ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Temporal Coordination of Articulatory Gestures in Nonnative Onset Clusters: Evidence From American English Speakers Using Electromagnetic Articulography

Hung-Shao Cheng,[a] Matthew Masapollo,[b] Christina Hagedorn,[c] and Adam Buchwald[a]

[a] Department of Communicative Sciences and Disorders, New York University, New York [b] Department of Speech, Language, and Hearing Sciences, University of Florida, Gainesville [c] Linguistics Program, College of Staten Island, City University of New York

ARTICLE INFO

Article History:
Received January 12, 2024
Revision received April 17, 2024
Accepted June 27, 2024

Editor-in-Chief: Cara E. Stepp
Editor: Defne Abur

https://doi.org/10.1044/2024_JSLHR-24-00025

ABSTRACT

Purpose: Research in cross-language speech production indicates that, although the production of nonnative consonant clusters is often difficult, speakers of American English can produce some nonnative clusters (e.g., /fn/) with high accuracy. This ease of production for select nonnative clusters may occur due to similarity of phonetic structure with native clusters (e.g., nonnative /fn/ and native /sm/ are both fricative-nasal sequences). The current study tested this hypothesis by examining the extent of transfer of articulatory coordination from phonetically similar native onset clusters (i.e., /fl/, /sm/) to nonnative /fn/ clusters.

Method: Using electromagnetic articulography, lip, tongue, and jaw movements were recorded in nine participants during the production of 22 nonwords (eight tokens per nonword) containing the native and nonnative clusters in different carrier phrases. We examined the temporal lags between each consonantal gesture in a cluster and the flanking vowel gesture, which were compared to the matched singleton conditions.

Results: Analyses revealed that, as in native speech, when the syllable onset became more complex (i.e., CV → CCV [C as consonant, V as vowel]), there was an increase in lag (less temporal overlap) between the leftmost consonantal gesture and the vocalic gesture, whereas there was a decrease in lag (more temporal overlap) between the rightmost consonant and the vocalic gesture (i.e., *C-center timing*). However, the segmental makeup of the cluster and type of carrier phrase used were also found to influence this change in temporal organization, raising new questions for future research.

Conclusions: By and large, the findings are in agreement with the idea that the temporal coordination of articulator movements may be transferred from native clusters to phonetically similar nonnative clusters. However, kinematic measures of a broader range of nonnative clusters in different contexts are needed to fully explore this position.

The widespread existence of accented speech among second language (L2) learners reflects the difficulty of producing sound sequences that are not found in one's native language (henceforth nonnative). Many nonnative consonant clusters are particularly challenging to produce in a fluent, coordinated, and natural-sounding way. Yet, at the same time, research has shown that speakers do not produce all nonnative clusters with equal accuracy (e.g.,

Correspondence to Matthew Masapollo: matthew.masapollo@mcgill.ca. Matthew Masapollo is now at the Department of Psychology, McGill University, Montreal, Quebec, Canada. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Davidson, 2006). Considerable research has sought to identify the factors that account for these accuracy differences. It has been suggested that the ease of production associated with select nonnative clusters occurs due to the similarity of phonetic structure with other native clusters (Davidson, 2006, 2010). If nonnative cluster production does indeed involve the application of a generalized representation of vocal tract dynamics, then articulatory coordination patterns for native clusters should transfer to nonnative clusters composed of phonetically similar movements. However, critical data needed to test this hypothesis are currently lacking since nonnative cluster production has always been studied acoustically. Detailed, kinematic analyses of nonnative cluster production, as in native cluster production, are scarce.

In this article, we examine the temporal coordination of articulatory movements in American English speakers during the production of nonnative /fn/ clusters and two phonetically similar native onset clusters (i.e., /fl/, /sm/) to improve our understanding of cross-language speech production. A finding that articulatory coordination patterns for native clusters transfer to nonnative clusters would help explain why speakers produce some nonnative clusters with better accuracy than others. Such findings could also be used to inform articulatory strategies utilized in L2 pedagogy. Our primary motivation for examining the temporal organization of /fn/ clusters stems from findings in the speech production literature that American English speakers can produce /fn/ clusters with high accuracy levels (relative to other nonnative stop–stop clusters, such as /gd/ or /bd/) when measured acoustically (Buchwald & Cheng, 2023; Davidson, 2006, 2010; Shaw & Davidson, 2011; Wilson et al., 2014), but that the component gestures affiliated with each consonant are produced with less overlapping coordination (see Buchwald et al., 2019; Buchwald & Cheng, 2023; Davidson, 2010). We here compare the temporal organization of /fn/ to those of /fl/ and /sm/, using electromagnetic articulography (EMA; Perkell et al., 1992; Rebernik et al., 2021), to determine the extent to which speakers transfer their articulatory knowledge associated with native onset cluster timing to the production of phonetically similar nonnative clusters. In the remainder of this introduction, we first review the literature relevant to nonnative cluster production that motivated the current experiment. We then discuss previous work on temporal coordination patterns observed with word-initial consonant clusters. Lastly, we provide an overview of the current study and the key research questions that were addressed. We adopt terminologies from multiple conceptual frameworks in the literature to facilitate the discussion, but the goal of the present study was not designed to evaluate any specific framework.

Production of Nonnative Onset Consonant Clusters

Numerous experimental studies in cross-language speech production have consistently shown that the initial production accuracy of nonnative consonant clusters varies widely depending on the segmental makeup of the cluster, even when speakers are not actively learning the language from which they originate (Davidson, 2006, 2010). For example, acoustic investigations have shown that, during nonword production, speakers of American English execute nonnative /f/−nasal clusters (e.g., /fm/, /fn/) more accurately than other types of nonnative clusters (e.g., stop–stop clusters; Buchwald & Cheng, 2023; Davidson, 2006, 2010; Shaw & Davidson, 2011; Wilson et al., 2014). While /f/−nasal clusters are not found word-initially in American English, these sequences can occur across syllable boundaries (e.g., [f#n] in Daphna). Thus, speakers of American English may be able to transfer their knowledge of producing the relative timing of heterosyllabic /f#n/ sequences to word-initial onset position (/#fn/; see, e.g., Ussishkin & Wedel, 2003). However, in another study, Davidson (2006) reported that there was no relationship between American English speakers' production accuracy of nonnative /f/, /z/, and /v/−initial obstruent clusters (e.g., /ft/, /zd/, /vb/) in pseudo-Czech words (which deviated significantly from ceiling) and the frequency of those sequences in other syllable positions across the English lexicon. Instead, Davidson suggested that the accuracy of nonnative cluster production is based on the phonetic similarity with other native clusters: /f/−nasal onset clusters display higher production accuracy compared to other types of nonnative clusters (e.g., stop–stop clusters) because the clusters are more phonetically similar to the native American English voiceless fricative-nasal (i.e., /sm/, /sn/) and voiceless fricative-liquid onset clusters (i.e., /fl/). In this view, speakers transfer their articulatory knowledge of native onset clusters to facilitate the production of nonnative /f/−nasal clusters.

While speakers of American English can execute /f/−nasal onset clusters with relatively high production accuracy, even with little to no motor practice, the most common error pattern found involves the presence of a vocoid (i.e., a vowel-like acoustic transition between consonants; Buchwald et al., 2019; Davidson, 2006, 2010; Wilson et al., 2014). Davidson (2010) further explored the nature of this "vowel insertion error" by comparing the acoustic properties of the vocoid segment to those of a lexical schwa produced in the same consonantal environment (e.g., /fən/ or /fəm/) during nonword production. Davidson reported that the vocoid segments were shorter in duration and had lower F1 than lexical schwa. These results demonstrate that the presence of a

vocoid likely reflects imprecise temporal coordination (i.e., a greater lag) among the flanking fricative and nasal gestures, rather than a pure insertion error of a lexical schwa to modify (or "repair") the nonnative sequences (but, see Shaw & Davidson, 2011).

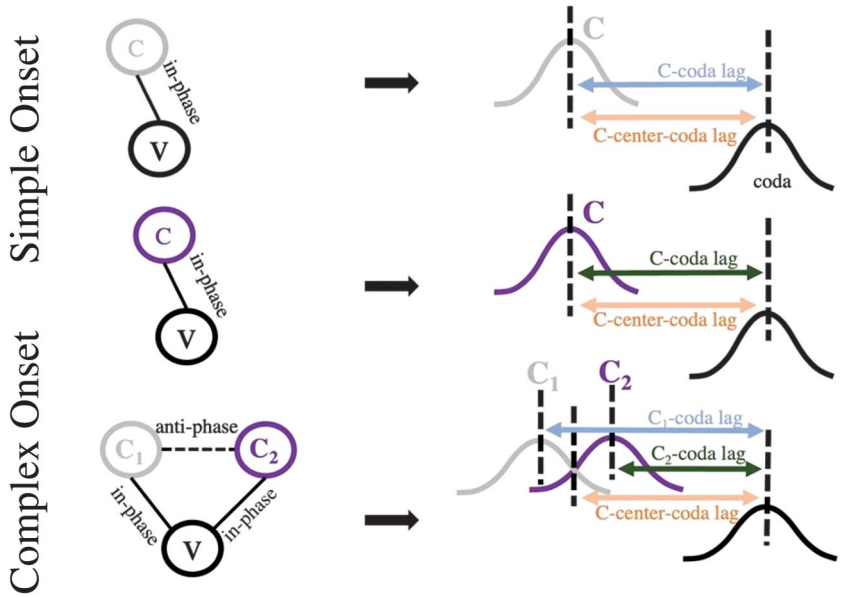
Another piece of evidence supporting the view advanced by Davidson (2006, 2010) comes from recent speech motor sequence learning studies (Buchwald et al., 2019; Buchwald & Cheng, 2023) in which native speakers of American English practiced producing nonwords with /fn/ onset clusters. Buchwald et al. (2019) reported that the acoustic duration of vocoids produced in /fn/ clusters significantly shortened throughout speech motor practice, likely reflecting improved temporal coordination indexed by increased temporal overlap of the gestures in question. In a subsequent study, Buchwald and Cheng (2023) also found that the acoustic duration of the nasal in the /f/−nasal clusters that were produced without acoustic evidence of a vocoid segment decreased significantly after a session of practice-based motor learning paradigm. Critically, this decrease in nasal duration could not simply be accounted for by a general increase in speaking rate throughout the course of production training. This finding is especially important in light of past studies showing that acoustic nasal duration is shorter in word-initial onset clusters (e.g., /#sm/) than in singletons (e.g., /#m/; Klatt, 1976; Marin & Pouplier, 2010). Thus, the incremental decrease in nasal duration observed by Buchwald and Cheng likely indicates that speakers continually refine the temporal coordination of the component consonantal gestures to become more natively "cluster-like" throughout motor practice. It is important to point out, however, that the aforementioned studies make inferences about the temporal coordination of nonnative /f/−nasal clusters based on acoustic analyses alone, which do not provide a direct measure of speech motor actions. In this study, we directly assessed the temporal coordination of nonnative /f/−nasal cluster production using EMA.

Temporal Coordination of Articulatory Gestures for Onset Clusters

While substantial research has examined the temporal coordination patterns for onset clusters, virtually all prior research has focused on the production of onset clusters in a speaker's native language. In a seminal study, Browman and Goldstein (1988) compared temporal coordination patterns of consonantal gestures affiliated with simple onsets (e.g., /sat/, /pat/, /lat/) and clusters (e.g., /splats/ vs. /spats/ vs. /plats/) in American English using X-ray microbeam data. In that study, the temporal lag between the maximum constriction of the primary articulator for each onset consonant (/s/, /p/, and /l/) and the leftmost coda consonant (i.e., /t/), which served as an anchor point to indirectly denote vowel offset, was measured. The results revealed that, regardless of onset complexity, the temporal lag between the midpoint of the onset as a whole and the coda anchor point remained relatively stable. This complex temporal coordination pattern was termed *C-center (consonant-center) timing* (Browman & Goldstein, 1988; see also Byrd, 1995); as onset complexity increases (CV ~ CCV [C as consonant, V as vowel]), the timing lag between the vowel-adjacent consonant and an anchor point decreases, indicative of an increased overlap with the vowel. Nam et al. (2009) proposed that onset-vowel articulatory timing patterns this way because all consonantal gestures in onset position, regardless of complexity, are triggered synchronously with the vocalic nucleus gesture (i.e., through in-phase coupling) while consonantal gestures are timed sequentially with respect to each other (i.e., through antiphase coupling; Browman & Goldstein, 2000; Goldstein et al., 2006; Nam & Saltzman, 2003; Saltzman & Byrd, 2000). This idea is schematized in Figure 1 (cf. Gafos et al., 2014; Shaw & Gafos, 2010, 2015). This model accounts for the finding that the temporal lag between the C-center and the anchor point remain relatively stable. In addition, this model further generates two testable predictions regarding this complex timing pattern. As the complexity of the onset increases from a simple onset to a cluster (CV ~ CCV), the gestures corresponding to the leftmost consonant in an onset cluster (C_1 in a #C_1C_2V sequence) are predicted to exhibit a decrease in temporal overlap with the vocalic gesture of the same syllable (henceforth, leftward shift), whereas the gestures corresponding to the rightmost consonant in a cluster (C_2 in a #C_1C_2V sequence) are predicted to exhibit increased temporal overlap with the vocalic gesture (henceforth, rightward shift), as schematized in Figure 1.

While "C_1V overlap" and "C_2V overlap" metrics may provide a useful diagnostic tool to examine the temporal coordination for onset clusters, several studies have demonstrated that there are cross-linguistic differences in the default temporal coordination pattern for onset clusters (Gafos et al., 2014; Hermes et al., 2017; Shaw et al., 2011). For example, in their study of Moroccan Arabic onset clusters, Shaw et al. (2011) reported a general pattern that when an onset became more complex (e.g., /lan/ ~ /flan/), the C_2V overlap was not larger in the cluster condition as compared to the matched singleton condition. In contrast to predictions based on C-center timing, their finding aligned more with a so-called "simplex timing" pattern, whereby the most stable temporal lag was between the rightmost consonant to an anchor point regardless of the number of consonants in the onset position (i.e., appending the leftmost consonant to a "simple onset").

Figure 1. Schematic representation of American English gestural coordination patterns for simple (CV) onsets (top) and complex cluster (CCV) onsets (bottom), relative to a constant anchor point denoting vowel offset. C = consonant; V = vowel.



Accordingly, one approach to examine the extent to which nonnative onset clusters (#C_1C_2) are temporally coordinated following C-center versus simplex timing is to compare the temporal lag between C_1 and an anchor point in the cluster condition to the corresponding lag in the C_1 singleton condition. The same comparison has been applied in studies of native clusters to examine the difference in the temporal lag between C_2 and an anchor point in the cluster condition to the C_2 singleton condition (Hermes et al., 2013; Marin, 2013; Marin & Pouplier, 2010; Pastätter & Pouplier, 2017; Pouplier & Beňuš, 2011). These two temporal overlap measures—here, *C_1V overlap* and *C_2V overlap*, respectively—index the degree to which the gestures corresponding to both the leftmost and the rightmost consonant in an onset cluster overlap with that of the flanking vocalic gesture relative to their singleton conditions.

The Current Research

In the present study, we used EMA (Perkell et al., 1992; Rebernik et al., 2021) to measure the temporal coordination patterns for nonnative /fn/ clusters produced by speakers of American English. Specifically, we used C_1V overlap and C_2V overlap measures to assess the temporal coordination of the component gestures affiliated with the consonants in /fn/ clusters as produced by American English speakers. While onset /f/−nasal clusters are not permitted in the phonotactics of American English, speakers generally are able to produce them in a perceptually acceptable way (Buchwald & Cheng, 2023; Davidson, 2006, 2010). However, the extent to which speakers apply C-center or simplex timing during their execution of this nonnative sequence remains unclear. We compared American English speakers' temporal coordination of the successive consonantal gestures affiliated with nonnative /fn/ clusters to that of native onset /fl/ or /sm/ clusters, given that the /fn/ cluster shares a high degree of phonetic similarity with these native onset clusters.

In our speaking task, participants read disyllabic nonwords (e.g., FNAGDEEP, /fnægdip/) embedded in various carrier phrases while their articulatory movements were recorded and tracked using EMA. While prior studies have elicited nonnative onset clusters using a repetition task (Pouplier et al., 2020), we used a reading task to avoid production errors that arise from perception, as the perception of nonnative clusters is known to be difficult (Berent et al., 2007; Dupoux et al., 2011) and the content of the auditory model can affect the temporal coordination of nonnative clusters (Pouplier et al., 2020; Wilson et al., 2014).

The goal of the current research was to determine how the articulatory gestures that comprise nonnative /fn/ clusters are temporally organized relative to those of native /fl/ and /sm/ clusters. We reasoned that if American English speakers are readily able to apply the temporal organization associated with native onset clusters (i.e., C-center timing; Browman & Goldstein, 1988) to nonnative clusters, then we would expect to observe the following as onset complexity increases (CV ~ CCV): (a) Across all three clusters, the temporal overlap of the leftmost

consonantal gesture and the vocalic gesture will decrease relative to their singleton forms (i.e., leftward shift); and (b) across all three clusters, the temporal overlap of the rightmost consonantal gesture and the vocalic gesture will increase as compared to their singleton forms (i.e., rightward shift).

Alternatively, if speakers append the leftmost consonant to a "simple onset" (i.e., simplex timing; Shaw et al., 2011) to facilitate the execution of articulatorily unfamiliar sequences, then we would expect to observe (a) the overlap of the leftmost consonantal gesture and the vocalic gesture in /fn/ clusters to decrease as compared to its singleton form (i.e., leftward shift) and (b) the overlap of the rightmost consonantal gesture and the vocalic gesture in /fn/ clusters to exhibit comparable temporal overlap to its singleton form (i.e., no rightward shift).

To determine the relative changes in C_1V and C_2V overlap between the cluster and singleton forms, we computed C_1V and C_2V overlap ratios. These ratios were computed based on the approach used in the work of Pastätter and Pouplier (2017) and were defined as the temporal lag between the maximum vocal tract constriction for the coda consonant and the maximum constriction for either the C_1 or C_2 consonant in a cluster divided by the temporal lag between the maximum constriction for the coda consonant and the maximum constriction for the consonantal in the singleton form (see more details below).

An important methodological point is that, while measuring C_2V overlap ratios, prior studies (e.g., Pastätter & Pouplier, 2017) have elicited clusters and their matched singletons using the same carrier phrase (e.g., "I say ____ again"). One potential confound in this approach lies in that the onset singletons are not preceded by a consonant, creating a mismatch largely corresponding to the position of the jaw (J) prior to the target C_2 segment in the cluster condition. That is, using /sm/ cluster as an example, the target C_2 is preceded by a consonant (i.e., "I say s[**m**]ug again"), resulting in the elevated J position prior to the GONS of the C_2. In contrast, the position of the J is at a lower place in the singleton condition (i.e., "I say [**m**]ug again"). This difference in the spatial configuration of the vocal tract prior to the production of the C_2 gesture could also lead to more C_2V overlap in the cluster condition if the C_2 is able to be achieved more quickly due to a relatively higher J position (Shaw & Chen, 2019; but see, Liu et al., 2022). One way to address this potential confound is to use a different carrier phrase for the singleton condition that matches the consonant that precedes the C_2 segment in the cluster condition (e.g., "I bus [**m**]ug again"), thereby largely controlling for the pre-C_2 J position. Several prior studies (Brunner et al., 2014; Jacobi, 2022) have observed C-center timing for native clusters even when the pre-C_2 J position is controlled. We here explored whether there was a measurable difference in C_2V overlap ratios using different types of carrier phrase for the matched singleton conditions (i.e., with and without controlling for the pre-C_2 J position).

Method

Participants

Twelve native speakers of American English (one male participant, 11 female participants; ages 19–21 years, M_{age} = 20 years), who were not aware of the specific purposes of the research, were originally recruited for this study. However, data from three participants (all female) were excluded from the final analyses because the EMA sensor located on the tongue body (TB; see details below) fell off during the recording session. The final sample consisted of nine speakers (one male and eight female participants; ages 19–21 years, M_{age} = 20 years). None of the participants reported a history of a speech, language, hearing, or other neurological disorder. Participants completed the Language Experience and Proficiency Questionnaire (Marian et al., 2007) to verify that they were functionally monolingual speakers. Participants had to meet the following inclusion criteria: (a) raised in a monolingual English home and educated in a monolingual English school, (b) no experience learning an L2 before 10 years of age, and (c) no experience conversing in an L2 on a regular basis (i.e., no consistent, daily exposure to an L2 while interacting with friends or family, watching TV, or listening to radio/music). All participants self-reported learning Spanish as an L2 in school. Critically, however, none of them had experience with Polish, Russian, Czech, German, or any other languages allowing onset fricative-nasal clusters that are nonnative in American English (e.g., /fn/).

Speech Stimuli

The speech stimuli consisted of nonwords with native onset clusters and nonnative onset clusters (referred to as native and nonnative CC nonwords, respectively). The syllable structure of each CC nonword was C_1C_2VC. CVC (e.g., /fnægdip/). The nonnative clusters were all fricative-nasal clusters (/fn/), and the native English clusters were fricative-nasal (/sm/) and fricative-liquid (/fl/) sequences. We selected the native clusters to be as phonetically similar as possible to the nonnative clusters in terms of their place and manner of articulation while also satisfying the constraint that each cluster be produced with heterorganic articulators, given the difficulty of identifying successive consonantal constrictions of homorganic

segments in EMA data. The /fl/ clusters matched the /fn/ clusters in their places of articulation (labiodental followed by alveolar), and the /sm/ clusters matched the /fn/ clusters in their manner of articulation (fricative followed by a nasal). We used two nonwords beginning with each cluster. For consistency, all nonwords had velar stops in the coda of the first syllable, had bilabial and alveolar stops in the second syllable, and were produced with a trochaic stress pattern, which facilitated identification of consonantal constrictions in the EMA time functions by maximizing lip, J, and tongue movement excursions in the vertical dimension.

For each CC nonword (e.g., /smægdip/), we included two corresponding singleton nonwords that contained the C_1 consonant (e.g., /sægdip/, henceforth, C_1 nonword) and the C_2 consonant (e.g., /mægdip/; henceforth, C_2 nonword) with the rest of the nonword being identical. In addition, we included several control stimuli that consisted of nonwords with each C_1C_2 sequence produced in a heterosyllabic context (e.g., /#sm/ vs. /s#m/; henceforth, $C_1\#C_2$ nonwords).

To control the vocal tract configuration prior to stimulus production, all nonwords were embedded in a carrier phrase. We used different sets of carrier phrases for the nonwords in each stimulus condition. For the C_1C_2, C_1, and C_2 nonwords, the carrier phrase used depended on the constriction location of the initial consonant. Nonwords that began with a vocal tract constriction location at the alveolar ridge (e.g., /smægdip/ or /sægdip/) were embedded in the carrier phrase, "I have a _____ again." C_1C_2, C_1, and C_2 nonwords that began with a constriction at a location other than the alveolar ridge (e.g., /flægdup/) were embedded in the carrier phrase, "I gotta _____ again." In this way, the consonant in the carrier sentence that preceded the initial consonant in the target word had a different constriction location, which, in turn, facilitated their identification in the obtained EMA time functions (see details below). The $C_1\#C_2$ nonwords were embedded in different carrier phrases. To create the /s#m/ environment, the nasal C_2 nonword (e.g., /mægdip/) was embedded in the carrier phrase, "I bus _____ again." Both the /f#l/ and /f#n/ sequences were created by embedding the corresponding C_2 nonword (e.g., /lægdup/ or /nægbit/) in the carrier phrase, "I puff _____ again."

All nonwords were visually cued with American English orthography without an auditory model. Thus, any observed differences in articulatory timing across clusters could not be attributable to online imitation of a model speaker. The orthographic model for each nonword was verified by native speakers of American English to ensure that the intended grapheme–phoneme correspondence would likely be elicited (see Table 1 for the full list of stimuli).

Table 1. International Phonetic Alphabet (IPA) transcription, American English orthography, and carrier phrase for each nonword in each experimental condition.

Condition	Onset	IPA	Orthography	Carrier phrase
Native CC nonwords	/sm/	smægdip	SMAGDEEP	I have a SMAGDEEP again
		smækput	SMACKPOOT	I have a SMACKPOOT again
	/fl/	flægdup	FLAGDOOP	I gotta FLAGDOOP again
		flækpit	FLACKPEET	I gotta FLACKPEET again
Nonnative CC nonwords	/fn/	fnægbit	FNAGBEET	I gotta FNAGBEET again
		fnæktup	FNACKTOOP	I gotta FNACKTOOP again
C_1 nonwords	/s/	sægdip	SAGDEEP	I have a SAGDEEP again
		sækput	SACKPOOT	I have a SACKPOOT again
	/f/	fægdup	FAGDOOP	I gotta FAGDOOP again
		fækpit	FACKPEET	I gotta FACKPEET again
C_2 nonwords	/m/	mægdip	MAGDEEP	I gotta MAGDEEP again
		mækput	MACKPOOT	I gotta MACKPOOT again
	/l/	lægdup	LAGDOOP	I have a LAGDOOP again
		lækpit	LACKPEET	I have a LACKPEET again
	/n/	nægbit	NAGBEET	I have a NAGBEET again
		næktup	NACKTOOP	I have a NACKTOOP again
C#C nonwords	/m/	mægdip	MAGDEEP	I bus MAGDEEP again
		mækput	MACKPOOT	I bus MACKPOOT again
	/l/	lægdup	LAGDOOP	I puff LAGDOOP again
		lækpit	LACKPEET	I puff LACKPEET again
	/n/	nægbit	NAGBEET	I puff NAGBEET again
		næktup	NACKTOOP	I puff NACKTOOP again

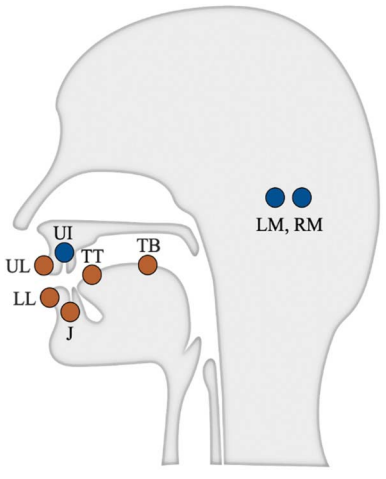
Note. C = consonant.

Instrumentation, Experimental Design, and Procedure

J, lip, and tongue movements were recorded using a Carstens AG501 EMA system (Carstens Medizinelektronik GmbH) in a quiet room. EMA tracks speech movements using electromagnetic fields oscillating at kilohertz rates, generated by a series of transmitter coils, to localize the positions of small sensors temporarily affixed on the surfaces of the articulators during a speaking task (Rebernik et al., 2021). The oscillating transmitter coils, positioned above the head of the speaker at different orientations, induce a current in the sensor coils whereby the coils are tracked with high spatial and temporal resolution.

Articulatory sensors were placed on the upper lip (UL), lower lip (LL), J, tongue tip (TT), and the TB with head reference sensors placed on the gingiva above the upper incisor (UI) and behind each ear on the left mastoid (LM) and right mastoid (RM) processes (see Figure 2). The UI-reference sensor was placed intraorally on the gingiva of the UIs using a piece of Stomahesive wafer. The LM- and RM-reference sensors were firmly placed using a medical tape. Prior to the speaking task, each participant's occlusal plane was obtained by attaching three sensors to the Carstens biteplane that the participant held between their upper and lower teeth while data were recorded for these three sensors, as well as the reference sensors. After the occlusal plane was established, sensors were attached to the midsagittal surface of the UL, LL, J, TT, and TB. The UL- and LL-movement sensors were placed on the vermillion border of the upper and lower lips, respectively, using a nontoxic dental glue (EPIGLU, Meyer-Haake). The J-movement sensor was placed intraorally on the gingiva of the lower incisors using a piece of Stomahesive wafer. The TT-movement sensor was placed approximately 1 cm behind the anatomical TT, and the TB-measurement sensor was placed approximately 4 cm behind the TT-movement sensor using a dental glue. A surgical ruler was used to make these measurements, and the measurements were performed with the tongue stretched outside of the mouth. The locations of the lingual sensors were marked on the lingual surface with a dental color applicator (Thompson Dental Company) in case the sensors detached mid-experiment. The tongue surface was dried prior to sensor placement, using a piece of dental gauze, to help improve adhesion (participants were also instructed to refrain from drinking and eating for at least 30 min prior to the experimental session). Participants were engaged in spontaneous conversation for approximately 10 min prior to the start of the experiment to allow some time to habituate to talking with the sensors mounted on the articulators (Dromey et al., 2018; Weismer & Bunton, 1999).

Figure 2. Placement of electromagnetic articulography sensors in the current experiment. Orange dots mark articulatory movement sensors; blue dots mark head reference sensors. LM = left mastoid; RM = right mastoid; UI = upper incisor; UL = upper lip; LL = lower lip; J = jaw; TT = tongue tip; TB = tongue body.



Participants were positioned under the transducer coils of the EMA system and in front of a computer monitor. Participants read randomized lists of the nonwords embedded in the carrier phrases, which were orthographically presented in the center of the screen. Participants produced the nonwords at a normal (self-determined) production rate and were instructed to produce each nonword as accurately as possible, making sure to produce all segments in the orthographic display. The mapping between the orthography and the intended phoneme string in each nonword was straightforward, and apart from the initial onset cluster, the rest of each word followed native English phonotactics (see Table 1). Any segmental errors or other inconsistencies in pronunciation were detected by our perceptual and acoustic analysis procedures (described below). Participants were given a series of practice trials prior to the main experimental test session. During these practice trials, participants were asked to produce several repetitions of four nonwords that contained either nonnative onset /ft/ clusters or /fm/ clusters. Participants were given feedback to confirm that they understood the instructions and were able to perform the speaking task. The stimuli used in these initial practice trials were not presented at any point in the rest of the study. The nonwords in the test trials were not auditorily modeled to the participants, nor was feedback given during the test trials, as this was not a motor training study (see, e.g., Buchwald & Cheng, 2023) and the content of auditory models has been shown to alter how speakers coordinate their articulatory movements (Pouplier et al., 2020; Wilson et al., 2014).

Participants were recorded while producing eight repetitions of each nonword, resulting in 288 tokens collected per subject. Nonwords were orthographically cued randomly. The stimuli were sequenced and presented using PsychoPy (Peirce, 2007). Simultaneous audio (.wav format files; sampled at 48 kHz with 16 bits per sample) was recorded using a shotgun microphone (t.bone EM9600) and the EMA sensor signals (approximately 0.3-mm spatial resolution at 250-Hz sampling rate) were recorded using Carstens' CS5RECORDER and CS5VIEW programs (Carstens Medizinelektronik GmbH).

Postprocessing and Analysis

Perceptual and acoustic analyses. Each recorded utterance that contained the nonnative /fn/ cluster was acoustically analyzed for production accuracy using Praat (Boersma & Weenink, 2023) by one trained research assistant. As the coder was not a native speaker of a language that contains the onset /fn/ cluster, we used an acoustically informed transcription approach because prior work has shown that nonnative listeners are poor at perceptually determining the production accuracy of many nonnative clusters (Davidson, 2011; Davidson & Shaw, 2012). We also quantified the extent to which participants produced the /fn/ cluster with an intervening vocoid when measured acoustically, as this behavior has been consistently reported in prior production experiments with speakers of American English (Buchwald & Cheng, 2023; Davidson, 2006, 2010). This was accomplished by identifying the presence of a vocoid structure, which was determined based on whether there were (a) at least two repetitive complex wave cycles and (b) a higher formant structure (i.e., F2 and F3) in the waveform and spectrogram (Wilson et al., 2014). We included /fn/ tokens that were either produced with or without a vocoid in the kinematic analyses (described below) as prior studies have suggested that the presence of a vocoid likely arises from a longer temporal lag between the consonantal gestures rather than insertion per se of a vowel segment (Buchwald et al., 2019; Davidson, 2006, 2010). Other phonemic errors such as substitutions (e.g., /fnægbit/ → [fmægbit]), metathesis (e.g., /fnægbit/ → [nfægbit]), and deletions (e.g., /fnægbit/ → [nægbit]) were also coded and verified in the acoustic signal. Utterances that were coded with having a phonemic error were excluded from the subsequent kinematic analyses (see below).

Additionally, to ensure that all target stimuli were produced with the trochaic stress pattern, the same coder perceptually evaluated the stress patterns for all utterances. As this was the only measurement based on perceptual judgments, the coding was verified by the first author of this study. We reported that only three utterances were produced with the incorrect stress pattern (i.e., iambic pattern) and were excluded from the kinematic analyses reported below.

Kinematic analyses. The following kinematic analyses were performed on utterances that were produced without phonemic errors (e.g., deletion, substitution) based on acoustic records as well as stress errors. This resulted in the removal of a total of 46 utterances across all participants (approximately 2% of the total utterances recorded). The raw kinematic data were processed using the MATLAB-based (The Mathworks, Inc., 2019, Version R2019b) *Mview* algorithms (Tiede, 2005, 2010). The raw kinematic data first underwent a series of standardized preprocessing steps to rotate and translate each position signal to a consistent maxillary frame of reference (based on the location of UI-reference sensor) and to correct for head motion artifacts (on the basis of the three static head reference sensors). The acoustic (.wav format files) and kinematic movement signals (.pos files) were then synchronized and visualized together (.mat files) in the *Mview* graphical user interface (Tiede, 2005, 2010).

The postprocessed movement time series obtained for each articulator were semi-automatically labeled using the *lp_findgest* function using *Mview*, which generates a set of labels (temporal landmarks in milliseconds) for a given vocal tract constricting action. These labels were visually inspected and manually edited for tracking errors. The time point of the maximum vocal tract constriction (in the vertical dimension) affiliated with each onset and coda consonant in the first syllable of the nonwords was manually identified in the movement time functions of the relevant articulator. Labiodental consonants (/f/) were defined on the basis of the maximum displacement of the LL-movement sensor (without J subtraction). Consonants having a primary constriction at the alveolar ridge (/s, n, l/) or at the soft palate (velum; /k, g/) were defined on the basis of the maximum vertical displacement of the TT- and TB-movement sensors, respectively. Bilabial consonants (/m/) were defined on the basis of lip aperture (LA), that is, the vertical Euclidean distance between the sensors attached to the UL- and LL-movement sensors. (Note that LA is a *constriction degree* measure [in this case, between the UL and LL], whereas all other articulatory measurements extracted were *constriction point* measures [maximum vertical displacement of a given sensor].) Then, once the maximum vertical constriction (MAXC) for a given consonant was manually identified and labeled in the appropriate articulator time series, *Mview* automatically labeled (a) the time point of the gestural onset (*GONS*; i.e., when the articulator first begins to accelerate toward the MAXC), defined as the time point when the velocity exceeds 20% of the peak velocity preceding the MAXC; (b) the time point of the onset of the gestural nucleus (*nucleus onset [NONS]*), defined as the time point when the velocity exceeded 80% of the peak velocity preceding MAXC; (c) the

time point of MAXC, defined as the time point at which the relevant articulator achieved its maximum vertical displacement (with reference to the UI)[1]; (d) the time point of the offset of the gestural nucleus (*nucleus offset [NOFFS]*), defined as the time point when the velocity dropped below 80% of the peak velocity following MAXC; and (e) the time point of gestural offset, defined as the time point when the velocity drops below 20% of the peak velocity preceding MAXC.

While various landmarks have been used in prior studies, including MAXC (Browman & Goldstein, 1988; Marin & Pouplier, 2010) and peak velocity (Pastätter & Pouplier, 2017; Pouplier et al., 2020), we opted to use MAXC because it is most readily identifiable by visual inspection after semi-automatic identification. The extracted MAXC labels were then used to calculate both C_1V and C_2V overlap ratios. These two ratios quantified the relative change in C_1V and C_2V overlap between each cluster and their corresponding singleton conditions. Both ratios were computed adopting the approach used in the work of Pastätter and Pouplier (2017), and here, we describe the calculation of C_2V overlap ratio as an example. As shown in Figure 3, the C_2V overlap was defined as the temporal lag between the MAXC of the coda consonant and the MAXC of the C_2 consonant in a cluster (C_2 – coda lag; e.g., [f**l**æ**g**dup])[2] and in their singleton form with (i.e., [f#**l**æ**g**dup]) and without (i.e., [#**l**æ**g**dup]) controlling for the difference in a prestimulus J position. Prior to computing the C_2V overlap ratios, outliers were identified and removed using the median absolute deviation (MAD) approach on a subject-by-subject and word-by-word basis (Leys et al., 2013). Each participant's C_2 – coda lag for each nonword was compared to their own median value. Values that fell 2 MAD away from the median were removed.

Given that there was a small number of tokens per nonword and per participant, we computed the C_2V overlap ratio for each cluster by comparing each C_2 – coda lag in the cluster condition to the entire distribution of the C_2 – coda lags in the singleton condition on a by-participant and by-word basis. Thus, the C_2V overlap ratio for each cluster was calculated as dividing each C_2 – coda lag in the cluster condition by each of the C_2 – coda lags in the corresponding singleton conditions ([MAXC $C_{2\ \text{complex onset}}$ – MAXC coda] / [MAXC $C_{2\ \text{singleton onset}}$ – MAXC coda]). C_2V overlap ratios of 1 would indicate no change in C_2V overlap in the cluster compared to the singleton condition, whereas a ratio of less than 1 would indicate an increase in C_2V overlap with increasing onset complexity.

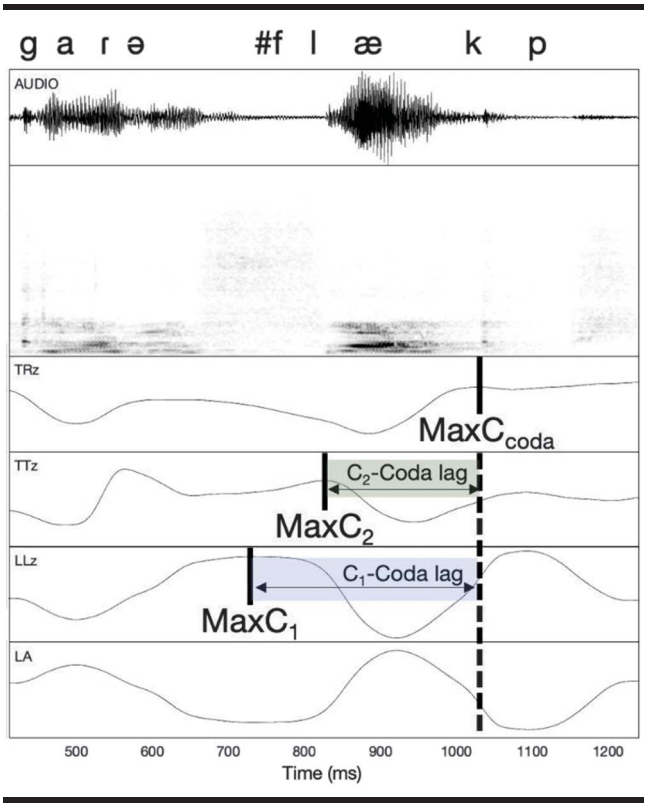


Figure 3. Illustration of articulatory timing measures taken from the electromagnetic articulography kinematic signals. FLACKPEET (/flækpit/) produced by one of the participants (in the carrier phrase "I gotta ____ again"). The panels show, from top to bottom: audio, spectrogram, tongue body (vertical; TRz), tongue tip (vertical; TTz), lower lip (vertical; LLz), and lip aperture (vertical; LA) positions. The green arrows (i.e., C_2 – coda lag) indicate the temporal interval between the MAXC of C_2 (/n/) and the coda (/g/), which served as an anchor point, and the blue arrows (C_1 – C_2 lag) indicate the temporal interval between successive MAXC of C_1 and C_2 in the /fn/ onset cluster.

Statistical Analyses

To first determine whether there were differences in temporal organization across the three consonant clusters, we built separate linear mixed-effects models for C_1V overlap ratios and C_2V overlap ratios, respectively, using R (R Core Team, 2017, Version 3.4.3) in RStudio (RStudio Team, 2023, Version 2023.09.0 + 463) with the *lme4* package (Bates et al., 2015). All categorical variables in the current analyses were sum-coded. To model C_1V overlap ratios, the categorical variable of sound represents each cluster type (/fl/ vs. /sm/ vs. /fn/) and was included as the main fixed-effect predictor in the model. To model C_2V overlap ratios, in addition to the sound variable, another categorical variable, J, which denoted whether the

[1]In some cases, an articulator achieved a maximum vertical displacement and plateaued for a period of time before lowering. In these cases, we laid the MAXC label at the temporal midpoint of the plateau.

[2]Note, however, that Pastätter and Pouplier (2017) computed C_2V overlap based on the intervals between the peak velocity preceding MAXC (rather than the MAXC itself) of the vowel-adjacent consonant and the peak velocity leading into the MAXC of the coda consonant for each cluster and singleton word.

J position prior to the C_2 consonant was controlled (yes vs. no), was also included as a fixed-effect predictor. The interaction term between sound and J was further included to examine whether there was a magnitude difference across all three clusters. Guided by the statistical approaches of Harel and McAllister (2019), we selected the optimal random-effects structure using the model comparison approach based on both the Akaike information criterion (AIC; Akaike, 1974) and Bayesian information criterion (BIC; Schwarz, 1978) values. To do so, we built two separate models that differed only in the random-effects structures while the fixed-effects structure remained constant. The first model included random intercept of participant and by-participant random slope for the fixed-effect predictor, sound. The second model included only random intercept of participant. The model that had the smallest AIC and BIC values was chosen as the final model.

Because all categorical fixed-effect predictors were sum-coded, after the best fitting model was selected, we used the *anova* function in R to test the significance of the fixed-effects predictors in the final selected model. In cases of significance, post hoc multiple comparisons were conducted using the *emmeans* package (Lenth, 2022) to compare the levels of each significant fixed-effect predictor. The alpha level was .05, and we adjusted all *p* values for multiple comparisons using the Holm method (Holm, 1979). The model assumptions were checked using the "performance" package (Lüdecke et al., 2021). All data organization, cleaning, and visualization were done using the *tidyverse* package (Wickham et al., 2019).

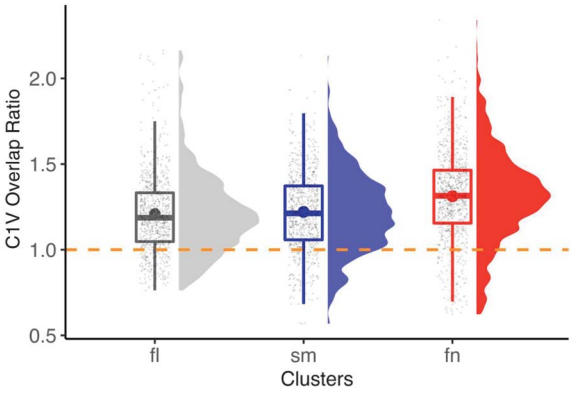
Results

This section consists of two parts. The first part reports on the overall accuracy rates for the /fn/ clusters based on our perceptual and acoustic analysis procedures. The second part reports on the results of the kinematic analyses focused on assessing the C_1V and C_2V overlap ratios for all three cluster types (/fn/ vs. /fl/ vs. /sm/).

Acoustic Analysis: Production Accuracy for Nonnative /fn/ Clusters

We first report the results of the perceptual and acoustic evaluation of all /fn/ utterances in our data set. Out of a total of 304 /fn/ tokens (pooled across all participants), only 29 utterances contained acoustic evidence of a vocoid, and only six utterances were coded with having a phonemic error (e.g., deletion). This resulted in an overall accuracy rate of 88% for the /fn/ cluster. This is consistent with prior studies that categorized tokens containing a vocoid as errors (Buchwald & Cheng, 2023). Again, only utterances that were coded with having a phonemic error were excluded from the kinematic analyses.[3]

Figure 4. Box plots of C_1V overlap ratios for onset clusters (/fl/ vs. /sm/ vs. /fn/) indicating the relative changes of C_1V overlap between the cluster and the singleton conditions.



Kinematic Analysis: Temporal Coordination of Gestures for Native and Nonnative Clusters

C_1V Overlap Ratios

A C_1V overlap ratio that is greater than 1 would indicate that the temporal lag between the target C_1 consonant and the vocalic gesture in the cluster condition (e.g., **f**na**g**beet) is longer than the temporal lag between the target C_1 consonant and the vocalic gesture in the matched singleton condition (e.g., **n**a**g**beet). If the onset clusters were temporally organized following either the C-center timing pattern (Browman & Goldstein, 1988; Pastätter & Pouplier, 2017) or the simplex timing pattern (Shaw et al., 2011), we would expect to observe the C_1V overlap ratio to be larger than 1, which would be indicative of a leftward shift. As shown in Figure 4, the native /fl/ cluster showed the smallest C_1V overlap ratio (M = 1.21, standard deviation [SD] = 0.23), followed by the native /sm/ cluster (M = 1.22, SD = 0.23). The nonnative /fn/ had the largest C_1V overlap ratio (M = 1.31, SD = 0.27). Overall, C_1V overlap ratios for all three clusters were larger than 1, reflecting that the temporal lag between the C_1 consonants and the anchor point was longer in the cluster condition than in the singleton condition. With respect to the inferential statistics, the best fitting linear-mixed effects model chosen by AIC and BIC was the model that contained both random intercept of participant as well as random slope of participant for the sound variable. The model revealed that there was not a main effect of sound, $F(2, 8) = 2.42$, $p = .15$, indicating that there are no statistically

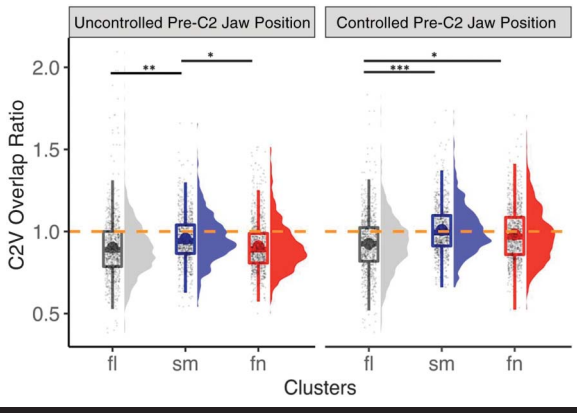
[3]Kinematic analyses were repeated with the exclusion of tokens containing an epenthetic vocoid; the same pattern of results were obtained.

reliable differences in the C_1V overlap ratios among the three clusters. Taken together, the results on C_1V overlap indicate that the C_1 consonants in the cluster condition were shifted leftward in terms of temporal lag relative to their matched singleton condition, aligning with the predictions of both C-center and simplex timing for onset clusters.

C_2V Overlap Ratios

Similar to the interpretation of C_1V overlap ratio, a C_2V overlap ratio that is greater than 1 suggests that the temporal lag between the C_2 consonant and the anchor point in the cluster condition (e.g., f**nag**beet) is longer than the temporal lag between the C_2 consonant and the anchor point in the matched singleton condition (e.g., **nag**beet). If the onset clusters display a C-center timing pattern (Browman & Goldstein, 1988; Pastätter & Pouplier, 2017), then we would expect the C_2V overlap ratio to be smaller than 1, representing a rightward shift of the C_2 consonant in the cluster condition. Alternatively, if the onset cluster displays a simplex timing pattern (Shaw et al., 2011), then we would expect the C_2V overlap ratio to hover around 1, indicating no leftward or rightward shift. We first present the C_2V overlap ratio that was calculated using the singleton condition that did not control for the pre-C_2 J position (i.e., C_2 nonwords). As shown in the left panel of Figure 5, the native /fl/ cluster showed the lowest C_2V overlap ratio (M = 0.90, SD = 0.19) and was followed by the C_2V overlap ratio for the nonnative /fn/ cluster (M = 0.91, SD = 0.15). The /sm/ cluster had the largest C_2V overlap ratio (M = 0.96, SD = 0.15). With respect to the inferential statistics, the best-fitting model that was chosen by AIC and BIC value was the model that included both random intercept of participant and random slope of sound for participant. The model revealed that there was a significant main effect of sound, $F(2, 7.9) = 10.33$, $p = .0006$, as well as a significant main effect of J, $F(1, 5414.8) = 130.15$, $p < .0001$. Additionally, there was a significant interaction between sound and J, $F(2, 5413.4) = 11.02$, $p < .0001$. Post hoc multiple comparisons showed that, when pre-C_2 J position was not controlled for the singleton condition, the C_2V overlap ratio for /fl/ was significantly smaller than the C_2V overlap ratio for /sm/ ($\beta = -0.06$, $SE = 0.02$, $p = .003$). The C_2V overlap ratio for /fn/ was significantly smaller than the C_2V overlap ratio for /sm/ ($\beta = -0.05$, $SE = 0.02$, $p = .05$). There was not a significant difference in the C_2V overlap ratio between /fl/ and /fn/ ($\beta = -0.01$, $SE = 0.02$, $p = .71$). With respect to the simple effect of J across each level of sound, C_2V overlap ratio with controlled pre-C_2 J position was significantly smaller than that without controlling for pre-C_2 J position for /fl/ ($\beta = -0.02$, $SE = 0.01$, $p = .003$), /sm/ ($\beta = -0.05$, $SE = 0.01$, $p < .0001$), and /fn/ ($\beta = -0.07$, $SE = 0.01$, $p < .0001$) cluster, respectively. As the C_2V overlap ratios for all the clusters were smaller than 1, the temporal lag between the C_2 consonant to the anchor point was shorter in the cluster condition than in the singleton condition. This result indicates a rightward shift for all three clusters and is compatible with predictions that follow for C-center timing.

Figure 5. Box plots of C_2V overlap ratios for onset clusters (/fl/ vs. /sm/ vs. /fn/) indicating the relative changes of C_2V overlap between the cluster and the singleton conditions when the pre-C_2 jaw position was uncontrolled (left panel) versus controlled (right panel). ***$p < .001$.



However, a different picture emerged when we examined the C_2V overlap ratio that was calculated using the singleton condition that did control for the pre-C_2 J position (i.e., C#C nonwords). As shown in the right panel of Figure 5, the native /fl/ cluster showed the lowest C_2V overlap ratio (M = 0.93, SD = 0.18) and was followed by the C_2V overlap ratio for the nonnative /fn/ cluster (M = 0.98, SD = 0.17). The /sm/ cluster had the largest C_2V overlap ratio (M = 1.01, SD = 0.15). The C_2V overlap ratio for the /fl/ cluster was significantly smaller than the C_2V overlap ratio for the /sm/ cluster ($\beta = -0.07$, $SE = 0.016$, $p < .0001$) and the C_2V overlap ratio for the /fn/ cluster ($\beta = -0.06$, $SE = 0.02$, $p = .03$). The magnitude of C_2V overlap ratio did not differ significantly between /sm/ and /fn/ clusters ($\beta = 0.03$, $SE = 0.02$, $p = .14$). Based on the values of C_2V overlap ratio, there was a rightward shift for the /fl/ clusters. In contrast, the C_2V overlap ratio for /sm/ and /fn/ were similar to each other and hovered around 1. This suggests a lesser degree of rightward shift as the temporal lag between the C_2 consonant to the anchor point in the cluster condition was similar to the temporal lag in the singleton condition. We will return to this point in the Discussion section.

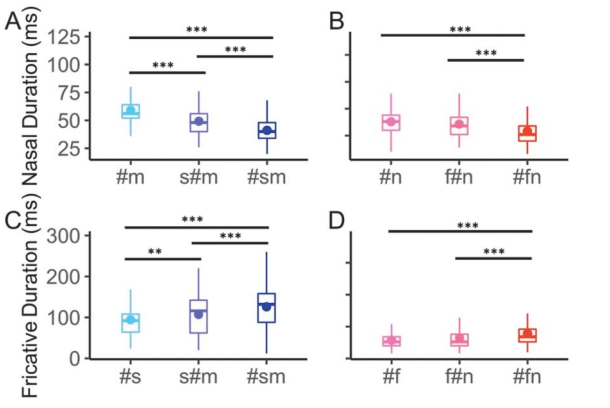
Post Hoc Analysis: Kinematic Nasal and Fricative Duration

The aforementioned kinematic analyses revealed that, when the pre-C_2 J position was controlled, there was

no statistically significant difference between the C_2V overlap ratios for fricative-nasal sequences that were orthographically cued as falling within the same syllable (/#sm/ and /#fn/) as opposed to those that spanned two syllables (/s#m/ and /f#n/; see Figure 5, right panel). To verify that our participants reliably distinguished between the two syllable affiliations in their productions, we performed a post hoc analysis to compare the kinematic duration of each nasal and fricative segment between the two types of sequences. Recall that the acoustic nasal duration has previously been reported to be shorter in clusters compared to singletons (Klatt, 1976; Marin & Pouplier, 2010). The kinematic duration of each segment was defined as the interval between the onset of the gestural nucleus (*NONS*), defined as the time point when the velocity exceeded 80% of the peak velocity preceding MAXC, and the offset of the gestural nucleus (*NOFFS*), defined as the time point when the velocity dropped below 80% of the peak velocity following MAXC.

The kinematic duration of the nasal segments in the native and nonnative sequences are given in Figures 6A and 6B. The kinematic duration of /m/ (see Figure 6A), on average, was the longest in the simple onset (#C) condition (M = 59 ms, SD = 13.3), followed by the heterosyllabic (C#C) condition (M = 49.2 ms, SD = 14.4). The /m/ in the onset cluster (#CC) condition had the shortest kinematic duration (M = 41.1 ms, SD = 9.34). The best-fitting model was the model that included random intercept of participant. The model revealed that the kinematic duration in the simple onset condition was significantly longer than the /m/ duration in the heterosyllabic condition (β = 9.95, SE = 1.31, p < .0001) and in the onset cluster condition (β = 17.83, SE = 1.31, p < .0001). The kinematic nasal duration in the heterosyllabic condition was significantly longer than that in the cluster condition (β = 7.88, SE = 1.32, p < .0001).

Figure 6. Box plots of kinematic nasal (A: /m/, B: /n/) and fricative duration (A: /s/, B: /f/). ***p < .001.



In terms of the kinematic duration of /n/ (see Figure 6B), on average, the longest duration was observed in the simple onset condition (M = 75.8 ms, SD = 23), followed by the heterosyllabic condition (M = 71.4 ms, SD = 27). Likewise, the kinematic duration of /n/ in the complex onset context had the shortest duration (M = 60.3 ms, SD = 35). The statistical model revealed that the kinematic duration of /n/ in the simple onset condition was significantly longer than that in the onset cluster condition (β = 16.12, SE = 3.14, p < .0001). No significant difference was found in duration between the singleton condition and the heterosyllabic condition (β = 5.01, SE = 3.13, p = .11). Kinematic nasal duration was longer in the simple heterosyllabic condition than that in the cluster condition (β = 11.11, SE = 3.12, p = .0008). Thus, across both sequence types, the nasal duration is significantly shorter in the cluster condition than in the singleton conditions.

The kinematic duration of the fricative segments in the native and nonnative sequences are given in Figures 6C and 6D. The kinematic duration of /s/ (see Figure 6C), on average, was the shortest in the simple onset (#C) condition (M = 94.3 ms, SD = 44.1), followed by the heterosyllabic (C#C) condition (M = 107 ms, SD = 51.8). The /s/ in the onset cluster (#CC) condition had the longest kinematic duration (M = 126 ms, SD = 52.7). This descriptive pattern was confirmed by the results of the statistical modeling. The model revealed that the kinematic duration in the simple onset condition was significantly shorter than that in both the heterosyllabic (β = −13.1, SE = 4.66, p = .005) and the onset cluster condition (β = −32.6, SE = 4.66, p < .0001). The kinematic /s/ duration in the heterosyllabic condition was significantly shorter than that in the onset condition (β = −19.44, SE = 4.65, p < .0001).

In terms of the kinematic duration of /f/ (see Figure 6D), on average, the shortest duration was observed in the simple onset condition (M = 57.6 ms, SD = 29), followed by the heterosyllabic condition (M = 64.1 ms, SD = 41.3). The kinematic duration of /f/ in the onset cluster condition had the longest duration (M = 73.1 ms, SD = 43). The statistical model revealed that the kinematic duration of /f/ did not differ significantly between the simple onset condition and the heterosyllabic condition (β = −5.06, SE = 3.67, p = .17). The kinematic duration of /f/ was significantly shorter than that in the onset cluster condition (β = −13.60, SE = 3.64, p = .0006). The /f/ duration in the heterosyllabic condition was significantly shorter than that in the onset cluster condition (β = −8.54, SE = 3.61, p = .04). Collectively, these results confirm that our participants reliably distinguished between these two sequence types in their production despite the similarity in the observed C_2V overlap ratios.

Discussion

In the present study, speakers of American English were orthographically cued to produce nonnative onset /fn/ clusters and phonetically similar native onset /fl/ and /sm/ clusters while their articulatory movements were recorded and tracked using EMA. The nonnative /fn/ cluster was chosen because past studies based on acoustics have consistently shown that speakers produce this nonnative cluster with high accuracy (relative to other nonnative clusters; Buchwald & Cheng, 2023; Davidson, 2006, 2010). Consistent with those prior investigations, we here found that participants' production accuracy for nonnative /fn/ clusters was well over 80% based on our perceptual and acoustic analysis procedures. This confirmed that our participants were able to produce this cluster with high levels of accuracy, even in the absence of extensive motor practice and while wearing EMA sensors on their articulators. Moreover, our findings reveal that participants produced the nonnative /fn/ cluster relatively well when presented with only orthographic models, in contrast to both auditory and orthographic models as used in some previous studies (Davidson, 2010; Pouplier et al., 2020; Wilson et al., 2014).

Next, we proceeded to investigate the extent to which speakers apply the temporal organization associated with native onset clusters to nonnative clusters. Prior studies have observed C-center timing for native onset clusters in American English (Browman & Goldstein, 1988; Byrd, 1995; Marin & Pouplier, 2010). That is, regardless of onset complexity, the temporal interval between the center of the onset cluster as a whole and the vowel remains stable. This complex temporal organization gives rise to the empirical observations that the leftmost consonant in the cluster (C_1) decreases in temporal overlap with the vocalic gesture, whereas the rightmost consonant (C_2) increases in temporal overlap with the vowel, as compared to their singleton counterparts (Pastätter & Pouplier, 2017). We examined how the nonnative /fn/ cluster is temporally organized with respect to the native cluster counterparts by quantifying the degree of leftward shift (i.e., C_1V overlap ratio) and rightward shift (i.e., C_2V overlap ratio). If speakers are readily able to transfer the C-center timing organization associated with the native clusters to nonnative clusters, we would expect to observe a C_1V overlap ratio that is larger than 1 and a C_2V overlap ratio that is smaller than 1. If, however, speakers apply a simplex timing organization (Shaw et al., 2011) to /fn/ clusters, in which speakers append the leftmost consonant to a simple onset, then we would expect to observe the C_1V overlap ratio to be larger than 1 and the C_2V overlap ratio to be equal to 1, reflecting no increase in rightward shift as compared to the singleton context.

The results on C_1V overlap ratios showed that the leftmost consonant (C_1) in the cluster condition reliably decreased in temporal overlap with the vocalic gesture as compared to its singleton counterpart, as indicated by the fact that the C_1V overlap ratios for each cluster were all greater than 1 (shown in Figure 4). Although there were numerical differences in the extent of leftward shift across the three clusters, these differences did not reach statistical significance. The results on C_2V overlap ratios showed that all three clusters exhibited a decrease in the temporal lag between the C_2 consonant and the anchor point in the cluster condition compared to the singleton condition when pre-C_2 J position was not controlled (see Figure 5, left panel). This indicates that when the onset became more complex, there was an increase in temporal overlap between the C_2 consonant and the vocalic gesture. This rightward shift further provides evidence that the onset clusters followed a C-center timing pattern, regardless of whether they were native or nonnative. Here, there was a significant difference in the magnitude of C_2V overlap ratios among the clusters (i.e., the ratios for /fl/ and /fn/ were reliably less than /sm/). These results are compatible with the notion that the gestures associated with native onset clusters in American English are temporally coordinated in-phase with the vocalic gesture and antiphase (in a competitive coupling relationship) with each other, resulting in C-center timing (Browman & Goldstein, 1988; Marin & Pouplier, 2010; Nam et al., 2009). Furthermore, this temporal organization may generalize to phonetically similar nonnative /fn/ clusters, lending support to the position advanced by Davidson (2006, 2010).

Crucially, however, a different pattern was observed when the C_2V overlap ratios were calculated using the singleton condition that matched for the pre-C_2 consonantal environment in the cluster condition (e.g., I say #fnægbit again vs. I puff #nægbit again). In this comparison (see Figure 5, right panel), only the native /fl/ cluster exhibited evidence of C-center timing in which there was a relatively shorter C_2V temporal overlap in the cluster condition compared to the singleton condition. In contrast, for both nonnative /fn/ and native /sm/ cluster, the magnitude of C_2V temporal overlap in the cluster condition was comparable to that in the singleton condition. Instead, they patterned more with the simplex timing pattern (Shaw et al., 2011). Post hoc analyses of the kinematic duration of the nasal and fricative segments (see Figure 6) suggest that participants reliably distinguished between fricative-nasal sequences that fell within a given syllable from those that spanned two different syllables in their productions despite the similarity in the C_2V overlap ratios. Thus, the difference in the overall pattern between the C_2V overlap ratios computed when the pre-C_2 J position was controlled versus uncontrolled is unlikely because participants did not

produce the intended syllable structure. Other possible accounts of this difference are discussed below.

Putting aside this issue, the kinematic findings of the present study are somewhat inconsistent with prior acoustic investigations. Specifically, Buchwald and Cheng (2023) reported that, for nonnative fricative-nasal clusters that were produced without acoustic evidence of errors, the acoustic duration of the nasal segments in the clusters decreased following a session of structured motor practice. This suggests that speakers might not have achieved this fine-grained temporal coordination that is natively cluster-like in the initial stage of production. However, here, we observed a comparable temporal coordination pattern in the nonnative /fn/ cluster and in the native clusters. In considering this apparent difference, there are a few points that we wish to highlight. First, the acoustic measure of nasal duration in Buchwald and Cheng (2023) largely corresponds to the C_2V overlap ratio in the current study. We indeed found that there was a decrease in C_2V overlap ratio in the cluster condition compared to the singleton condition. While this patterns with the characteristics of cluster timing (e.g., shorter nasal duration in cluster than in singleton; Klatt, 1976; Marin & Pouplier, 2010), we cannot make a definitive claim on whether this finding reflects the end state of temporal coordination for the /fn/ cluster produced by native speakers of American English, as there was no training component in the current experiment. Second, there is a difference in the elicitation modality between the current study and previous motor sequence learning studies. Whereas in the current study, participants were asked to read the target nonwords aloud, participants were presented with both auditory and orthographic models in previous production learning studies. More importantly, the auditory models were produced by a model speaker who is a simultaneous bilingual of American English and another language that contains the clusters not found in American English. Therefore, the auditory models used in the production learning studies could contain subphonemic cues for the nonnative clusters that drive the changes in duration; previous studies have shown that speakers adapt their temporal coordination of onset clusters to be aligned with the cluster targets presented in the auditory models (Pouplier et al., 2020).

Although our results are consistent with the idea that the temporal coordination of articulator movements may be transferred from native clusters to phonetically similar nonnative clusters, they provide no information about whether and how articulatory strategies may change following extended speech motor practice. In the current experiment, we only recorded eight productions of each nonword, which may vary in different ways. Therefore, splitting the data set to compare the initial and final productions would not yield representative differences. Future, larger scale studies are needed to collect appropriate data sets (with more repetitions per word) with which to address the articulatory correlates of cluster learning.

Limitations of the Present Study

The difference in the overall pattern between the C_2V overlap ratios computed when the pre-C_2 J position was controlled versus uncontrolled is somewhat inconclusive. There are several potential confounds that may have contributed to this particular finding. Consider first how the nature of the speaking materials themselves may have contributed to this finding. To match the pre-C_2 consonant environment, we used the sentences, "I bus _____ again" and "I puff ____ again" for the singletons. These sentences contain a relatively low-frequency verb, whereas the carrier phrases that were used to elicit the cluster condition contained relatively high-frequency verbs (e.g., have, gotta). This difference in carrier phrase lexical item frequency may have affected the temporal coordination of the target nonwords. However, if this were the case, then it would be difficult to account for the pattern exhibited by the /fl/ cluster given that the same stimulus design was used. In addition, previous studies have demonstrated that prominence location and prosodic boundaries can affect temporal lags (Sotiropoulou & Gafos, 2022). The use of different carrier phrases in the present study might introduce variation in prosody. However, we think that the prosodic influence on temporal lags is likely to be minimized in our data. First, all target nonwords were presented in capital letters to the participants, which ensured that the placement of focus stress was on the targets. Indeed, only three tokens were produced with the iambic stress pattern based on the perceptual analysis and were excluded from the kinematic analyses. Second, all target nonwords were embedded at the phrasal medial position in the carrier phrases, ensuring that there were no differences in prosodic boundaries across different carrier sentences.

Next, consider how the place and manner of articulation of the rightmost consonant may have modulated C_2V overlap ratios. In a prior study with speakers of Polish, Pastätter and Pouplier (2017) found that the extent of C_2V overlap ratios depended on the degree of "coarticulatory resistance," in which less temporal overlap was observed when the onset consonants and vowel involved placed competing demands on the same articulator (i.e., the tongue; Recasens & Espinosa, 2009). Specifically, results revealed that the onset clusters that contained a labial C_2 consonant (e.g., /sp/, /ʃm/) had the smallest C_2V overlap ratio (i.e., more rightward shift), whereas the clusters that contained an alveolar sibilant (/ps/, /pʃ/) had the largest ratio (i.e., less rightward shift). Critically, however, the present findings cannot be accounted for by the same

principle of coarticulatory resistance found in Polish because the /sm/ cluster, as the only labial C_2 in the current study exhibited the least amount of temporal overlap with the vocalic gesture. As Pastätter and Pouplier (2017) speculated, however, coarticulatory resistance may be a language-specific factor, and our findings are consistent with this idea. Our findings suggest that, in American English, both place (e.g., labial) and manner (e.g., nasal) of articulation affect the temporal overlap between consonants and vowels. This also aligns with the results of Marin and Pouplier (2010) in which the /sm/ cluster was produced with the smallest amount of rightward shift as compared to both the /sp/ and /sk/ clusters in American English. Although, it is important to note that we are not able to fully answer whether the source of this restriction on consonant–vowel temporal overlap is derived from the /m/ segment having a labial gesture, or a velum lowering gesture, or a combination of the two. Follow-up studies could address this question by including the onset /sp/ cluster to provide evidence of the exact hierarchy of coarticulatory resistance in American English.

However, another possible explanation for a greater amount of C_2V overlap in /fl/ and /fn/ compared to /sm/ is that both the /fl/ and /fn/ clusters have a front-to-back articulatory transition, whereas the /sm/ cluster has a back-to-front transition. Back-to-front sequences have been argued to allow less overlap because the acoustic burst signaling the release of the first, posterior constriction will be obliterated if the second (more anterior) constriction is already in place. This perceptual account aligns with the results of Chitoran et al. (2002), which revealed that Georgian front-to-back stop–stop clusters (e.g., /bg/) were produced with a higher degree of C_1C_2 temporal overlap than back-to-front clusters (e.g., /gb/). The current findings provide additional evidence that front-to-back articulatory transitions also may lead to a higher amount of CV temporal overlap. Collectively, our results highlight the need for detailed articulatory descriptions of a broader range of native and nonnative (e.g., /zm/, /vm/, /gd/, /bd/) clusters elicited in different phonetic contexts.

Conclusions

In summary, articulatory coordination patterns for native clusters appear to transfer to nonnative clusters composed of phonetically similar movements, although the segmental composition of the cluster and type of carrier phrase used can modulate the degree of transfer, raising new questions for future research. If evidence of transfer is found with a broader range of cluster types, then that would inform phonetic theories that seek to explain why speakers produce some nonnative clusters with better accuracy than others (Davidson, 2006, 2010). Such findings could also potentially be used to inform articulatory strategies utilized in L2 pedagogy.

Ethical Approval

All experiments complied with the principles of research involving human subjects as stipulated by the University of Florida.

Informed Consent

Informed consent was obtained from all individual participants included in the study.

Data Availability Statement

The data that support the findings of this study are openly available in the Open Science Framework online repository: https://osf.io/dxeka/.

Acknowledgments

The research reported in this article was supported by a New Investigator Research Grant from the American Speech-Language-Hearing Foundation (PI: M. Masapollo) and R01 DC-018589 (PI: A. Buchwald). The second author (M. Masapollo) was also supported by R01 DC-017439 (PI: D.J. Ostry) during the preparation of this article. Thanks to Jessica Goel, who assisted with data collection; Jessica Smith and Grant Oberle, who assisted with coding of the kinematic data; and Maria Cano Cuervo and Khamiya Nisbett, who assisted with coding of the acoustic data. We would also like to thank Mark Tiede for developing and sharing his MView toolbox, which facilitated the articulatory analyses reported here. Finally, this work benefited from helpful discussions with, or comments from, Jason Shaw, Dani Byrd, Lisa Goffman, and other audience members of the 12th International Seminar on Speech Production (held in Providence, RI) and the 181st Meeting of the Acoustical Society of America (held in Seattle, WA).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

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

Berent, I, Steriade, D, Lennertz, T, & Vaknin, V (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition, 104*(3), 591–630. https://doi.org/10.1016/j.cognition.2006.05.015

Boersma, P., & Weenink, D. (2023). Praat: Doing phonetics by computer (Version 6.3.17) [Computer software]. http://www.praat.org/

Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica, 45*(2–4), 140–155. https://doi.org/10.1159/000261823

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

Brunner, J., Geng, C., Sotiropoulou, S., & Gafos, A. (2014). Timing of German onset and word boundary clusters. *Laboratory Phonology, 5*(4), 403–454. https://doi.org/10.1515/lp-2014-0014

Buchwald, A., Calhoun, H., Rimikis, S., Lowe, M. S., Wellner, R., & Edwards, D. J. (2019). Using tDCS to facilitate motor learning in speech production: The role of timing. *Cortex, 111,* 274–285. https://doi.org/10.1016/j.cortex.2018.11.014

Buchwald, A., & Cheng, H.-S. (2023). Factors affecting nonnative consonant cluster learning. *Journal of Speech, Language, and Hearing Research, 66*(8S), 3038–3051. https://doi.org/10.1044/2022_JSLHR-22-00322

Byrd, D. (1995). C-centers revisited. *Phonetica, 52*(4), 285–306. https://doi.org/10.1159/000262183

Chitoran, I., Goldstein, L., & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (7th ed., pp. 419–448). Mouton de Gruyter.

Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics, 34*(1), 104–137. https://doi.org/10.1016/j.wocn.2005.03.004

Davidson, L. (2010). Phonetic bases of similarities in cross-language production: Evidence from English and Catalan. *Journal of Phonetics, 38*(2), 272–288. https://doi.org/10.1016/j.wocn.2010.01.001

Davidson, L. (2011). Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 270–282. https://doi.org/10.1037/a0020988

Davidson, L., & Shaw, J. A. (2012). Sources of illusion in consonant cluster perception. *Journal of Phonetics, 40*(2), 234–248. https://doi.org/10.1016/j.wocn.2011.11.005

Dromey, C., Hunter, E., & Nissen, S. L. (2018). Speech adaptation to kinematic recording sensors: Perceptual and acoustic findings. *Journal of Speech, Language, and Hearing Research, 61*(3), 593–603. https://doi.org/10.1044/2017_JSLHR-S-17-0169

Dupoux, E, Parlato, E, Frota, S, Hirose, Y, & Peperkamp, S (2011). Where do illusory vowels come from? *Journal of Memory and Language, 64*(3), 199–210. https://doi.org/10.1016/j.jml.2010.12.004

Gafos, A. I., Charlow, S., Shaw, J. A., & Hoole, P. (2014). Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics, 44,* 152–166. https://doi.org/10.1016/j.wocn.2013.11.007

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system*. Cambridge University Press. https://doi.org/10.1017/CBO9780511541599.008

Harel, D., & McAllister, T. (2019). Multilevel models for communication sciences and disorders. *Journal of Speech, Language, and Hearing Research, 62*(4), 783–801. https://doi.org/10.1044/2018_JSLHR-S-18-0075

Hermes, A., Mücke, D., & Auris, B. (2017). The variability of syllable patterns in Tashlhiyt Berber and Polish. *Journal of Phonetics, 64,* 127–144. https://doi.org/10.1016/j.wocn.2017.05.004

Hermes, A., Mücke, D., & Grice, M. (2013). Gestural coordination of Italian word-initial clusters: The case of 'impure s.' *Phonology, 30*(1), 1–25. https://doi.org/10.1017/S095267571300002X

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Jacobi, J. (2022). *Coordination and timing of speech gestures in Parkinson's disease* [Doctoral dissertation, University of Groningen]. https://doi.org/10.33612/diss.238268145

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

Lenth, R. (2022). emmeans: Estimated marginal means, aka least-squares means (R package Version 1.7.2.) [Computer software]. https://CRAN.R-project.org/package=emmeans

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Liu, Z., Xu, Y., & Hsieh, F. (2022). Coarticulation as synchronised CV co-onset—Parallel evidence from articulation and acoustics. *Journal of Phonetics, 90,* Article 101116. https://doi.org/10.1016/j.wocn.2021.101116

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software, 6*(60), Article 3139. https://doi.org/10.21105/joss.03139

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Marin, S. (2013). Organization of complex onsets and codas in Romanian: A gestural approach. *Journal of Phonetics, 41*(3–4), 211–227. https://doi.org/10.1016/j.wocn.2013.02.001

Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control, 14*(3), 380–407. https://doi.org/10.1123/mcj.14.3.380

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupé (Eds.), *Approaches to phonological complexity* (pp. 297–328). Walter de Gruyter. https://doi.org/10.1515/9783110223958.297

Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator of syllable structure. *Proceedings of the XVth International Congress of Phonetic Sciences, 3,* 2253–2256.

Pastätter, M., & Pouplier, M. (2017). Articulatory mechanisms underlying onset-vowel organization. *Journal of Phonetics, 65,* 1–14. https://doi.org/10.1016/j.wocn.2017.03.005

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Perkell, J. S., Cohen, M. H., Svirksy, M. A., Matthies, M. L., Garabieta, I., & Jackson, M. T. T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America, 92*(6), 3078–3096. https://doi.org/10.1121/1.404204

Pouplier, M., & Beňuš, Š. (2011). On the phonetic status of syllabic consonants: Evidence from Slovak. *Laboratory Phonology, 2*(2), 243–273. https://doi.org/10.1515/labphon.2011.009

Pouplier, M., Lentz, T. O., Chitoran, I., & Hoole, P. (2020). The imitation of coarticulatory timing patterns in consonant clusters for phonotactically familiar and unfamiliar sequences. *Laboratory Phonology, 11*(1), Article 1. https://doi.org/10.5334/labphon.195

Pouplier, M., Marin, S., Hoole, P., & Kochetov, A. (2017). Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics, 64,* 108–126. https://doi.org/10.1016/j.wocn.2017.01.006

R Core Team. (2017). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing.

Rebernik, T., Jacobi, J., Jonkers, R., Noiray, A., & Wieling, M. (2021). A review of data collection practices using electromagnetic articulography. *Laboratory Phonology, 12*(1), Article 6. https://doi.org/10.5334/labphon.237

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

RStudio Team. (2023). R Studio: Integrated development for R [Computer software]. RStudio. PBC.

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

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Shaw, J., & Chen, W. R. (2019). Spatially conditioned speech timing: Evidence and implications. *Frontiers in Psychology, 10,* Article 2726. https://doi.org/10.3389/fpsyg.2019.02726

Shaw, J., & Davidson, L. (2011). Perceptual similarity in input–output mappings: A computational/experimental study of non-native speech production. *Lingua, 121*(8), 1344–1358. https://doi.org/10.1016/j.lingua.2011.03.003

Shaw, J., & Gafos, A. I. (2010). Quantitative evaluation of competing syllable parses. In *Proceedings of the 11th Meeting of the ACL Special Interest Group on Computational Morphology and Phonology* (pp. 54–62). Association for Computational Linguistics.

Shaw, J., & Gafos, A. I. (2015). Stochastic time models of syllable structure. *PLOS ONE, 10*(5), Article e0124714. https://doi.org/10.1371/journal.pone.0124714

Shaw, J., Gafos, A. I., Hoole, P., & Zeroual, C. (2011). Dynamic invariance in the phonetic expression of syllable structure: A case study of Moroccan Arabic consonant clusters. *Phonology, 28*(3), 455–490. https://doi.org/10.1017/S0952675711000224

Sotiropoulou, S., & Gafos, A. (2022). Phonetic indices of syllabic organization in German stop-lateral clusters. *Laboratory Phonology, 13*(1). https://doi.org/10.16995/labphon.6440

The MathWorks Inc. (2019). MATLAB (Version: 9.13.0) (R2022b). https://www.mathworks.com

Tiede, M. (2005). MVIEW: Software for visualization and analysis of concurrently recorded movement data [Computer software]. Haskins Laboratories.

Tiede, M. (2010). MVIEW: Multi-channel visualization application for displaying dynamic sensor movement. Haskins Laboratories.

Ussishkin, A., & Wedel, A. (2003). Gestural motor programs and the nature of phonotactic restrictions: Evidence from loanword phonology. In G. Grading & M. Tsujimira (Eds.), *Proceedings of WCCFL, 2003* (Vol. 22, pp. 505–518). Cascadilla Press.

Weismer, G., & Bunton, K. (1999). Influences of pellet markers on speech production behavior: Acoustical and perceptual measures. *The Journal of the Acoustical Society of America, 105*(5), 2882–2894. https://doi.org/10.1121/1.426902

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), Article 1686. https://doi.org/10.21105/joss.01686

Wilson, C., Davidson, L., & Martin, S. (2014). Effects of acoustic–phonetic detail on cross-language speech production. *Journal of Memory and Language, 77,* 1–24. https://doi.org/10.1016/j.jml.2014.08.001