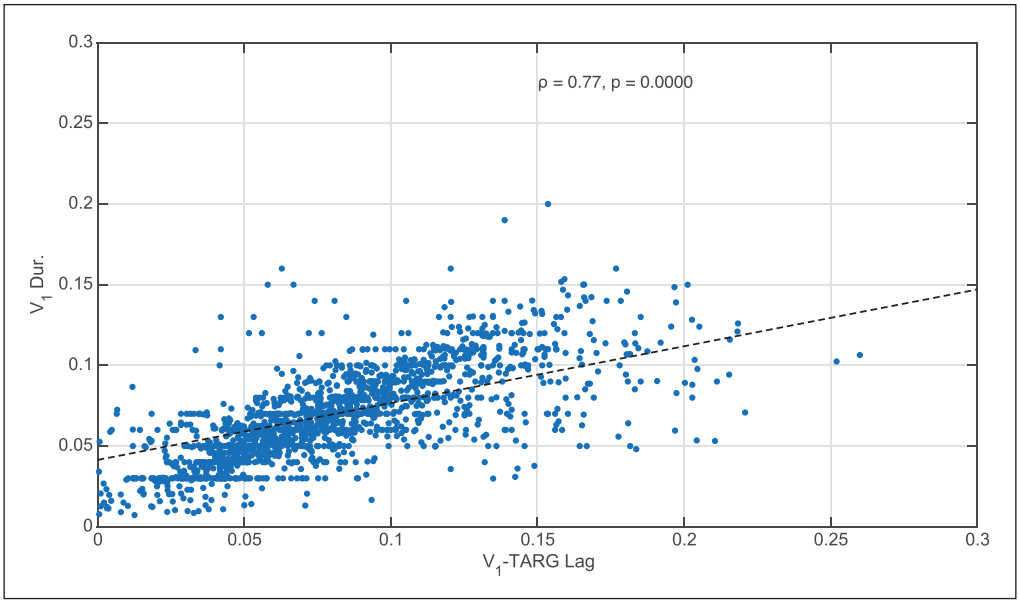

Figure 7: Spearman's Rank Correlation between acoustic V_1 duration and V_1-TARG Lag.

We found no effects of POA or gemination on the V_1-CLO Lag, which is estimated at –17 ms (SE 11), **Figure 6** top right.

For V_1-TARG Lag, **Figure 6** bottom left, we found an intercept of 47 ms (SE 14 ms) for alveolar, bilabials, dorsals and palatals. Geminates have longer lags by +39 ms (SE 7 ms). Interactions between POA and gemination were also observed: the effect is basically absent for dorsal geminates, where the effect size is estimated at –33 ms (SE 8 ms).

For the V_1-V_2 Lag, **Figure 6** bottom right, we found an intercept of 111 ms (SE 12 ms) for alveolar, bilabials, and dorsals. The lag is longer for palatals by +35 ms (SE 15). When the intervening consonant is a geminate the V_1-V_2 Lag is longer by +107 ms (SE 7 ms). No interactions between gemination and POA were observed.

Given the findings presented above, we hypothesized that the longer duration of vowels preceding geminates may be due to a longer time to target compared to singletons, effectively allowing for a longer period for vowel production. Evidence in favor of this hypothesis comes from a robust Spearman's Rank Correlation ($\rho = .77$, $p < 1e$–307) observed between time to target and preceding vowel duration, cf. **Figure 7**. In other words, the longer the time to target, the longer the acoustic duration of the preceding vowel.

4. Discussion

4.1. Summary of findings

In this paper, we investigated the production of TJ geminate consonants. Building on our previous work (Burroni et al., 2025), we extended our analyses to a broader range of places of articulation

to determine whether the intra-gestural, kinematic, and inter-gestural differences observed for alveolar geminates generalize to other places of articulation.

In terms of intra-gestural properties, we found that TJ geminates, across all places of articulation, are characterized by slightly longer (27 ± 7 ms) closing phases and much longer plateau phases (75 ± 10 ms). The importance of longer plateau phases of geminates is also evident when considering the warping functions of singletons to geminates, which are characterized by local "stretches" of time during plateau phase. The longer plateaux of geminates robustly correlate with their longer acoustic durations. For release phases, on the other hand, we observed no substantial durational differences between singletons and geminates.

In terms of kinematic or spatial properties, we found that TJ geminates, across all places of articulation, are produced by more extreme articulator positions, greater movement amplitude, slower peak velocity, and lower stiffness.

Finally, in terms of inter-gestural timing properties, we found that TJ geminates, across all places of articulation, are characterized by longer preceding vowels V_1 Dur. (the effect being weaker for dorsals), longer V_1-TARG Lags (again, the effect being weaker for dorsals), and V_1-V_2 Lags. V_1-CLO Lags are not different for singletons and geminates, suggesting a similar timing between V_1 and the following consonant, regardless of gemination.

While we identified articulatory properties of TJ geminates that generalize to all places of articulation, two place-specific effects should be acknowledged. First, for the palatal place of articulation, geminate plateau lengthening is reduced by approximately one-third of its overall effect (–26 ms against an overall effect of +75 ms). A possible explanation lies in the high degree of articulatory precision and coarticulatory resistance that characterizes palatal consonant production (cf. Recasens, 1990). Under the Degree of Articulatory Constraint (DAC) framework, palatals are among the most constrained consonants, requiring tight tongue body-palate constriction (Recasens & Espinosa, 2009), a fact that may limit the degree to which their production is susceptible to durational modulations.

Second, for the dorsal place of articulation, the observed differences are attenuated or sometimes absent. This behavior of dorsals is likely due to difficulties in isolating consonantal gestures produced with the tongue dorsum, as this articulator is also used for vowels. The resulting temporal and spatial properties of dorsal gestures thus invariably reflect both consonantal and vocalic articulation, and consequently, their measurement and interpretation are difficult. Nonetheless, many of the properties that are most clearly related to consonantal articulation, such as intra-gestural properties, more extreme articulator position, greater amplitude, and lower velocity, also emerged for dorsals. Future work should investigate further the properties of TJ dorsal geminates, likely with flanking low vowels to facilitate the identification of the tongue dorsum trajectory more closely connected with consonantal articulation. Nonetheless, we would like to emphasize that the shared articulator problem can never be entirely bypassed.

Despite these considerations regarding dorsal gestures, in general, our findings align strikingly with our previous results (Burroni et al., 2025), **Table 3**. In this sense, the current study replicated the previous findings with a new experimental pool.

Property	Measure	B&Al25	Current study
Intra-gestural Durational	CLO Dur.	G > S	G > S
	PLAT Dur.	G > S	G > S
	REL Dur.	=	=
Kinematic Spatial	CD Max.	G > S	G > S
	Mov. Amp	G > S	G > S
	Peak Vel.	=	G < S
	Stiffness	G < S	G < S
Inter-gestural Timing	V_1 Dur.	G > S	G > S
	V_1-CLO Lag	=	=
	V_1-TARG Lag	G > S	G > S
	V_1-V_2 Lag	G > S	G > S

Table 3: Summary of findings of the present study compared with Burroni et al. (2025) (B&Al25).

The observed differences between singletons and geminates, in terms of intra-gestural, kinematic, and inter-gestural timing properties, are virtually identical in the two studies. The present work differs from our previous study (Burroni et al., 2025) only in peak velocity. Our previous study found no difference, while the present work suggests that geminates may be slower. There are at least three possible sources for this discrepancy.

One issue may be the statistical modeling of peak velocity. Our previous study did not explicitly model differences across segment types (coronal stops, fricatives, affricates); manner was only indirectly accounted for through random effects of speaker and item. Moreover, peak velocity was not z-scored by segment type and speaker; in the present study, z-scoring was necessary to compare articulators with different spatial ranges. Thus, any gemination effects on peak velocity may have been missed due to the lack of fixed effects for manner and/or z-score normalization.

A second issue is that we cannot exclude the possibility that the differences may be specific to either a different speaker pool or different lexical items.

Third, in our previous study, we also employed a two-step smoothing procedure on the articulatory trajectories using a robust algorithm based on the Discrete Cosine Transform (Garcia, 2010) and Savitzky-Golay filtering. These steps may also have affected our estimates of peak velocity.

Ultimately, only additional work can clarify whether TJ geminates have lower peak velocity. We believe, however, that it is quite likely that TJ geminates are, on average, slower gestures than singletons due to their longer plateau phases, where articulators display little to no movement, as also reported in previous work (Löfqvist, 2006). Whether this also applies to a specific velocity instant, like the peak velocity time point, requires further investigation.

Notwithstanding the presence of possible peak velocity differences, the present study and our previous one agree strikingly in both qualitative and quantitative terms on the intra-gestural, kinematic, and inter-gestural timing patterns that characterize the production of geminate consonants in TJ.

More broadly, the present findings are also consistent in direction with previous EMA work on TJ geminates (Löfqvist, 2005, 2006, 2007), which reported longer movement paths, lower average speeds, and lower stiffness. Unfortunately, direct magnitude comparisons were unsuitable due to various factors, such as differences in sample size, statistical approach, and the lack of effect size reports in previous work.

We now turn to the relationship among the intra-gestural, kinematic, and inter-gestural dimensions that characterize TJ geminate production, and to the articulatory and phonological accounts that may explain them.

4.2. Articulatory control in TJ geminate production

Our findings show that TJ geminate production involves intra-gestural, kinematic, and inter-gestural specifications. Specifically, intra-gestural effects include longer closing phases and, especially, plateau durations. Geminate production is also generally characterized by more extreme articulator positions, greater movement amplitude, and lower velocity, resulting in less stiff gestures. Finally, inter-gestural timing signatures are also present. As geminates are less stiff, the V_1-TARG Lag is also longer in geminates. The V_1-V_2 Lag is also longer, a fact that follows from the increase in plateau duration for geminates.

A question that arises is whether intra-gestural, kinematic, or inter-gestural properties hold similar importance in the production of singleton and geminate consonants or whether some properties may “drive” the others.

The question is not easy to assess directly. We can, however, have at least an indirect answer based on how well these properties can be used to classify singleton and geminate production as a proxy for their importance in production. As in previous work (Al-Tamimi & Khattab, 2015; Burroni et al., 2020; Hirata & Amano, 2012; Idemaru & Guion, 2008), we trained Linear Discriminant Analysis (LDA) models to classify singletons and geminates to gauge the importance of different properties. We trained three models based on properties that differentiate the two classes along the intra-gestural (C Acoustic Dur., CLO Dur., PLAT Dur., REL Dur.), kinematic (CD

Max., Mov. Amp, Peak Vel., Stiffness), and inter-gestural (V_1 Dur., which strongly correlates with V_1-CLO Lag, V_1-TARG Lag, V_1-V_2 Lag) dimensions. The accuracy we report below was obtained from testing using five-fold cross-validation. The LDA models were trained on the three sets of properties separately. Specifically, each model was trained using 80% of the data and tested on the remaining unseen 20%. The procedure was repeated five times over the entire dataset to obtain a more balanced picture of accuracy. The accuracy over the five folds are displayed in **Figure 8**. The average accuracy for intra-gestural properties was 94.6%, for kinematic properties it was 66.16%, and for inter-gestural properties accuracy it was 96.7%.

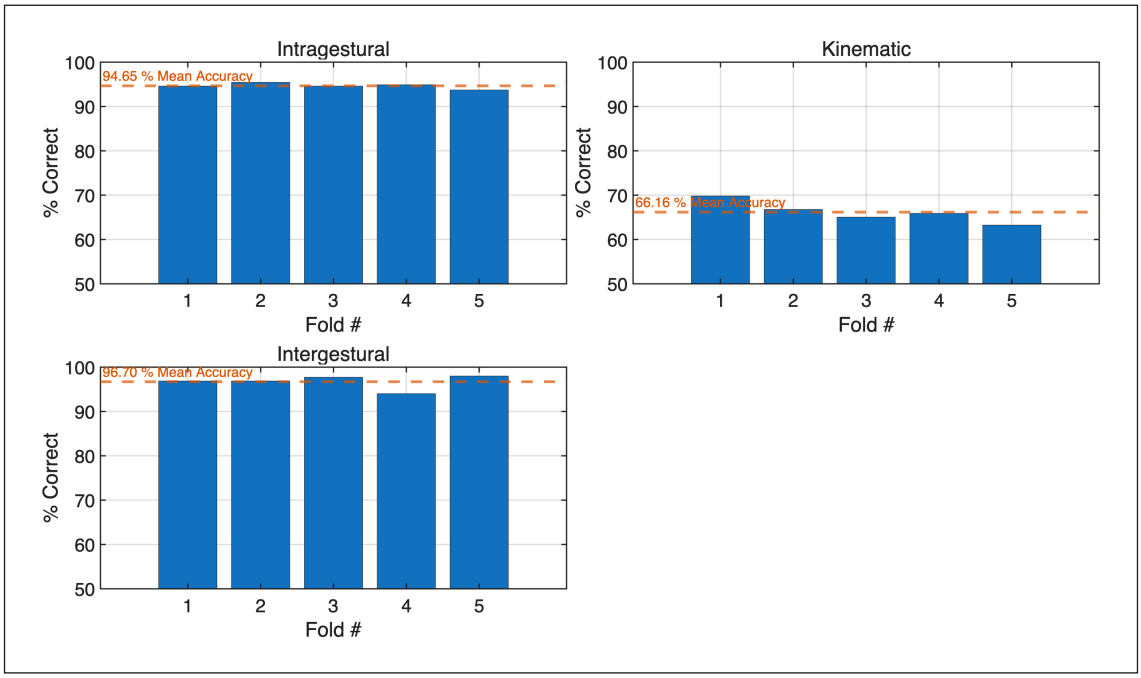

Figure 8: Accuracy in testing of LDA models obtained from five-fold cross-validation. Each bar represents a fold, the orange dashed line indicates average accuracy over all five folds. **Top left**: Accuracy of the LDA model trained using intra-gestural properties. **Top Right**: Accuracy of the LDA model trained using kinematic properties. **Bottom Left**: Accuracy of the LDA model trained using inter-gestural properties.

These results suggest that the production of geminates relies more robustly on intra- and inter-gestural timing properties compared to kinematic properties. That is, even though spatial differences between singletons and geminates were observed, these do not seem robust enough to support the hypothesis that longer durations are the consequence of spatial control, for example, of more extreme or virtual targets (Gafos et al., 2011; Löfqvist, 2005). In particular, the higher velocity predicted to accompany more extreme or virtual targets is not observed (Löfqvist, 2005). On the contrary, TJ geminates seem to be produced either more slowly or with velocity similar to

that of singletons. These findings align with previous work (Löfqvist, 2005), which also suggested that geminates are likely not simply epiphenomenal to more extreme spatial properties.

It should be noted that the primacy of temporal over non-temporal correlates has also been documented in other languages, with a caveat that, in these languages, non-durational correlates appear to be stronger predictors than in TJ. For example, in Lebanese Arabic, non-temporal cues yield better classification accuracies than kinematic parameters do in TJ, underscoring that the weighting of temporal and non-temporal dimensions in the singleton-geminate contrast may be language-specific (Al-Tamimi & Khattab, 2015, 2018).

If spatial control is unlikely to be the correct option, how can TJ speakers control their articulators to produce geminate consonants? To answer this question, we consider in more detail the production strategies we have reported. Our results show that TJ speakers produce geminate consonants by altering the duration of closing phases and, especially, plateau phases and by further modifying the timing of overlapping vocalic gestures. Both characteristics, which emerged from our statistical analyses, can readily be observed in the production of singleton and geminate tokens for each individual word pair by different speakers. This is illustrated in **Figure 9** (top panels), where time 0 represents the articulatory closing phase onset. In this figure, each trajectory represents a token containing a medial bilabial singleton/geminate consonant with a flanking a_i or i_a vocalic environment with color coded singleton or geminate trajectories. This is the environment where the production can more readily be seen since the consonant and the vowel are produced with maximally independent articulators, the lips and the tongue. The longer plateaux of geminates are evident in the top panels of **Figure 9** in the form of longer stretches of lower Lip Aperture (LA) values, starting around timepoints 0.05 to 0.1 for both singletons and geminates. Note that for geminates the stretch of lower values is longer, indicating longer bilabial closures. The different vocalic gesture articulation is evident in the bottom panels of **Figure 9** in the form of slower and delayed a-to-i tongue body raising (panels 1 and 2) and i-to-a tongue body lowering (panel 3) movements happening roughly at the same time as the bilabial closure.

The observations above suggest a generalized slowdown of both consonantal and vocalic articulation, as noted in previous work (Ishii, 1999; Löfqvist, 2006, 2007). In addition, clear, albeit not nearly as robust, spatial differences were also observed between singleton and geminate contexts for consonants. However, previous work suggests that spatial differences exist for vocalic gestures, too (Löfqvist, 2005; Morimoto et al., 2025; Takada, 1985). This seems to be the case at least for TB height for aC:i vs. aCi in the bottom panels of columns 1 and 2 of **Figure 9**. If so, the production of TJ geminate consonants relies not only on inter- and intra-gestural temporal adjustments, together with more subtle kinematic differences, but also on effects that extend across consonantal and vocalic articulators. In the following section, we discuss how these effects could be explained.

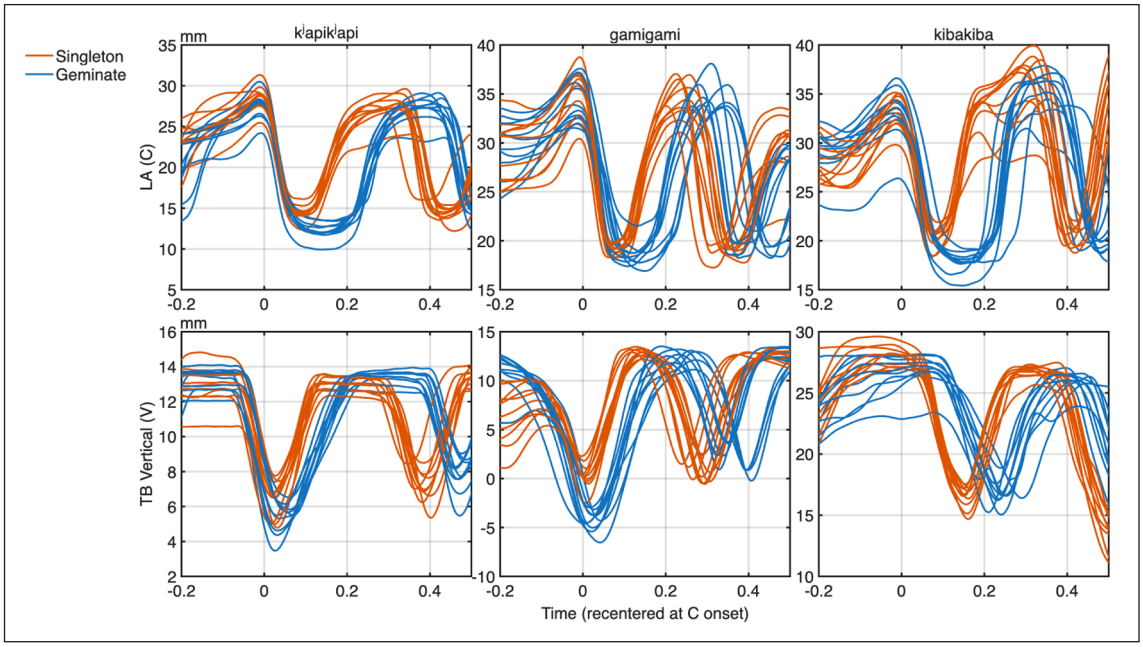

Figure 9: Kinematic trajectories of consonantal and vocalic articulation for different words (titles) produced by different speakers (in parenthesis). **Top**: Trajectories represent Lip Aperture, the 3-D distance between the lips. Lower values indicate constrictions. **Bottom**: Trajectories represent tongue dorsum height, a correlate of vowel height, lower values indicate lower tongue positions. For both top and bottom panels time zero represents the articulatory onset of consonantal gestures estimated from our landmarking procedure.

4.3. Phonological and phonetic models of geminate articulatory control

Most phonological approaches, both for TJ and cross-linguistically (Ham, 2002; Kawagoe, 2015; Lahiri & Hankamer, 1988; Ridouane, 2003; Shibatani, 1990; Vance, 1987), treat duration as the primary or even the only dimension of geminate contrasts. This approach underspecifies the exact phonetic realization, including whether there will be accompanying kinematic or inter-gestural differences. Longer phonetic durations could be due to the presence of an additional moraic phoneme /Q/ or the presence of an additional feature [+long] in the phonological representation. Additional phonetic properties cannot be captured directly, so the matter must be relegated to TJ language-specific phonetics.

A step towards more isomorphic phonological and phonetic representation can be offered, we believe, by a more direct encoding of the dependency between temporal and spatial properties in geminate production. One possibility is to consider temporal properties primary, and spatial properties a result of enhancement (Keyser & Stevens, 2006). Such an approach was developed by Ridouane (2007), who proposed that geminate production in Tashlhiyt Berber involves primary temporal correlates, while a post-lexical tensification rule is responsible for enhanced spatial

correlates. Such an account may be, in part, viable for TJ, as well. It would explain intra-gestural and kinematic modifications. However, it cannot readily explain differences in inter-gestural timing between singletons and geminates, in particular the longer durations of vowels preceding geminates. If TJ geminates were [+tense], we may expect *shortening* of the preceding vowel rather than lengthening, and we may also expect greater velocity or stiffness for the consonantal movements (Ridouane, 2003). Neither phenomenon is observed.

Gestural models, like Articulatory Phonology (Browman & Goldstein, 1990a) and Task Dynamics (Saltzman & Munhall, 1989), provide a more complete isomorphism between phonetics and phonology. This approach may offer a natural way to capture both temporal (intra- and inter-gestural) and spatial modifications. Since the framework operates with multiple simultaneous articulatory tiers and not with linearly arranged vowels and consonants, it could also capture inter-gestural effects of gemination.

As noted, in Articulatory Phonology, the duration of articulatory gestures is indirectly captured either by longer activation times (Saltzman & Munhall, 1989) or differences in articulatory parameters, for example, changes in stiffness and targets (Browman & Goldstein, 1990b). In other words, while there is no direct way to specify gemination contrasts with a single [+long] feature, gestural models have more flexibility. This flexibility can accommodate highly regular, language-specific phonetic details observed for gemination contrasts. In addition, gestural models can directly relate spatial and temporal modifications via prosodic modulation gestures (Byrd & Krivokapić, 2021; Katsika et al., 2014; Saltzman et al., 2008).

Viewed from this angle, TJ gemination may be a form of (micro-) prosodically-modulated articulation (Shaw, 2022), similar to the production of articulatory gestures in stressed syllables or near prosodic boundaries (Byrd & Saltzman, 1998, 2003; Katsika, 2016; Saltzman et al., 2008). Thus the approach would rely on the presence of prosodic modulation gestures, μ (mu) gestures, of which the better-known prosodic boundary modulation is a special case, the so-called π (pi) gesture (Burroni & Tilsen, 2022; Byrd & Krivokapić, 2021; Byrd & Saltzman, 1998, 2003). We also note that the idea that geminates are modulated or slowed down articulation is not new, but it has also been entertained in phonological work on Japanese (cf. Vance, 1987, especially the references in Chapter 5). We now introduce the elements that would be necessary to flesh out such an account.

In gestural frameworks, the initiation of articulatory gestures is controlled by a system of gestural planning oscillators (Browman & Goldstein, 2000; Goldstein et al., 2009; Nam, 2007b; Saltzman & Byrd, 2000). Furthermore, the activation of gestures, which controls their duration, is proportional to the frequency of the oscillator's periods (e.g., Burroni, 2023; Kirkham & Strycharczuk, 2024; Saltzman et al., 2008; Tilsen, 2018). Saltzman et al. (2008) proposed temporal

(μ_T) and spatial (μ_S) modulation gestures. μ_T gestures alter online the frequency of planning oscillators, causing gestures (and larger units like syllables and feet) to increase in duration. μ_S gestures change the spatial parameters of co-active articulatory gestures. Both classes of gestures have subsequently been adopted to model many suprasegmental phenomena, including rhythmic shortenings, the kinematics of stressed syllables, the scope of boundary effects, etc. (cf. Byrd & Krivokapić, 2021, for a review). μ_T alone may explain the temporal and spatial characteristics of geminate production in TJ and possibly explain global effects on inter-gestural timing, as we illustrate in the simulation below.

The model below closely follows previous approaches, especially Saltzman et al. (2008). Following standard work in Articulatory Phonology (Burroni, 2023; Byrd & Saltzman, 1998), the shape of activation is continuously ramped rather than step-like (Saltzman & Munhall, 1989). In this model, the unfolding of articulatory gestures is controlled by gestural planning oscillators which are modeled as second-order hybrid oscillatory systems, as in Equation 1 for the i^{th} oscillator:

$$(1) \quad \ddot{x}_i = -\alpha_i \dot{x}_i - \beta_i x_i^2 \dot{x}_i - \gamma_i \dot{x}_i^3 - \omega_{0i}^2 x_i$$

This model follows Saltzman et al. (2008) and other works in using a realistic duration for the gestures of interest (e.g., Kirkham & Strycharczuk, 2024); for singleton consonants we assume a frequency ω_0 of 10 Hz equal to a duration of 0.1s. The other parameters are defined following previous work (Saltzman et al., 2008): $-\alpha = \beta\,\omega_{0i}$ and $\gamma = \frac{1}{\omega_{0i}}$. Following the approach in Saltzman et al. (2008), a prosodic modulation gesture (μ_T) is defined as a gesture that warps the frequency of the i^{th} oscillator according to Equation2:

$$(2) \quad \omega_{0i}^* = \left(1 - \delta\alpha_{\mu_T}(t)\right)\omega_{0i}$$

The parameter δ, set to 0.5 following Saltzman et al., defines a "strength" coefficient for the time-dependent activation of the μ_T gesture α_{μ_T}. The higher the strength parameter δ, the stronger the warping of an oscillator's frequency.

Similar to the assumption made for stressed syllables (Saltzman et al., 2008), μ_T is assumed to be co-active with the initiation of the gestural planning oscillator cycle associated with a geminate consonant. Once μ_T is applied in this fashion (**Figure 10** top panel, dashed line) to a singleton-like gestural planning oscillator (**Figure 10** top panel, blue line), the resulting gestural planning oscillator period is increased (**Figure 10** top panel, blue double arrow). The increased period translates into longer activation intervals (**Figure 10** mid panel). If other parameters remain fixed (specifically, α, β and γ), as in previous models (Saltzman et al., 2008), the temporal modulation gesture μ_T also induces a *spatial* effect on the oscillator: it increases its amplitude during the slowed down period. We assume that just as oscillator frequency influences activation duration,

oscillator amplitude influences activation strength. For example, if the oscillator amplitude exceeds its normal value, so the activation strength of the associated gesture will proportionally exceed the upper activation limit of 1 (Saltzman & Munhall, 1989). If so, then μ_T directly affects both the duration and activation strength of its co-active consonantal gestures. Note that the first activation wave for singleton closure (in orange, mid panel of **Figure 10**) has lower activation and shorter duration than the one for geminate closure (in blue, **Figure 10**). The second activation wave of both singletons and geminates represents a release, as the simulations follow the split-gesture model of Articulatory Phonology (Browman, 1994; Burroni, 2022; Nam, 2007a, 2007b; Tilsen, 2017) in assuming that releases are actively controlled (the second activation waves in orange for singleton and in blue for geminates in **Figure 10** mid panel). Identical releases for both singleton and geminate gestures are posited, since no singleton/geminate differences were observed.

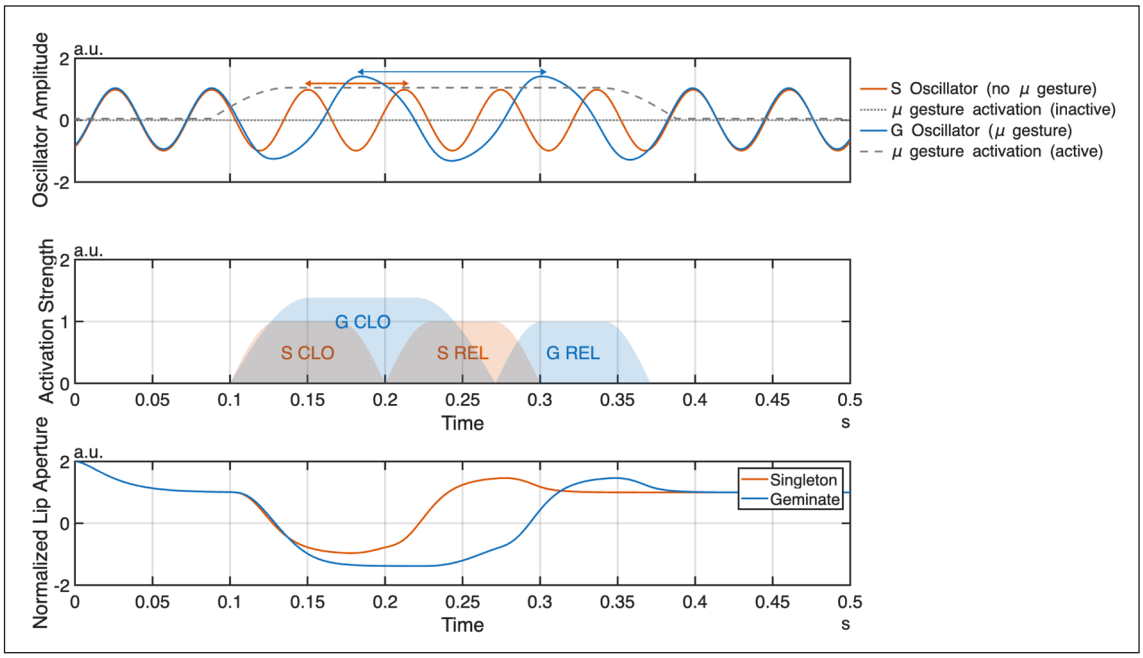

Figure 10: Top: Solid lines indicate gestural planning oscillators for singleton and geminate gestures, dashed lines mark the inactive or active μ_T gesture, double arrows mark duration of one period, which is proportional to activation duration, like amplitude is to activation strength. **Mid**: Activation for geminates and singletons; note the different duration and level of activation for the first series of waves marking closure activations. The second activation wave for both singletons and geminates is release activation. **Bottom**: Predicted Lip Aperture trajectories for singletons and geminates.

As a final step, the activation waves, differently modulated by μ_T, together with the *identical* gestural parameters, enter into the equation of the Task Dynamic model (Saltzman & Munhall, 1989), which regulates gestural execution via a second-order critically damped forced harmonic oscillator in Equation 3:

$$\ddot{z}_i = m^{-1} \cdot \left(-b\dot{z} - k\left(t\right)\left(z - z_0\left(t\right)\right)\right) \quad (3)$$

Again, the model follows standard assumptions of unit mass $m = 1$, critical damping $b = 2\sqrt{km}$, and time-dependent stiffness *k(t)* and target $z_0(t)$ that depend on the currently active gestures. We assume, following Byrd and Saltzman (1998), that damping, stiffness, and target parameters all scale with activation strength. If multiple gestures are active, the time-dependent target and stiffness are derived via activation-weighted averaging of the parameters of all active gestures (Saltzman & Munhall, 1989).

When the activation patterns are entered into the model using *identical* parameter specifications, i.e., the same stiffness, target, the model outputs the kinematic trajectories in the bottom panel of **Figure 10**. The one that illustrates the geminate displays a longer closing phase and, especially, plateau, in addition to slightly more extreme spatial correlates. The difference emerges despite geminates being *underlying in all respects identical to singletons*, except for prosodic modulation.

The μ_T gesture thus provides a principled mechanism to account for both the intra-gestural and kinematic differences observed in TJ geminate production. By locally modulating gestural planning oscillator frequency, the μ_T gesture slows down the activation dynamics of the consonantal gestures, producing a longer plateau and closing phase interval (**Figure 10**, bottom row). At the same time, because the same modulation also affects the strength of activation dynamics, spatial correlates emerge as a by-product. As activation levels are also stronger during slowed-down cycles, more extreme articulatory targets are approached with greater displacement and lower kinematic stiffness, consistent with the characteristics of geminates observed in our data. In addition, since the vocalic gestures are co-active with μ_T, we can consider whether they are also modulated. There are two analytical possibilities. One is that the temporal modulation gesture we have proposed only modulates the consonantal gesture. The alternative is that the temporal modulation gesture is like the π gesture, proposed for boundary adjacent-prosodic lengthening. The π gesture applies the same temporal slowdown to all overlapped gestures. If μ_T also applies to all co-active gestures, predicted effects are lower average speed (for vowels just like for geminates) and spatial effects on vowel articulation, which have been reported in the literature, at least based on formant values (Morimoto et al., 2025).

To assess whether μ_T impacts overlapped gestures, we studied the average speed of the tongue dorsum in the vertical dimension during the consonantal gesture interval (gestural onset

to offset). The peak velocity occurs during the consonantal closing phase (**Figure 8**), when the μ_T is expected to be active. A linear mixed-effects regression model on this measure revealed that the vocalic articulation, accomplished by the tongue dorsum, is indeed slower during geminate production. We found an intercept average speed of 1.4 cm/s (SE 0.9 cm/s) for alveolar, palatal, and dorsal singletons. Tongue dorsum average speed is faster during bilabials, with an effect size of 2.2 cm/s (SE 1.06 cm/s). Crucially, for geminates, the tongue dorsum average speed is reduced with an effect of –0.55 cm/s (SE 0.21 cm/s). No interaction of gemination with POA was found. We refer the reader to our additional files for more details on this model. Given the wide variety of vocalic context observed in our data, it is not possible to check spatial effects on vowels. This is a task that we leave open for future work based on more controlled vocalic contexts. The slower vocalic movement that occurs during the geminate suggests that the type of temporal modulation gesture needed is trans-gestural modulation, the same type proposed to account for articulatory slowdown at prosodic boundaries (Byrd & Saltzman, 2003) and for rhythmic effects (Saltzman et al., 2008).

In sum, the μ_T gesture framework offers a way to explain why spatial strengthening follows from temporal modulation rather than constituting an independent or even primary control parameter. In addition, this account brings another suprasegmental phenomenon, phonological length, within the purview of prosodic modulation gestures, thus increasing the number of empirical domains that may be subsumed under gestural accounts of suprasegments and prosody (Byrd & Krivokapić, 2021). Finally, we also note that gestural models provide a particularly flexible and explicit account of the interaction between durational and non-durational parameters. This is an aspect on which the broader literature on articulatory strength (e.g., Al-Tamimi & Khattab, 2015, 2018; Jessen, 2001; Kochetov & Kang, 2017; Kohler, 1984) has independently converged, providing important empirical and theoretical groundwork that gestural accounts can now build upon.

Widening the scope to geminate production cross-linguistically, a key question is whether the μ_T account proposed here extends to other languages, like Italian. Evidence from companion studies on Italian (Burroni et al., 2024, under review) suggests that, in this language, geminates may involve not only intra-gestural and kinematic modifications—the latter arguably stronger than in TJ—but also qualitatively different inter-gestural timing signatures. Specifically, Italian geminates seem to be characterized by earlier consonantal initiation relative to the preceding vowel and earlier V_2 initiation, both of which differ from TJ, where the V_1-CLO lag shows no clear gemination effect and V_1-V_2 lags are lengthened rather than shortened. If confirmed, these cross-linguistic differences suggest that geminate production is better viewed as a complex, layered articulatory task: languages like Italian and Japanese appear to share a universal core, longer plateau durations compatible with a μ_T modulation, while also differing in how additional articulatory dimensions, including kinematic strengthening and inter-gestural reorganization,

are further recruited to accomplish the length contrast "task." A rigorous account of how these differences may reflect language-specific interplays between μ_T and other mechanisms, like inter-gestural coupling differences, is an issue that requires the collection and rigorous modeling of a cross-linguistic dataset currently not available. In this way, we may also be able to clarify whether the differences in geminate production may indeed be connected to rhythmic properties, as previously hypothesized (Smith, 1992, 1995).

In conclusion, the μ_T gesture approach offers a potential way to unify the apparent dual nature of gemination in TJ—its temporal and spatial signatures—under a single dynamic mechanism. The model also captures why temporal cues dominate in the classification results (**Figure 8**): the primary control resides in the temporal domain, while kinematic and spatial effects are emergent. This provides an explanation of our findings that does not require multiple independent phonological mechanisms, such as multiple [+ long], [+ tense], or different moraic/autosegmental specifications with language-specific phonetic realizations. Instead, gemination arises as a prosodic modulation of articulatory timing, an instance of local slowing of gestural planning oscillators, akin to prosodic boundary lengthening or stress-related temporal expansion. This idea has also been considered in Japanese phonology work (Vance, 1987), albeit not in a gestural framework. Crucially, the model can also capture cross-articulator effects, an aspect that is harder to account for in more linear and non-gestural phonological frameworks and has been consistently reported in previous work (Löfqvist, 2006; Morimoto et al., 2025). Although this account is more complex than positing a single [+ long] feature (or a combination of features), we believe that it has important advantages. It brings closer phonological representations and highly-regular, language-specific phonetic details. Future work should also address the issue of whether μ_T accounts could be tied to general motor control accounts, where modulations in activation could arise from feedback mechanisms (Burroni & Tilsen, 2022; Tilsen, 2022) and/or selection dynamics in a planning field (Chaturvedi & Shaw, 2025; Kirkham & Strycharczuk, 2024; Shaw, 2025; Tilsen, 2018).

4.4. Limitations

Some limitations of the present study should be acknowledged. First, the data were collected in a citation-style speech context and used mimetic words. Future work comparing TJ geminate production, both lexical and mimetic words, across different speaking rates and prosodic contexts, would clarify whether the proposed temporal modulation mechanisms generalize across all contexts.

Second, our temporal analyses were based primarily on single articulatory sensors for each POA and on acoustic landmarks. While these methods provide reliable temporal measures, they cannot capture the full dynamics of the vocal tract and are especially limited if we wish to observe global constriction patterns and changes in tongue shapes typical of vowels. Techniques

that combine high spatial and temporal resolutions, such as real-time MRI or ultrasound, would enable a more detailed examination of such properties.

Third, our dataset included six speakers and a limited number of manners of articulation (oral and nasal stops). Expanding the dataset to include more speakers and a fuller set of manners (especially fricatives and affricates) across places of articulation would allow for stronger generalization of the observed temporal and kinematic patterns. In particular, it would be important to examine how μ_T modulation may interact with manner of articulation and aerodynamic requirements. Affricates are especially informative in this regard, since, cross-linguistically, phonological gemination seems to target their closure rather than their frication (Pycha, 2009), consistent with our finding that release phases show no modulation. In TJ, the realization of /z/ as fricative or affricate has itself been shown to depend on consonant duration (Maekawa, 2010), suggesting complex interactions that deserve further investigation.

Finally, while our modeling provided a qualitative account of how temporal modulation gestures can generate geminate-like patterns, a more quantitative computational modeling approach is needed. Incorporating multiple articulators and parameterizing the strength and timing of modulation gestures would allow for explicit testing of model predictions against empirical data, offering a more comprehensive account of intra-gestural, kinematic, and spatial properties of TJ geminates.

5. Conclusion

This study investigated how speakers of Tokyo Japanese (TJ) produce and control consonantal length contrasts, integrating articulatory evidence across bilabial, alveolar, palatal, and dorsal mimetic geminates. Replicating and extending our prior work, we found that the primary intra-gestural correlate of gemination is a robust lengthening of the gestural plateau, accompanied by a smaller increase in closing phase duration and no change in release duration. Kinematically, geminates showed more extreme articulator positions, larger movement amplitudes, lower stiffness, and slower peak velocity, while inter-gestural timing revealed longer preceding vowels, longer V_1-TARG Lags, and longer V_1-V_2 Lags, with no substantial differences in V_1-CLO Lags. Classification analyses further demonstrated that temporal cues (intra- and inter-gestural) strongly outperform purely spatial/kinematic cues in distinguishing singletons from geminates.

We proposed a unified gestural account in which a temporal prosodic modulation gesture (μ_T) slows the relevant planning oscillator(s), directly yielding longer plateaux and indirectly producing the observed spatial/kinematic correlates via activation dynamics. This approach obviates the need for multiple independent phonological specifications and explains why temporal properties dominate in the discriminability of geminates. Within this framework, spatial strengthening is emergent rather than primary.

Situating TJ in a cross-linguistic perspective, we argued that TJ constitutes a case where local temporal control drives the length contrast and reorganizes surrounding vocalic timing, contrasting with languages such as Italian, where durational control often reflects different timing regimes (e.g., greater overlap/anticipation and concomitant pre-geminate shortening). This typological contrast may be consistent with differences in higher-level temporal organization (e.g., syllable/mora timing), but more research is needed in this regard.

Overall, kinematic studies of geminate production reveal a more general lesson: similar acoustic contrasts can arise from distinct underlying articulatory control strategies, and thus articulatory investigations are necessary to understand how speakers of the same or different languages implement phonological contrasts in the vocal tract.

Additional files

The following additional files for this article can be found as follows: https://osf.io/4tk8q/overview.

Acknowledgements

We thank Teerawee Sukanchanon and Piyapath Srisomyos for their help with data annotation, as well as Ulrike Rupprecht for help with data acquisition. We also thank Phil Hoole for help in setting up the GUI used to collect the data and Michele Gubian for helpful discussion concerning statistical analyses.

Ethics approval

The authors obtained ethical approval from the institutional review board of the University of Munich. Informed consent was obtained from all participants.

Competing interests

The authors have no competing interests to declare.

Author contributions

Francesco Burroni: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing – Original Draft (Sections 1.1, 1.2, 1.3, 1.4, 1.5, 2.1, 2.3, 2.4, 3, 4.1, 4.2, 4.3, 5), Writing – Review & Editing, Visualization, Project administration. **Lia Saki Bučar Shigemori**: Conceptualization, Methodology, Validation, Investigation, Resources, Data Curation, Writing – Original Draft (Sections 1.3, 2.2, 4.4, 5), Writing – Review & Editing, Project administration. **Shigeto Kawahara**: Conceptualization, Methodology, Writing – Review & Editing, Supervision, Funding acquisition. **Jason A. Shaw**: Conceptualization, Methodology, Writing – Review & Editing, Supervision.

References

Abramson, A. S. (1998). The complex acoustic output of a single articulatory gesture: Pattani Malay word-initial consonant length. In U. Warotamasikkhadit and T. Panakul (Eds.), *Papers from the Fourth Annual Meeting of the Southeast Asian Linguistics Society: Vol. 4* (pp. 1–20). https://sealang.net/sala/seals/htm/4/volume.htm

Al-Tamimi, J., & Khattab, G. (2015). Acoustic cue weighting in the singleton vs geminate contrast in Lebanese Arabic: The case of fricative consonants. *Journal of the Acoustical Society of America*, *138*(1), 344–360. https://doi.org/10.1121/1.4922514

Al-Tamimi, J., & Khattab, G. (2018). Acoustic correlates of the voicing contrast in Lebanese Arabic singleton and geminate stops. *Journal of Phonetics*, *71*, 306–325. https://doi.org/10.1016/j.wocn.2018.09.01030

Atal, B. S., Chang, J. J., Mathews, M. V., & Tukey, J. W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *Journal of the Acoustical Society of America*, *63*(5), 1535–1555. https://doi.org/10.1121/1.381848

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Boersma, P., & Weenink, D. (2025). PRAAT: Doing phonetics by computer [Computer program]. https://www.praat.org/

Browman, C. P. (1994). Lip aperture and consonant releases. In P. A. Keating (Ed.), *Phonological structure and phonetic form: Papers in laboratory phonology III* (pp. 331–353). Cambridge University Press. https://doi.org/10.1017/CBO9780511659461.019

Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*(2–4), 140–155.

Browman, C. P., & Goldstein, L. (1990a). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, *18*(3), 299–320. https://doi.org/10.1016/S0095-4470(19)30376-6

Browman, C. P., & Goldstein, L. (1990b). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology* (pp. 341–376). Cambridge University Press. https://doi.org/10.1017/CBO9780511627736.019

Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP. Bulletin de la communication parlée*, *5*, 25–34.

Burroni, F. (2022). A split-gesture, competitive, coupled oscillator model of syllable structure predicts the emergence of edge gemination and degemination. In A. Ettinger, T. Hunter, & B. Prickett (Eds.), *Proceedings of the Society for Computation in Linguistics 2022* (pp. 11–22). https://aclanthology.org/2022.scil-1.2/

Burroni, F. (2023). *Dynamics of f0 planning and production: Contextual and rate effects on thai tone gestures* [Doctoral dissertation, Cornell University]. https://doi.org/10.7298/rj8r-kg22

Burroni, F., Kawahara, S., & Shaw, J. A. (2025). Articulatory correlates of consonantal length contrasts: The case of Japanese mimetic geminates. *JASA Express Letters*, *5*(1), 015201. https://doi.org/10.1121/10.0034762

Burroni, F., Lau-Preechathammarach, R., & Maspong, S. (2022). Unifying initial geminates and fortis consonants via laryngeal specification: Three case studies from Dunan, Pattani Malay, and Salentino. *Proceedings of the 2021 Annual Meeting on Phonology*. https://doi.org/10.3765/amp.v9i0.5200

Burroni, F., Maspong, S., Benker, N., Hoole, P. A., & Kirby, J. (2024). Spatiotemporal features of bilabial geminate and singleton consonants in Italian. *Proceedings of ISSP*, 181–184. https://doi.org/10.21437/issp.2024-46

Burroni, F., Maspong, S., Hoole, P., & Kirby, J. (under review). Phonological control of time in speech: Italian consonantal length contrasts as an articulatory task.

Burroni, F., Maspong, S., Pittayaporn, P., & Kochaiyaphum, P. (2020). A new look at Pattani Malay initial geminates: A statistical and machine learning approach. *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, 21–29. https://aclanthology.org/2020.paclic-1.3/

Burroni, F., & Tilsen, S. (2022). The online effect of clash is durational lengthening, not prominence shift: Evidence from Italian. *Journal of Phonetics*, *91*, 101124. https://doi.org/10.1016/j.wocn.2021.101124

Burroni, F., & Tilsen, S. (2025). Thai speakers time lexical tones to supralaryngeal articulatory events. *Journal of Phonetics*, *108*, 101389. https://doi.org/10.1016/j.wocn.2024.101389

Byrd, D., & Krivokapić, J. (2021). Cracking prosody in Articulatory Phonology. *Annual Review of Linguistics*, *7*(1), 31–53 https://doi.org/10.1146/annurev-linguistics-030920-050033

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, *26*(2), 173–199. https://doi.org/10.1006/jpho.1998.0071

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2), 149–180. https://doi.org/10.1016/S0095-4470(02)00085-2

Celata, C., Meluzzi, C., & Bertini, C. (2022). Acoustic and kinematic correlates of heterosyllabicity in different phonological contexts. *Language and Speech*, *65*(3), 755–780. https://doi.org/10.1177/00238309211065789

Chaturvedi, M., & Shaw, J. A. (2025). A dynamic neural model of tonal downstep. *Proceedings of the Annual Meetings on Phonology*, *1*(1). https://doi.org/10.7275/amphonology.3021

DiCanio, C. T. (2012). The phonetics of fortis and lenis consonants in Itunyoso Trique. *International Journal of American Linguistics*, *78*(2), 239–272. https://doi.org/10.1086/664481

Dunn, M. H. (1993). *The phonetics and phonology of geminate consonants: A production study* [Doctoral dissertation, Yale University].

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, *8*(1), 113–133. https://doi.org/10.1016/S0095-4470(19)31446-9

Fuchs, S., Perrier, P., & Hartinger, M. (2011). A critical evaluation of gestural stiffness estimations in speech production based on a linear second-order model. *Journal of Speech, Language, and Hearing Research*, *54*(4), 1067–1076. https://doi.org/10.1044/1092-4388(2010/10-0131)

Fujimoto, M., Funatsu, S., & Hoole, P. (2015). Articulation of single and geminate consonants and its relation to the duration of the preceding vowel in Japanese. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, S.-S. Stuart-Smith, & J. M. Scobbie (Eds.), *18th International Congress of Phonetic Sciences, ICPhS 2015, Glasgow, UK, August 10–14, 2015*. University of Glasgow.

Gafos, A. & Goldstein, L. (2011). Organization of phonological elements: Articulatory representation. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford handbook of laboratory phonology* (pp. 219–253). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199575039.013.0010

Gahl, S., & Baayen, R. H. (2024). Time and thyme again: Connecting English spoken word duration to models of the mental lexicon. *Language*, *100*(4), 623–670. https://doi.org/10.1353/lan.2024.a947037

Garcia, D. (2010). Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics & Data Analysis*, *54*(4), 1167–1178. https://doi.org/10.1016/j.csda.2009.09.020

Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In C. G. M. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in phonetics and speech science* (pp. 239–249). The Commercial Press.

Greca, P., Burroni, F., & Harrington, J. (2025). Does VtoV coarticulation always require "look-ahead"? Evidence from an EMA and acoustic study of Campanian Italian. *6th Phonetics and Phonology in Europe. PaPE 2025 Book of Abstracts*.

Guenther, F. H. (2016). *Neural control of speech*. MIT Press. https://doi.org/10.7551/mitpress/10471.001.0001

Guenther, F. H., & Hickok, G. (2015). Role of the auditory system in speech production. In M. J. Aminoff, F. Boller, & D. F. Swaab (Eds.), *Handbook of clinical neurology* (pp. 161–175, Vol. 129). Elsevier. https://doi.org/10.1016/B978-0-444-62630-1.00009-3

Hagedorn, C., Proctor, M., & Goldstein, L. (2011). Automatic analysis of singleton and geminate consonant articulation using real-time magnetic resonance imaging. *Interspeech 2011*. International Speech Communication Association. https://doi.org/10.21437/interspeech.2011-162

Ham, W. (2002). *Phonetic and phonological aspects of geminate timing*. Routledge. https://doi.org/10.4324/9781315023755

Hayes, B., & Steriade, D. (2004). Introduction: The phonetic bases of phonological markedness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 1–33). Cambridge University Press.

Hirata, Y., & Amano, S. (2012). Production of single and geminate stops in Japanese three-and four-mora words. *The Journal of the Acoustical Society of America*, *132*(3), 1614–1625. https://doi.org/10.1121/1.4730975

Idemaru, K., & Guion, S. G. (2008). Acoustic covariants of length contrast in Japanese stops. *Journal of the International Phonetic Association*, *38*(2), 167–186. https://doi.org/10.1017/S0025100308003459

Ishii, T. (1999). A study of the movement of the articulatory organs in Japanese geminate production-ANX-ray microbeam analysis. *Nippon Jibiinkoka Gakkai Kaiho*, *102*(5), 622–634. https://doi.org/10.3950/jibiinkoka.102.622

Jessen, M. (2001). Phonetic implementation of the distinctive auditory features [voice] and [tense] in stop consonants. In T. A. Hall (Ed.), *Distinctive feature theory* (pp. 237–294). De Gruyter Mouton. https://doi.org/10.1515/9783110886672.237

Katsika, A. (2016). The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics*, *55*, 149–181. https://doi.org/10.1016/j.wocn.2015.12.003

Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., & Goldstein, L. (2014). The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics*, *44*, 62–82. https://doi.org/10.1016/j.wocn.2014.03.003

Kawagoe, I. (2015). The phonology of sokuon, or geminate obstruents. In H. Kubozono (Ed.), *Handbook of Japanese phonetics and phonology* (pp. 79–120). De Gruyter Mouton. https://doi.org/10.1515/9781614511984.79

Kawahara, S. (2013). Emphatic gemination in Japanese mimetic words: A wug-test with auditory stimuli. *Language sciences*, *40*, 24–35. https://doi.org/10.1016/j.langsci.2013.02.002

Kawahara, S. (2015). The phonetics of sokuon, or geminate obstruents. In H. Kubozono (Ed.), *Handbook of Japanese phonetics and phonology* (pp. 43–78). De Gruyter Mouton. https://doi.org/10.1515/9781614511984.43

Kawahara, S., & Matsui, F. M. (2017). Some aspects of Japanese consonant articulation: A preliminary EPG study. *ICU Working Papers in Linguistics (ICUWPL)*, *2*, 9–20.

Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, *14*(1), 29–59. https://doi.org/10.1016/S0095-4470(19)30608-4

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

Kirkham, S., & Strycharczuk, P. (2024). A dynamic neural field model of vowel diphthongisation. In C. Fougeron & P. Perrier (Eds.), *Proceedings of ISSP 2024 – 13th International Seminar on Speech Production* (pp. 205–208). https://doi.org/10.21437/issp.2024-52

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, *59*(5), 1208–1221. https://doi.org/10.1121/1.380986

Kochetov, A. (2025). Linguopalatal contact differences between Japanese geminates and singletons across different places and manners. *Journal of the International Phonetic Association*, 1–30. https://doi.org/10.1017/S0025100325100662

Kochetov, A., & Kang, Y. (2017). Supralaryngeal implementation of length and laryngeal contrasts in Japanese and Korean. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, *62*(1), 18–55. https://doi.org/10.1353/cjl.2017.0002

Kohler, K. J. (1984). Phonetic explanation in phonology: The feature fortis/lenis. *Phonetica*, *41*(3), 150–74. https://doi.org/10.1159/000261721

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. https://doi.org/10.1111/lnc3.12373

Kubozono, H. (2017). *The phonetics and phonology of geminate consonants*. Oxford University Press. https://doi.org/10.1093/oso/9780198754930.001.0001

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell Oxford.

Lahiri, A., & Hankamer, J. (1988). The timing of geminate consonants. *Journal of Phonetics*, *16*(3), 327–338. https://doi.org/10.1016/S0095-4470(19)30506-6

Local, J., & Simpson, A. (1999). Phonetic implementation of geminates in Malayalam nouns. *Proceedings of the 14th ICPhS*, 595–598.

Löfqvist, A. (2005). Lip kinematics in long and short stop and fricative consonants. *Journal of the Acoustical Society of America*, *117*(2), 858–878. https://doi.org/10.1121/1.1840531

Löfqvist, A. (2006). Interarticulator programming: Effects of closure duration on lip and tongue coordination in Japanese. *Journal of the Acoustical Society of America*, *120*(5), 2872–2883. https://doi.org/10.1121/1.2345832

Löfqvist, A. (2007). Tongue movement kinematics in long and short Japanese consonants. *Journal of the Acoustical Society of America*, *122*(1), 512–518. https://doi.org/10.1121/1.2735102

Löfqvist, A. (2017). Articulatory coordination in long and short consonants: An effect of rhythm class? In H. Kubozono (Ed.), *The phonetics and phonology of geminate consonants* (pp. 118–129). Oxford University Press. https://doi.org/10.1093/oso/9780198754930.003.0006

Maddieson, I. (1984). Phonetic cues to syllabification. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged*. Academic Press.

Maekawa, K. (2010). Coarticulatory reinterpretation of allophonic variation: Corpus-based analysis of /z/ in spontaneous Japanese. *Journal of Phonetics*, *38*(3), 360–374. https://doi.org/10.1016/j.wocn.2010.03.001

Morimoto, M. (2020). *Geminated liquids in Japanese: A production study* [Doctoral dissertation, University of California, Santa Cruz].

Morimoto, M., & Kitamura, T. (2019). Articulation of geminated liquids in Japanese. *Proceedings of the 19th International Congress of Phonetic Sciences*, 2811–2815.

Morimoto, M., Mizoguchi, A., Li, W., & Arai, T. (2025). Tongue contours during pre-geminate vowel lengthening in Japanese: A case study with voiceless alveolar plosive /t/. *Journal of the Phonetic Society of Japan*, *29*(1), 103–113. https://doi.org/10.24467/onseikenkyu.29.1_103

Nam, H. (2007a). Articulatory modeling of consonant release gesture. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 625–628.

Nam, H. (2007b). *A gestural coupling model of syllable structure* [Doctoral dissertation, Yale University].

Neiberg, D., Ananthakrishnan, G., & Engwall, O. (2008). The acoustic to articulation mapping: Non-linear or non-unique? *Interspeech*, *2008*, 1485–1488. International Speech Communication Association. https://doi.org/10.21437/interspeech.2008-427

Pycha, A. (2009). Lengthened affricates as a test case for the phonetics-phonology interface. *Journal of the International Phonetic Association*, *39*(1), 1–31.

Recasens, D. (1990). The articulatory characteristics of palatal consonants. *Journal of Phonetics*, *18*(2), 267–280. https://doi.org/10.1016/S0095-4470(19)30393-6

Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of the Acoustical Society of America*, *125*(4), 2288–2298. https://doi.org/10.1121/1.3089222

Ridouane, R. (2003). *Suites de consonnes en Berbère: Phonétique et phonologie [Consonant clusters in Berber: Phonetics and phonology]* [Doctoral dissertation, Université de la Sorbonne Nouvelle, Paris III]. https://theses.hal.science/tel-00143619

Ridouane, R. (2007). Gemination in Tashlhiyt Berber: An acoustic and articulatory study. *Journal of the International Phonetic Association*, *37*(2), 119–142. https://doi.org/10.1017/S0025100307002903

Ridouane, R. (2010). Geminates at the junction of phonetics and phonology. In F. Cécile, K. Barbara, D. Mariapaola, & V. Nathalie (Eds.), *Laboratory Phonology 10* (pp. 61–90). De Gruyter Mouton. https://doi.org/10.1515/9783110224917.1.61

Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, *19*(4), 499–526. https://doi.org/10.1016/S0167-9457(00)00030-0

Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*(4), 333–382. https://doi.org/10.1207/s15326969eco0104_2

Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. *Speech Prosody*, *2008*, 175–184. https://doi.org/10.21437/SpeechProsody.2008-3

Shaw, J. A. (2022). Micro-prosody. *Language and Linguistics Compass*, *16*(2), e12449. https://doi.org/10.1111/lnc3.12449

Shaw, J. A. (2025). Unifying phonological and phonetic aspects of speech in dynamic neural fields: The case of laryngeal patterns in Japanese. *Phonological Studies*, *28*, 57–68.

Shaw, J. A., & Kawahara, S. (2018). The lingual articulation of devoiced /u/ in Tokyo Japanese. *Journal of Phonetics*, *66*, 100–119. https://doi.org/10.1016/j.wocn.2017.09.007

Shaw, J. A., & Kawahara, S. (2019). Effects of surprisal and entropy on vowel duration in Japanese. *Language and Speech*, *62*(1), 80–114. https://doi.org/10.1177/0023830917737331

Shibatani, M. (1990). *The languages of Japan*. Cambridge University Press.

Smith, C. L. (1992). *The timing of vowel and consonant gestures* [Doctoral dissertation, Yale University].

Smith, C. L. (1995). Prosodic patterns in the coordination of vowel and consonant gestures. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence: Papers in laboratory phonology IV* (pp. 205–222). Cambridge University Press. https://doi.org/10.1017/CBO9780511554315.015

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*(1), 3–45. https://doi.org/10.1016/S0095-4470(19)31520-7

Takada, M. (1985). On some articulatory characteristics of the mora obstruent (sokuon). *National Institute for Japanese Language and Linguistics. Kenkyu Kokokusyo*, *6*, 17–40. https://doi.org/10.15084/00001095

Takeyasu, H., & Giriko, M. (2017). Effects of duration and phonological length of the preceding/following segments on perception of the length contrast in Japanese. In H. Kubozono (Ed.), *The phonetics and phonology of geminate consonants* (pp. 85–117). Oxford University Press. https://doi.org/10.1093/oso/9780198754930.003.0005

Tilsen, S. (2017). Exertive modulation of speech and articulatory phasing. *Journal of Phonetics*, *64*, 34–50. https://doi.org/10.1016/j.wocn.2017.03.001

Tilsen, S. (2018). Three mechanisms for modeling articulation: Selection, coordination, and intention. *Cornell working papers in phonetics and phonology*. Cornell Phonetics Lab.

Tilsen, S. (2022). An informal logic of feedback-based temporal control. *Frontiers in Human Neuroscience*, *16*, 851991. https://doi.org/10.3389/fnhum.2022.851991

Tilsen, S., & Hermes, A. (2020). Nonlinear effects of speech rate on articulatory timing in singletons and geminates. *12th International Seminar on Speech Production*, 56–59. https://hal.science/hal-03510889

Tomaschek, F., Tucker, B. V., Fasiolo, M., & Baayen, R. H. (2018). Practice makes perfect: The consequences of lexical proficiency for articulation. *Linguistics Vanguard*, *4*(s2), 20170018. https://doi.org/10.1515/lingvan-2017-0018

Turk, A., & Shattuck-Hufnagel, S. (2020a). *Speech timing: Implications for theories of phonology, speech production, and speech motor control*. Oxford University Press. https://doi.org/10.1093/oso/9780198795421.001.0001

Turk, A., & Shattuck-Hufnagel, S. (2020b). Timing evidence for symbolic phonological representations and phonology-extrinsic timing in speech production. *Frontiers in Psychology*, *10*, 2952. https://doi.org/10.3389/fpsyg.2019.02952

Türk, H., Lippus, P., & Šimko, J. (2017). Context-dependent articulation of consonant gemination in Estonian. *Laboratory Phonology*, *8*(1), 26. https://doi.org/10.5334/labphon.117

Vance, T. J. (1987). *An introduction to Japanese phonology*. State University of New York Press.

Zmarich, C., Fivela, B. G., Perrier, P., Savariaux, C., & Tisato, G. (2011). Speech timing organization for the phonological length contrast in Italian consonants. *International Speech Communication Association*, *2011*, 401–404. https://doi.org/10.21437/Interspeech.2011-160