


Rapid movements at segment boundaries^{a)}

Malin Svensson Lundmark^{b)} 

Centre for Languages and Literature, Lund University, Lund, Sweden

ABSTRACT:

This paper reports on a one-to-one aspect of the articulatory-acoustic relationship, explaining how acoustic segment boundaries are a result of the rapid movements of the active articulators. In the acceleration profile, these are identified as *acceleration peaks*, which can be measured. To test the relationship, consonant and vowel segment durations are compared to articulatory *posture intervals* based on acceleration peaks, and time lags are measured on the alignment of the segment boundaries to the acceleration peaks. Strong relationships and short time lags are expected when the acceleration peaks belong to *crucial* articulators, whereas weak relationships are expected when the acceleration peaks belong to *non-crucial* articulators. The results show that lip posture intervals are indeed strongly correlated with [m], and tongue tip postures are strongly correlated with [n]. This is confirmed by the time lag results, which also reveal that the acoustic boundaries precede the acceleration peaks. Exceptions to the predictions are attributed to the speech material or the joint jaw-lip control unit. Moreover, the vowel segments are strongly correlated with the consonantal articulators while less correlated with the tongue body, suggesting that acceleration of crucial consonantal articulators determines not only consonant segment duration but also vowel segment duration.

© 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1121/10.0017362>

(Received 15 June 2022; revised 17 January 2023; accepted 8 February 2023; published online 1 March 2023)

[Editor: Jianjing Kuang]

Pages: 1452–1467

I. INTRODUCTION

The relationship between articulation and acoustics is central to the development of articulatory phonetics. This paper addresses a specific aspect of the articulatory-acoustic relationship, a one-to-one connection between movement characteristics of an articulator, i.e., how it reaches its steady-states, and the resulting acoustic outcome, the acoustic segment boundaries.

A. Segmental articulations

Phonemes or speech sound units are commonly known as being represented in speech by segments. The basic function of speech segments is to make the sounds differentiate from each other and create distinctiveness (Jakobson *et al.*, 1969; Ohala, 1992). Segments can be likened to articulatory *steady-states*, which is when the movement does not change over time. For that reason, segments have also been referred to as *dead intervals* as the steady-state serves no new information (Ohala, 1992).

Segments are temporally coordinated articulatory movements and they are delimited by segment transitions. The segment transitions are used by the listener to differentiate

between sounds as they hold most of the information of the coordinated articulatory movements (Stevens and Blumstein, 1978; Ohala, 1992). Indeed, studies on vowel perception (the so-called *silent-center* paradigm, i.e., Strange, 1987; Jenkins *et al.*, 1999) have shown that listeners use information of the transitions in vowel-identification tasks.

The segment transitions consist of rapid articulatory movements that result in the large acoustic changes that we refer to as segment boundaries (Fant and Lindblom, 1961; Gårding, 1967; Jakobson *et al.*, 1969; Zsiga, 1994). Rapid articulatory movement changes at the segment transitions are local and instant, and the consequence of those movements are the segment boundaries. It follows that this causal relationship can be described such that

Rapid articulatory movements → acoustic changes

→ acoustic segment boundaries.

1. Peak acceleration

Rapid articulatory movements are large acceleration and deceleration changes, and they occur when an active articulator changes position. For example, in a bilabial stop, the lips accelerate when opening the lips and decelerate just before closing the lips. Likewise, for an alveolar consonant, tongue tip (TT) speed changes can be observed just before and after contact with the palate (Svensson Lundmark, 2020). The large acceleration and deceleration changes occur because, for the articulator to move from one steady-state to the next, it needs to accelerate, that is, change its

^{a)}Portions of this work were presented in “Evidence of segmental articulations: Acceleration determines vowel segment duration in Swedish Word Accents,” Proceedings of TAI, Sønderborg, Denmark, December, 2022; in “Rapid movements at segment boundaries—Preliminary reports on manner,” Proceedings of FONETIK 2022, Stockholm, Sweden, June, 2022; and in “Peak acceleration determines segment boundary” at Speech Motor Control, Groningen, The Netherlands, August, 2022.

^{b)}Also at: Centre for Industrial Electronics, University of Southern Denmark, Odense, Denmark. Electronic mail: malin.svensson_lundmark@ling.lu.se

velocity. This is in accordance with dynamical systems theory (DST), which are measurable functions that are bound by relationships of different natural phenomena (for an overview, see Iskarous, 2016). An *acceleration peak* denotes the moment when acceleration is at its highest, which is when velocity changes the most. Thus, we find an acceleration peak between a steady-state and *peak velocity* (Fig. 1). Peak velocity is the moment an articulator travels the fastest, which is when it keeps its steady course toward the target while position is changed. Thus, as a function of time, velocity is the change in position, whereas acceleration is the change in velocity. In other words, acceleration is the second derivative to position, and it occurs as a result of added force (Eager et al., 2016). The amount of force added determines the speed of the articulator, and depending on the type of force added, the result of the change in velocity may be peak acceleration or *peak deceleration* (Eager et al., 2016). In other words, one type of force is used when the articulator moves away from the constriction, at which peak acceleration occurs. Another type of force causes the articulator to slow down before a constriction, and that is when we, instead, find peak deceleration (Fig. 1).

An acceleration peak entails a change in acceleration, a *jerk*, which means that not only will the position change but the direction will also change (Eager et al., 2016). Hence, an acceleration peak denotes a time when an articulator jerks, i.e., moves rapidly, as velocity, position, and direction change at the same time (Eager et al., 2016). As this results in rapid movement changes of the articulator, there is a direct connection between acceleration and the segment boundary (as proposed in Svensson Lundmark, 2020,

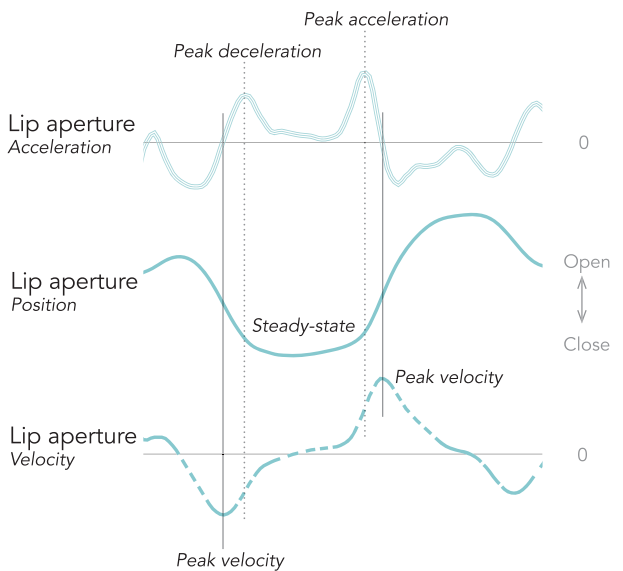


FIG. 1. (Color online) Lip aperture (LA; the distance between upper and lower lip, middle curvature), its velocity profile (bottom), and its acceleration profile (top). Notice that the top of the middle curve represents when the lips are open, and bottom represents when the lips are closed. The vertical dotted lines show *peak deceleration* and *peak acceleration*, respectively, which occur between *peak velocity* (the solid lines) and a *steady-state* (here, it is a bilabial constriction).

following Fant and Lindblom, 1961; Gårding, 1967; Jakobson et al., 1969; Zsiga, 1994). Hence, we can add an acceleration peak to the causal relationship of an articulator and its acoustic outcome and create a hypothesis to be tested in this study,

- Acceleration peak → rapid articulatory movements
- acoustic changes
- acoustic segment boundaries.

2. Articulatory posture intervals of consonants

The steady-state of the articulator is found between peak deceleration and peak acceleration, where the active articulator stays at its target position and has no active movement in any direction. This forms an articulatory interval which is hereafter referred to as a *posture interval* (Svensson Lundmark, 2020). A posture interval is essentially articulatory and only related to one crucial active articulator. Moreover, it is defined as delimited by the two articulatory landmarks: peak deceleration and peak acceleration (Fig. 2).

Since peak deceleration and peak acceleration correlate with the rapid movements at each end of a constriction, a posture interval would, in theory, be equal to the duration of the constriction, i.e., the segment (Fig. 2). As such, a posture interval measured on the lips would correlate with a bilabial constriction. Likewise, a posture interval measured on the movements of the TT would correlate with an alveolar

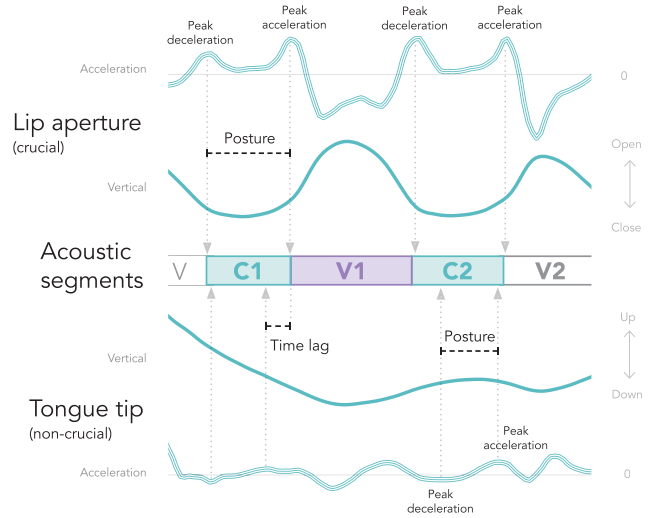


FIG. 2. (Color online) The acceleration profiles and vertical positions of LA (top curvatures) and TT (bottom curvatures) during a CVCV sequence (acoustic segments, middle), where the lips are the crucial articulator. Notice that the bottom of the LA position curve represents when the lips are closed. The two measurements of the present study are included: (1) *posture interval*, which is framed by the peak deceleration and the peak acceleration, and when measured on the crucial articulator is hypothesized to correlate with the segment duration (C1 and C2); and (2) *time lag*, which measures timing of the acoustic segment boundary in relation to the acceleration peak (a positive time lag indicates the segment boundary follows the acceleration peak).

constriction. In fact, we see, yet, a possible causal relationship, this time between a posture interval of any given active articulator and the resulting acoustic segment:

Lip posture interval → a bilabial constriction.

Tongue tip posture interval → an alveolar constriction.

Although constrictions, in general, involve several active articulators (the velum, the jaw, the laryngeal articulators, etc.), the proposed causal relationship above presupposes only the action of the crucial articulator. The posture interval, thus, refers more specifically to the place of articulation. As for the manner of articulation, the pattern is less straightforward, and the type of constriction may affect how well the segment boundary and an acceleration peak align (Svensson Lundmark, 2022a).

This paper proposes that there is an obvious link between articulation and acoustics in the segmental division of consonants and vowels, which is that the acceleration peaks of the crucial articulators generate segmental articulations. These segmental articulations either consist of active movements, so-called *active intervals*, when the articulators move toward or away from a position, or they consist of fairly stationary posture intervals, a steady-state, when the active articulator stays at its target position, e.g., a constriction (Svensson Lundmark, 2020). Between the steady-states of the posture intervals, when there is no constriction and the active articulators move somewhere, there are either only silent pauses, or what we know as vowels, emerge.

3. Vowels

As proposed in Öhman (1966), consonantal articulation is of an instantaneous nature, layered on top of diphthongal vowels. This has been further developed into various models, including the C/D model by Fujimura (2000), and articulatory phonology (AP) by Browman and Goldstein (1988, 1989), following work by Fowler (1986, 1996). The idea behind this notion of overlapping gestures is that the consonantal gestures are separate in shape and structure from the vowels. The vowels consist of slow continuous movements, whereas the consonants are fast and as a result end up as shorter than the vowels. Moreover, different features are assumed not only between consonants and vowels but also between consonants in syllable onset and coda position (Fujimura, 2000; Browman and Goldstein, 1988). Although AP and the C/D model assume different rules to govern the consonants in onset and coda, in AP, the difference lies in how the articulatory gestures are timed with one another. Simplified, in onset, the consonantal and vocalic gesture are synchronized, whereas, in coda, the gestures are sequentially timed with one another (Browman and Goldstein, 1988, 2000). Cross-linguistic work suggests that the timing of the gestures, the coarticulation, is not only syllabic specific but also language specific, as well as feature specific (Fowler and Saltzman, 1993; Byrd, 1996; Pouplier, 2012; Bombien

et al., 2013; Marin, 2013; Svensson Lundmark *et al.*, 2021). The C/D model, instead, assumes that the consonantal movements in onset vs coda behave differently simply because they are either at the beginning or end of the *syllable pulse* (Fujimura, 2000). The syllable pulse is related to prominence, and it determines the syllable magnitude, which, in turn, is directly linked to jaw displacement (Fujimura, 2000; Erickson *et al.*, 2014). Thus, according to the C/D model, the amount of jaw opening would, in theory, be related to the strength of the syllable magnitude and, in turn, to the level of prominence such that the more jaw displacement there is, the more prominence there is.

Following the reasoning of the C/D model, a consonant-vowel-consonant (CVC) syllable is the result of the jaw displacement, and syllable onset takes place during the jaw opening while coda takes place when the jaw closes. As orofacial movements are all closely linked and connected, how constrictions occur depends heavily on the degree of jaw displacement (Gracco, 1988; Lindblom, 1983; Mooshammer *et al.*, 2007; Bose and van Lieshout, 2012; Kawahara *et al.*, 2014). Thus, it is reasonable to assume very different types of movements of the active articulator in concordance with the jaw opening (syllable onset) as opposed to the jaw closing (syllable coda): starting positions of active articulators would vary between an open and closed mouth and distance to target position would vary as well. Between a movement in onset vs in coda, there would possibly also be different trajectories and speeds of the articulator. This calls for different articulatory strategies between consonants in onset vs coda, which, in turn, could explain the different articulatory timing patterns found in the literature. However, irrespective of the underlying articulatory strategies, the movements of the consonants that occur in onset and coda are on either side of the vowel. Therefore, the basic idea remains that the consonants are superimposed on the vocalic diphthongal gestural movement (Öhman, 1966). Hence, the resulting acoustic vowel segment is limited by when either of its neighboring consonantal constrictions are made no matter the type of articulatory strategy.

a. Distance between two posture intervals. After a constriction, as the active articulator accelerates and leaves its position, peak acceleration marks the end of the consonantal acoustic segment (Fig. 2). This is the proposed acoustic-articulatory one-to-one relationship of the present paper. As the consonant segment in a consonant-vowel (CV) sequence ends, the acoustic vowel segment appears. Thus, the rapid movements of the active articulator, peak acceleration, would not only denote the end of the consonant segment but also the start of the vowel segment. Likewise, in a vowel-consonant (VC) sequence, when the acoustic vowel segment ends, peak deceleration of the active articulator before a constriction denotes the end of the acoustic vowel segment (Fig. 2). Hence, in a consonant-vowel-consonant sequence (C₁V₁C₂), as exemplified in Fig. 2, the acoustic vowel segment is really a result of the timing of the two nonlocal but still neighboring articulatory posture intervals.

An acoustic vowel segment duration is, therefore, proposed to be specified by the timing of consonantal acceleration or, more specifically, by the timing of peak acceleration at the end of the word-initial consonant (C1) and the timing of peak deceleration at the start of the word-medial consonant (C2; Fig. 2). Ongoing work suggests that this pattern is present no matter what the vowel quality or vowel quantity is (Svensson Lundmark, 2022a).

B. Hypothesis and predicted result

The present paper tests the hypothesis that the acceleration peaks of the rapid articulatory movements at segment transitions define the acoustic segment boundaries. As a result, acoustic segment duration of consonants and vowels should be equal to the articulatory posture interval between two acceleration peaks. As a first step to test the hypothesized causal relationship, this study uses a twofold approach to investigate the correlation between the two variables. That is, this study (1) examines the strength of the relationship between the acoustic segment duration and the articulatory posture intervals, and (2) examines timing by calculating time lags between an acoustic segment boundary and an acceleration peak.

The proposed causal relationship is tested on consonants and vowels, although with slightly different approaches. For the consonants, acceleration peaks are collected on the active TT and lip articulators. In this study, articulatory measurements are made on movements of *crucial* and *non-crucial* active articulators. Active articulators are voluntary articulatory movements, and they include, e.g., the lower lip, tongue body (TB), and TT. Crucial articulator refers to an active articulator where the onset and coda of a syllable is articulated (Fujimura, 2000; Erickson et al., 2014). For example, between [n] and [m], where velum is active in both, the crucial articulators differ; the TT is crucial for [n], whereas the lower lip is crucial for [m]. The expectation is to find a strong relationship between the acoustic segment and articulatory posture interval *only* when the posture interval involves a *crucial* active articulator between, e.g., [m] and a posture interval measured on the lips. Furthermore, the relationship is expected to be weak when the posture interval involves a *non-crucial* active articulator between a bilabial [m] and a TT posture interval, and so forth (Table I). In conjunction with this, timing of the segment boundaries with peak deceleration and peak acceleration, respectively, is evaluated by calculating time lags, where the expected result is a better alignment to crucial articulators than to non-crucial articulators (Table I).

The correlation for the vowels is also assumed to be strong between a crucial consonantal articulator and an acoustic vowel segment. What complicates the prediction somewhat in terms of the vowels is that we need to take two different articulatory posture intervals into account. For example, in the target word onset /man/, we expect to find a very strong relationship with the calculated interval between a lip and a TT posture (lip - TT), but not with the TT - lip

TABLE I. The expected results (1) on the strength of the relationship between a segment and posture interval and (2) on the time lags between the segment boundary and the acceleration peaks. The expected outcome is heavily based on whether the articulator is crucial or not.

	Lip posture	TT posture
Bilabial consonant	(1) Strong positive relation (2) Short not varied time lags	(1) Weak positive relation (2) Long varied time lags
Alveolar consonant	(1) Weak positive relation (2) Long varied time lags	(1) Strong positive relation (2) Short not varied time lags

posture (as both articulatory landmarks are measured on non-crucial articulators). However, we may find a semi-strong relationship for the vowel when one of the articulatory landmarks is collected from a crucial consonantal articulator. Furthermore, to evaluate the hypothesis that the acceleration of the consonants determines the acoustic vowel segment duration, posture intervals on the TB have also been calculated. The approach is similar to that of the articulation of the consonants in this study in that peak deceleration and peak acceleration of the place of articulation of the vowel frame the TB posture interval. This approach is based on the idea of vowel constriction location, which is analogous to place of articulation of consonants (Stevens and House, 1955; Wood, 1979). The expected result is no or a weak correlation between a TB posture and the acoustic vowel segment. Thus, the stated hypothesis of this study is predicted to be falsified for the TB acceleration peaks. Timing of the vowel segment boundaries with the acceleration peaks of the TB is calculated using time lags, expecting similar results as for the non-crucial articulators of the consonants.

II. METHOD

Acoustic and kinematic data of word onset CVC sequences are included in the present study. This entails acoustic segment duration of consonants and vowels (C1, V1, and C2) and kinematic measurements on articulatory movements, such as lip aperture (LA), TT, and TB acceleration.

A. Speech material

The material was initially recorded for a dissertation project (Svensson Lundmark, 2020). The full corpus includes electromagnetic articulography (EMA) data on 21 speakers, contains approximately 3000 tokens, and consists of 18 disyllabic target words (9 Swedish word accent pairs) placed in a low-prominence context. Low-prominence context means that leading questions ensure that a narrow focus (contrastive) is placed on the last element of the target sentence instead of on the target word.

The aggregated data set used for this study is comprised of four of the target words: /nan:a/, /man:a/, /mam:a/, and /nam:nar/. The four disyllabic target words are embedded in a set of similarly structured target sentences. The vowel-consonant-vowel (VCV) sequence preceding the target words is identical across all of the sentences (/ade/). All of

TABLE II. The target word onsets of the study and number of tokens.

Word onset	Number of tokens
/nan/	139
/man/	143
/mam/	137
/nam/	139

the target words are produced with the Swedish word accent 2 (a tonal rise throughout the stressed first vowel), hence, there are no tonal differences between them.

The CVC word onsets of the four target words consist of an open vowel [a], and either a bilabial [m] or alveolar [n] in the word-initial (C1) and word-medial (C2) position (Table II). C2 is always a mora-sharing geminate, except for in /nam:nar/ in which the mora-sharing geminate, [mː], is also part of a cluster, [mːn] (Sec. IID 1). The aggregated data set amounts to 558 tokens.

B. Speakers

For different reasons, three speakers were excluded from subsequent analysis. One speaker had a deviant dialect; for one speaker, the acoustic quality was not satisfactory (because the position of the microphone was not controlled for); and for one speaker, one of the target words was missing. The aggregated data set used for this study contains data from 18 speakers (12 female). All of the speakers are of the South Swedish dialect of Scania (the southernmost region of Sweden), and the speakers' ages ranged from 23 to 75 years old (\bar{x} = 40 years of age, standard deviation (sd) = 12.8 years). All of the speakers grew up with South Swedish speaking parents. However, three of the speakers grew up with one parent with a first language other than Swedish. The participants were unaware of the purpose of the study, and each participant provided written consent to participate in this study.

C. Procedure

The speakers were recorded with EMA, a Carstens AG501 (Carstens Medizinelektronik GmbH, Bovenden, Germany) at the Lund University Humanities Laboratory. The sentences (leading questions + target sentences) were read from a computer screen in a random order with each set appearing eight times. Articulatory data were recorded at 250 Hz, and audio was recorded simultaneously using an external condenser microphone (a t.bone EM 9600, Thomann GmbH, Burgebrach, Germany) at a sampling rate of 48 kHz.

Articulatory data were collected from six sensors. Two sensors were placed on the midline of the upper and lower lips at the vermilion border, one sensor was placed on the lower incisor, and three sensors were placed on the midline of the tongue. For the tongue sensors, the first sensor, corresponding roughly to the tongue dorsum, was placed on the tongue where the participant made a bite mark after having stretched out his or her tongue as far as possible. The second tongue sensor, corresponding roughly to the tongue blade,

was placed between the sensor at the back and a third tongue sensor, which was placed approximately 1 cm from the TT. To correct for head movements, three additional sensors were used: one behind each ear and one on the nose ridge. The occlusal plane was not controlled for other than by observing and correcting the speaker's position during the recording. Before analysis, the data were corrected for head movements with the Carstens software, using the three reference sensors mentioned above, and then transferred to *R* (R Core Team, 2015). For the purpose of the study, the positions of four of the sensors were used in the subsequent analysis: two sensors on the lips to calculate LA, one sensor on the midline of the TT, and one at tongue dorsum (hereafter referred to as TB).

D. Data analysis

The author segmented the acoustic data manually in Praat (Boersma and Weenink, 2018) using ProsodyPro (Xu, 2013). Segment boundaries were established by examining differences in high-frequency intensity, shape of the waveforms, and formant transitions (Machač and Skarnitzl, 2009). In case of doubt, segmentation was led by perceptual judgment. The text grid files were used to collect consonant and vowel segment duration with ProsodyPro (Xu, 2013). The textgrid files were also later used in *R* as reference time windows for collection of the articulatory data (Sec. IID 2).

An inter-annotation agreement (IAA) was conducted on the segmented acoustic data. The second annotator, a trained phonetician, segmented the target words in 10% of the aggregated data set (60 tokens). The tokens were randomly collected from ten of the speakers. As an indicator of an IAA between the two annotators, an agreement was determined if 90% of the segment boundaries differed by less than 10 ms (Machač and Skarnitzl, 2009). The four segment boundaries of the target word onsets (CVC) were calculated. Results of the IAA showed that 93.4% of the CVC segment boundaries (213 out of 228) were within 10 ms. On average, placement of the segment boundaries differed by 2.4 ms between the two annotators.

1. Heterorganic consonant cluster

One of the target word onsets, /nam/, is part of a heterorganic cluster combination, [mːn] (the target word /nam:nar/). The cluster entails a temporal overlap of the two consonantal gestures as they belong to two different articulators, which proved to be difficult to separate in the acoustic analysis (Fig. 3). As a result, the full heterorganic consonant cluster segment, [mːn], is included in the analysis on word-medial position (C2). Moreover, to make a better judgment of the correlation between the acoustic segment and articulation, the study includes a comparison to a combined posture interval (Sec. IID 3).

2. Articulatory data

The articulatory data were further processed in *R* (R Core Team, 2015). The acceleration landmarks were extracted from

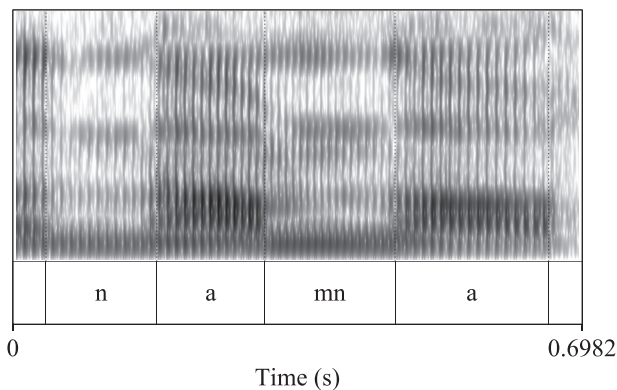


FIG. 3. Example of the target word onset, /nam/, by a male speaker. The overlapping gestures of the heterorganic cluster with the mora-sharing geminate, [m:n], complicates the segmentation.

three-dimensional positions of the lips and two-dimensional positions of the TT and TB. LA was first calculated in *R* using the three-dimensional Euclidian distance between the sensors on the upper and lower lip. LA is used as a measure of the active articulator in this study as, together, the upper and lower lip form an articulator. The acceleration landmarks are, thus, extracted from the calculated LA, whereas for the TT and TB, they are based on two-dimensional positions.

To be able to extract landmarks from second derivative (acceleration), the position data need to be filtered and smoothed. The signal has been filtered with low-pass filter, using the *R* function *loess* with $\text{span} = 0.1$ (Fig. 4). The value was determined by visually inspecting the result. The filtered sequence is the target word, and the fitting is performed locally, which means no time delay. The acceleration signal is simplified by using *loess*, but it should be noted that the signals have already been filtered during the data processing with the Carstens software.

Articulatory landmarks were automatically collected from the acceleration profile of the vertical LA calculation and the two-dimensional TT and TB positions. Ten different peak deceleration and peak acceleration landmarks were collected, according to Table III. During the collection of the landmarks, the Praat textgrid files were used as reference for the time windows (for example, the script was to search

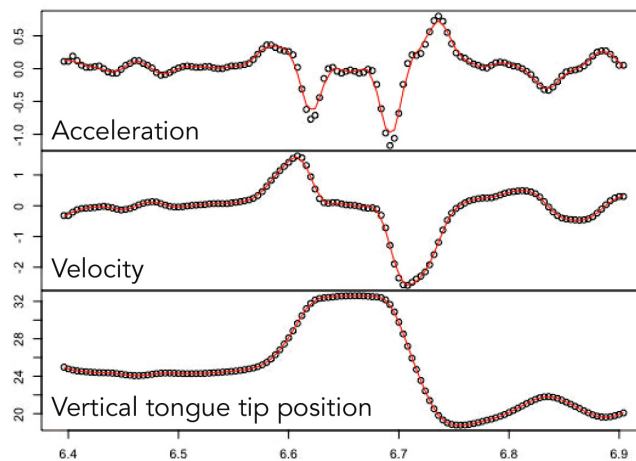


FIG. 4. (Color online) An example of the smoothed articulatory data in a target word that includes an alveolar segment. The acceleration (top), velocity, and position signals (bottom) of the TT sensor vertical dimension are depicted. As the TT is a crucial articulator in this word, we observe that the two acceleration peaks denote the edges of the TT constriction (the vertical position). The red solid line is the filtered signal after using the *R* function *loess* ($\text{span} = 0.1$).

for *peak acceleration in the vicinity of the C1 offset*, etc). The time windows were specified as approximately 80–160 ms time windows, which depended on a number of factors, such as speaker, speech rate, type of landmark, and target word. The time windows occasionally had to be manually adjusted after inspection. As the acceleration signal was smoothed, there was not much distortion. The acceleration landmarks (peak acceleration and peak deceleration) were, therefore, visually prominent in *R* (see Figs. 2 and 4 for reference) and could also be retrieved by the script without much manual adjustment.

3. Measurements and calculations

The collected acceleration landmarks (Table III) were used to calculate the posture intervals (for correlation with the consonant and vowel segments) and the distance between two posture intervals (for correlation with the vowel segments). As Table III shows, the posture intervals of the consonants were calculated for word-initial position

TABLE III. The ten acceleration landmarks on LA, TT, and TB that are used to (1) calculate the posture intervals (some landmarks were used multiple times) for comparison to the consonant and vowel (C1, V1, and C2) segments, and (2) measure time lags to the acoustic segment boundaries. Note that some segment boundaries are duplicated in the table, e.g., C1 offset and V1 onset refer to the same acoustic segment boundary.

Articulatory posture intervals	Acoustic segment boundary			
	C1 onset	C1 offset/V1 onset	C2 onset/V1 offset	C2 offset
Lip	LA peak deceleration	LA peak acceleration	LA peak deceleration	LA peak acceleration
TT	TT peak deceleration	TT peak acceleration	TT peak deceleration	TT peak acceleration
Lip + TT	—	—	LA peak deceleration	TT peak acceleration
Lip - lip	—	LA peak acceleration	LA peak deceleration	—
TT - TT	—	TT peak acceleration	TT peak deceleration	—
Lip - TT	—	LA peak acceleration	TT peak deceleration	—
TT - Lip	—	TT peak acceleration	LA peak deceleration	—
TB	—	TB peak deceleration	TB peak acceleration	—

(C1) and word-medial position (C2), which was either as a distance between two LA landmarks, lip posture (as in Fig. 2), or between the two TT landmarks (TT posture; see Fig. 4 for an example of two TT acceleration peaks). A combined posture interval on the distance between LA peak deceleration and TT peak acceleration (lip + TT posture) was added specifically for comparison to the cluster segment [m:n] in /nam/.

The posture intervals of the vowels, on the other hand, were calculated in two ways: (1) as the distance between the two posture intervals of the word-initial (C1) and word-medial (C2) consonants, and (2) as the distance between the two TB acceleration landmarks (Table III). Because of the four target word onsets, /mam/, /man/, /nan/, and /nam/, we find four possible combinations of lip and TT landmarks: between two lip posture intervals, which would mean the landmarks LA peak acceleration and LA peak deceleration (as in Fig. 2); between two TT posture intervals (TT peak acceleration - TT peak deceleration); and between a posture interval of a bilabial (LA peak acceleration) and an alveolar (TT peak deceleration), or vice versa (Table III).

Time lags were measured as the distance between a segment boundary and an acceleration peak of an articulator: the lips, TT, and TB, whether they were crucial articulators or not (Fig. 2). For example, the alignment of the acoustic segment boundary C1 onset is measured to LA peak deceleration and TT peak deceleration (Table III). A positive time lag indicates that the acceleration peak precedes the segment boundary. In the case of a negative time lag, this means that the segment boundary comes before the acceleration peak and not after. The start of an acoustic consonant segment is expected to be aligned specifically with *peak deceleration*, whereas the end of the consonant segment is expected to be aligned specifically with *peak acceleration* (Table III). This is also the case for the acoustic vowel segment boundaries and TB acceleration peaks (e.g., V1 onset to TB peak deceleration and V1 offset to TB peak acceleration). However, regarding the alignment of the vowel segment to the consonantal articulators, the opposite pattern is calculated: the acoustic V1 onset is aligned to LA and TT peak acceleration, while V1 offset is aligned to peak deceleration (Table III).

E. Statistical analysis

As a first step of testing, the predicted one-to-one articulatory-acoustic relationship, correlation tests, and linear mixed effects models were performed. For the correlation test, the strength of the relationship is dependent on whether it is tested on acoustic segments and articulatory posture intervals on a *crucial* articulator (expecting strong positive relation) or segments and posture intervals on a *non-crucial* articulator (expecting a weak relation; Table I). Pearson correlation tests were used to assess the relationship between segment duration and the segmental articulations.

The linear mixed effects models were used on the time lags measurements as by-speaker variation is expected in articulatory data. Inter-speaker variability is also expected

because of the multiple responses by each of the 18 speakers. To represent the two levels of the independent variable, *place of articulation* (/m/ or /n/) was set as fixed effects. *Speaker* was set as random effects; random intercept and random slope when warranted (*word/item* was, however, not included as a random effect as it only contains four levels). In addition, a likelihood ratio test was performed to test whether adding the complexity of random slope was warranted. The complexities were added when the model comparison showed that the Akaike information criterion (AIC) value of the more complicated model was lower than the value of the less complicated model by at least two (following [Wieling and Tiede, 2017](#)). The linear mixed effects models were, thus, established as [time lag] ~ place of articulation + (place of articulation^R | speaker) (^R denotes when random slope was added to the model; see Table IV). The models were run in R using the lme4-package ([Bates et al., 2015](#)). *P*-values were obtained by using the lmerTest-package in R ([Kuznetsova et al., 2017](#)). Outliers were excluded by the use of time frames during the automatic data collection in R, and by replacing all of the negative values of the posture intervals and the segments with missing values. All of the statistical tests were performed in R (R Core Team, 2015).

III. RESULT

The first report of results is on the strength of relationship between the acoustic segment duration and the articulatory posture intervals.

A. Correlation between segment duration and posture interval

1. Consonants

The results on the word-initial consonants partly confirm the predictions. First, the correlation is very strong between the lip posture and word-initial C1 segment, [m] (Fig. 5): there is a near to perfect relationship in /man/ ($r=0.94$) and /mam/ ($r=0.89$). The word-initial segment, [n], shows similar tendencies. Between C1 and the TT posture, there is a strong positive relationship in /nam/ ($r=0.82$) and /nan/ ($r=0.79$; Fig. 5).

The word-medial consonants (C2) also follow the predictions of strong relationships: we find a very strong correlation between the word-medial, [n:], and the TT posture in /man/ ($r=0.91$) and /nan/ ($r=0.92$; Fig. 5). As expected, the relationship is also strong between the lip posture and word-medial C2 segment, [m:], where we find a strong positive relationship in /mam/ ($r=0.92$; Fig. 5). However, in /nam/, where the lips are a crucial articulator in word-medial position, the relationship with the lip posture is only moderately positive ($r=0.67$) while the relationship with the TT posture is very strong ($r=0.84$). Furthermore, as can be observed in Fig. 5, the acoustic segment duration is visibly longer (y axis) than the interval duration (x axis). However, these results are indeed expected since the acoustic word-medial segment in this particular target word onset

TABLE IV. Linear mixed effects model results on time lag measurements. Fixed effects, place of articulation; random effects, speakers (intercept); ^R; random slopes by speaker added to the model; *, significant effects. SE = standard error; df = degrees of freedom.

Articulator	Time lag		Estimate	SE	df	t-value	p-value	
LA	C1 onset ^R	(Intercept) /m/	-10.875	0.843	527.780	-12.902	0.000	
		Place of articulation /n/	-22.032	2.344	18.717	-9.399	0.000*	
	C1 offset	(Intercept) /m/	0.542	1.277	31.793	0.425	0.674	
		Place of articulation /n/	0.424	1.273	544.609	0.333	0.739	
	C2 onset ^R	(Intercept) /m/	-14.731	1.060	17.940	-13.896	0.000	
		Place of articulation /n/	1.895	1.291	17.112	1.468	0.160	
	C2 offset ^R	(Intercept) /m/	28.129	1.575	303.500	17.855	0.000	
		Place of articulation /n/	-2.880	3.260	20.238	-0.884	0.387	
	C2 offset ^R	(Intercept) /nam/	56.707	2.664	18.557	21.29	0.000	
Place of articulation /mam/		-56.380	3.090	19.265	-18.24	0.000*		
TT	C1 onset ^R	(Intercept) /m/	-13.795	2.784	17.736	-4.955	0.000	
		Place of articulation /n/	0.369	3.316	17.673	0.111	0.913	
	C1 offset ^R	(Intercept) /m/	0.438	2.141	17.141	0.205	0.840	
		Place of articulation /n/	-10.278	3.251	19.231	-3.161	0.005*	
	C2 onset	(Intercept) /m/	-15.856	1.300	532.000	-12.196	0.000	
		Place of articulation /n/	-13.215	1.882	532.000	-7.023	0.000*	
	C2 offset ^R	(Intercept) /m/	-9.896	1.956	18.260	-5.060	0.000	
		Place of articulation /n/	11.147	1.901	21.662	5.864	0.000*	
	C2 offset ^R	(Intercept) /nam/	-1.914	1.610	250.232	-1.189	0.236	
		Word onset /mam/	-17.051	4.217	24.845	-4.044	0.000*	
	TB	V1 onset ^R	(Intercept) /m/	-28.294	4.373	18.003	-6.470	0.000
			Place of articulation /n/	-20.435	3.742	17.985	-5.460	0.000*
V1 offset		(Intercept) /m/	19.896	2.136	23.629	9.315	0.000	
	Place of articulation /n/	10.990	1.461	520.330	7.520	0.000*		

includes the full heterorganic cluster, [m̥n] (Sec. IID 1). Therefore, as Fig. 6 shows, when correlated with the combined lip + TT posture (i.e., the two crucial articulators that together create the consonant cluster), we find an almost perfect relationship ($r = 0.95$). This relationship is also very strong for the word-medial C2 segment, [n̥], in /nan/ ($r = 0.86$) and /man/ ($r = 0.83$), where the combined posture interval is partly based on the peak acceleration of the TT (Fig. 6).

The expectation of a weak relationship with a *non-crucial* articulator is not entirely met. Instead, the relationship is stronger than expected between the lip posture interval and the segment [n]: moderately positive with C1 in /nan/ ($r = 0.56$) and /nam/ ($r = 0.6$) and C2 in /man/ ($r = 0.64$) and /nan/ ($r = 0.69$; Fig. 5). As for the correlation with the TT posture, the prediction of a weak relationship with the segment [m] is confirmed for C1 in /man/ ($r = 0.21$) and C2 in /mam/ ($r = 0.22$), whereas a moderately positive relation is found for C1 in /mam/ ($r = 0.49$; Fig. 5).

2. Vowels

The vowel results on the strength of the relationship include four different calculated distances between two posture intervals, as well as the TB posture (Table III). For the calculated distance between intervals, the predicted outcome is that the acoustic vowel segment should display a very strong relationship only when the two posture intervals are both measured on a crucial articulator. Indeed, an expected very strong relationship is found between the acoustic vowel segment and all four of the calculated intervals between two

crucial articulators (Fig. 7). This can be seen between the lip - lip posture and the acoustic vowel in /mam/ ($r = 0.83$), the TT - TT posture and the vowel in /nan/ ($r = 0.84$), the lip - TT posture with the vowel in /man/ ($r = 0.86$), and the TT - lip posture with the acoustic vowel segment in /nam/ ($r = 0.88$; Fig. 7). However, no weak relationships with the acoustic vowel segment are found when the two posture intervals are measured on the non-crucial articulators. Instead, the lip - lip posture is strongly correlated with the vowel segment in /nan/ ($r = 0.70$); the TT - TT posture is moderately correlated with the vowel in /mam/ ($r = 0.52$); the lip - TT posture is moderately correlated with the acoustic vowel in /nam/ ($r = 0.54$); and the TT - lip posture is very strongly correlated with the vowel segment in /man/ ($r = 0.86$; Fig. 7). Moreover, when one of the postures include a crucial articulator, we find quite a strong relationships as well: a strong relationship is found between the lip - lip posture and the vowel segment in /man/ ($r = 0.75$) and /nam/ ($r = 0.70$; Fig. 7). There is also a strong relationship between The TT - TT posture and the acoustic vowel in /nam/ ($r = 0.67$) and a very strong relationship with the vowel in /man/ ($r = 0.90$; Fig. 7). Between the lip - TT posture and the acoustic vowel segment, there is a strong correlation in /mam/ ($r = 0.72$) and /nan/ ($r = 0.74$); and similar results are found between the TT - lip posture and the vowel in /mam/ ($r = 0.55$) and /nan/ ($r = 0.80$; Fig. 7).

On the other hand, the relationship between the acoustic vowel segment and the TB posture is mostly weak, but positive. We find the weakest relationship in /mam/ ($r = 0.15$),

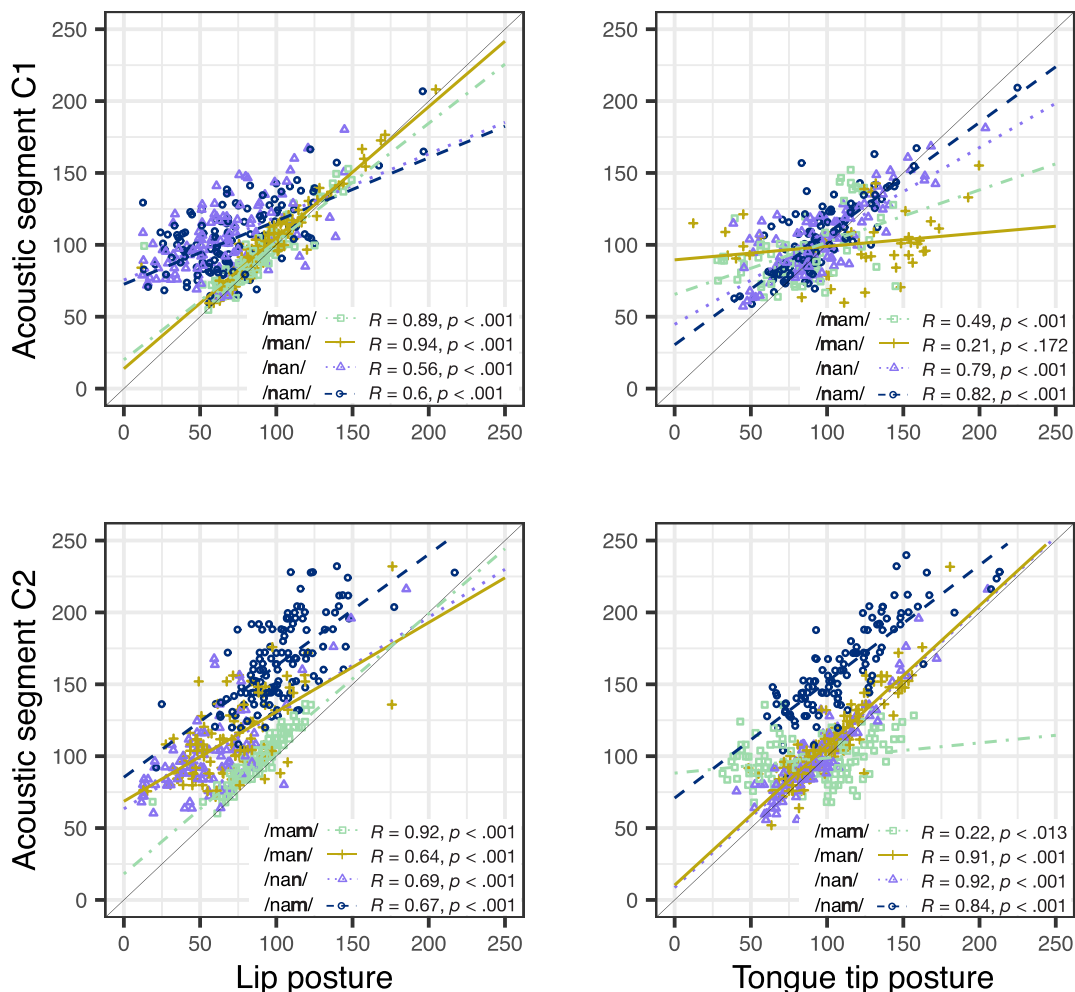


FIG. 5. (Color online) Correlation plots with correlation coefficients on the relationship between the acoustic consonant segments (y axis, in ms) and articulatory lip and TT posture intervals (x axis, in ms). The diagonal line represents a perfect correlation.

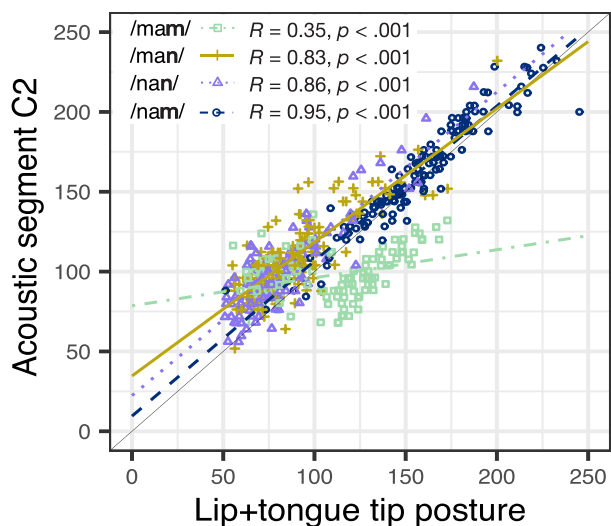


FIG. 6. (Color online) Correlation plots with correlation coefficients on the word-medial consonant (y axis, in ms) and combined lip lip + TT posture interval (x axis, in ms). The diagonal line represents a perfect correlation.

where the lips are the crucial articulators in both onset and coda position (Fig. 8). Of the four target word onsets the strongest relationship is found in /nan/ ($r=0.44$) where the TT is active both before and after the presumed TB constriction location of the vowel (Fig. 8). Moreover, the correlation between the TB posture and the acoustic vowel segment is weak in /man/ ($r=0.29$), but moderately positive in /nam/ ($r=0.40$) (Fig. 8). Judging by Fig. 8, the TB posture intervals are shorter than the acoustic vowel segments.

B. Time lag between segment boundary and acceleration peak

Time lag measurements are used to see whether an acoustic segment boundary is temporally aligned with an acceleration peak. Hence, expected results are short time lags when measured on a crucial articulator while longer and more varied time lags are expected when measured on a non-crucial articulator.

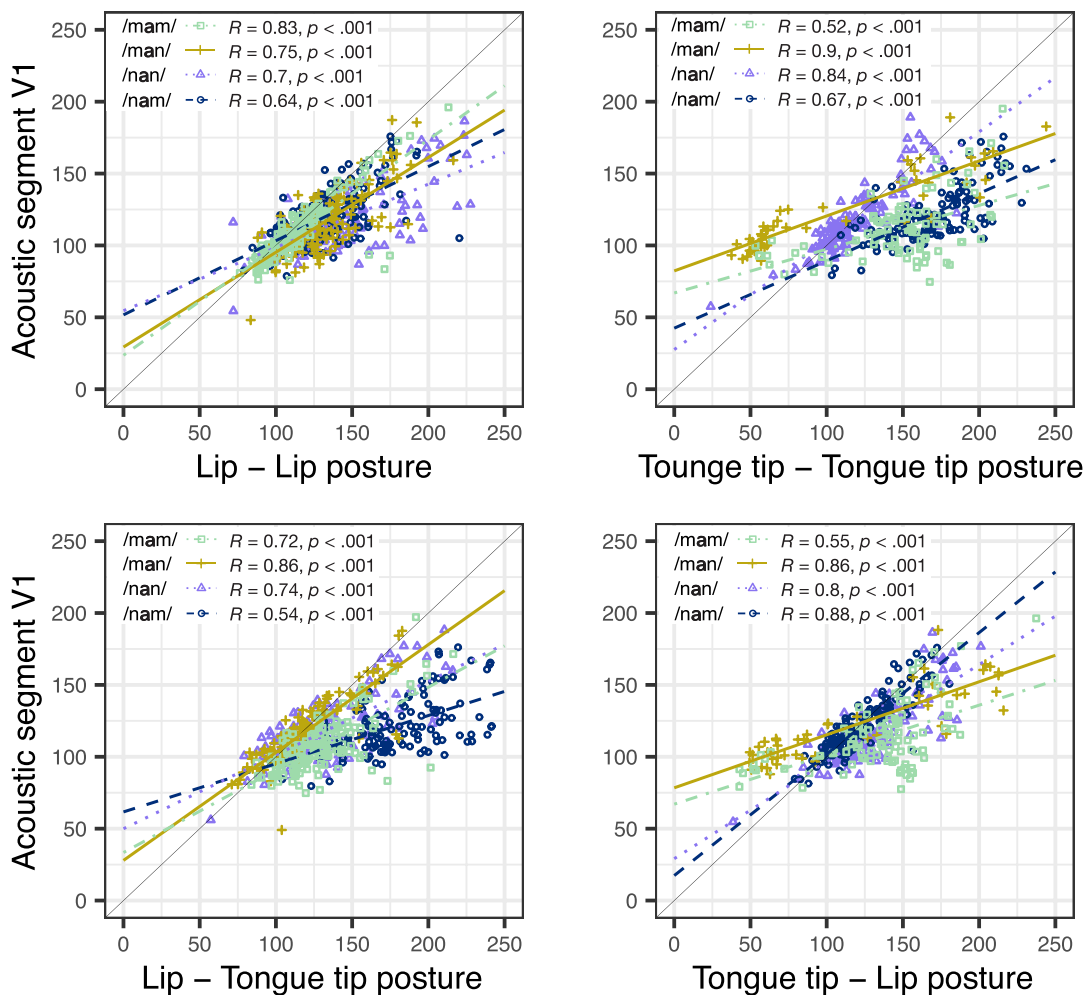


FIG. 7. (Color online) Correlation plots with correlation coefficients on the acoustic vowel segments (y axis, in ms) and calculated distances between two neighboring but non-adjacent lip and/or TT posture intervals (x axis, in ms). The diagonal line represents a perfect correlation.

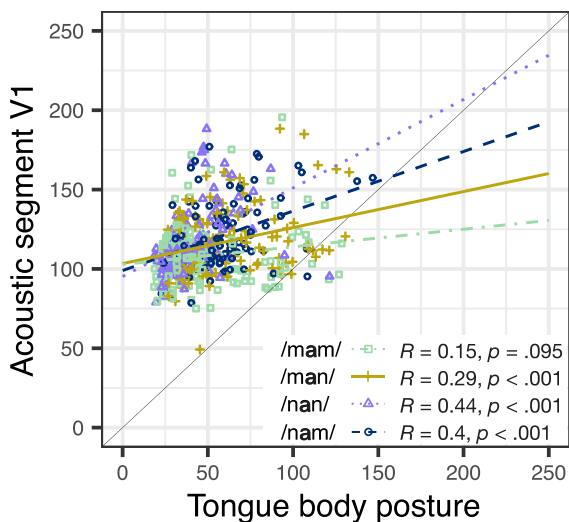


FIG. 8. (Color online) Correlation plots with correlation coefficients on the acoustic vowel segment (y axis, in ms) and the articulatory TB posture (x axis, in ms). The diagonal line represents a perfect correlation.

1. LA

Figure 9 displays time lags on the LA acceleration landmarks, where zero represents the acceleration peaks, and timing of the acoustic segment boundaries are represented by the boxes. At C1 onset, we see better aligned segment boundaries in /mam/ and /man/, where the lips are a crucial articulator, than in the words with a word-initial [n], where the TT is the crucial articulator. In /nan/ and /nam/, the time lags are instead longer and more varied (Fig. 9). According to the linear mixed effects model, which includes random slopes by speaker, the time lag differences between the two places of articulation are statistically significant ($t = -9.4, p < 0.001$; Table IV). A similar pattern seems to be present at the word-medial C2 onset: when place of articulation is the lips, the time lags appear shorter and slightly less varied in /mam/ and /nam/ (Fig. 9). However, the mixed effects model displays no statistically significant effect by place of articulation on the time lags at C2 onset ($t = 1.5, p < 0.16$; Table IV).

As for the end of the word-initial segments, the time lags at C1 offset appear more varied in /nan/ and /nam/,

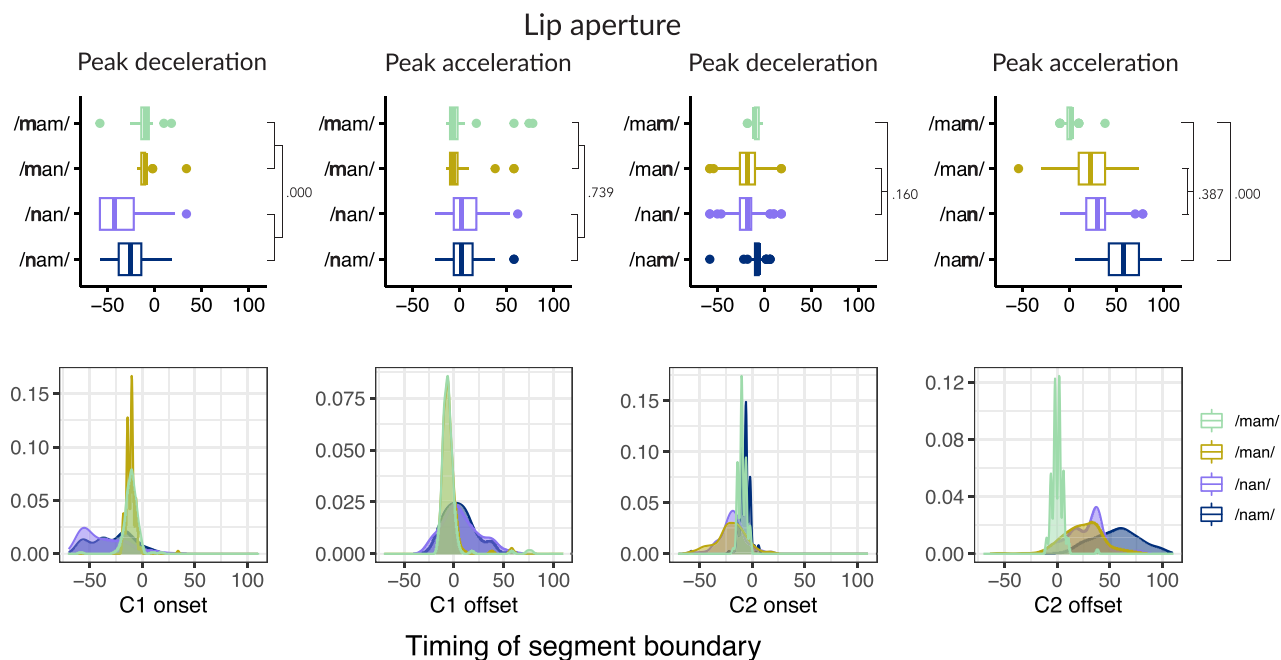


FIG. 9. (Color online) Boxplots and density plots on the time lags between LA acceleration landmarks and acoustic segment boundaries (x axis, in ms). A positive time lag indicates that the segment boundaries (boxes) follow the acceleration peaks (the reference point).

where the lips are non-crucial, however, the difference between places of articulation is not significant ($t = 0.33$, $p < 0.739$; Table IV). At the end of the word-medial segment (C2 offset), there is no time lag in /mam/, indicating a perfect alignment of segment boundary to peak acceleration. However, since the word-medial segment in /nam/ includes the heterorganic cluster, [m̥n], the segment boundary at C2 offset will align with the TT acceleration landmark instead of the lip landmark (Fig. 10). A significant difference is,

therefore, only found when comparing individual target word onsets /mam/ to /nam/ ($t = -18.2$, $p < 0.001$; Table IV).

Although the time lags are very short when measured on crucial articulators, the direction of the time lags is unexpected. The negative lags indicate that the segment boundaries at C1 and C2 onset seem to occur slightly before the peak deceleration landmarks and at C1 offset before peak acceleration in /man/ and /mam/ (Fig. 9). Only at C2 offset do we find the expected pattern of a segment boundary

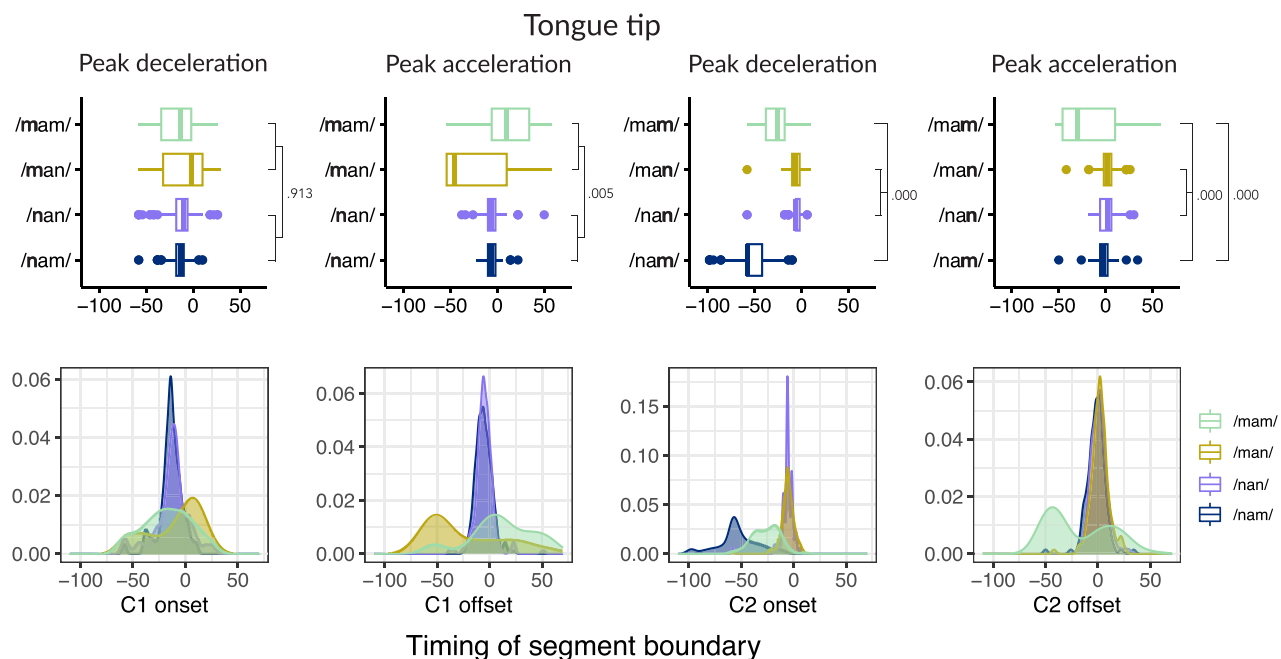


FIG. 10. (Color online) Boxplots and density plots on the time lags between TT acceleration landmarks and acoustic segment boundaries (x axis, in ms). A positive time lag indicates that the segment boundaries (boxes) follow the acceleration peaks (the reference point).

following the peak acceleration landmark (Fig. 9). None of the density plots on the crucial articulator, where we expect short or no time lags, show skewed results, although there are signs of multimodal distribution, which could be an indicator of a low sample size (Fig. 9).

2. TT

When the TT is a crucial articulator, there are no time lags or very short time lags between the TT peak deceleration and the start of the segments (C1 and C2 onset) and between peak acceleration and the end of the segments (C1 and C2 offset; Fig. 10). The differences are statistically significant between the places of articulation at C1 offset ($t = -3.2, p < 0.01$), C2 onset ($t = -3, p < 0.01$), and C2 offset between place of articulation ($t = -5.9, p < 0.001$) and between, specifically, /mam/ and /nam/ ($t = -4.0, p < 0.001$; Table IV). Figure 10 displays a similar result for the start of the word-initial consonant (C1 onset), but the difference is not significant between places of articulation ($t = 0.1, p < 0.913$). However, we can see a pattern that is similar to the results on the lips: the segment boundary seems to be approximately 13 ms before the acceleration peak and not after as would be expected given the prediction of a causal relationship (Table IV). The density plots are not skewed other than at C2 onset, where there are signs of a multimodal distribution.

3. TB

As the correlation results showed weak relationships between the TB posture and vowel acoustic segment, long and varied time lags are expected between vowel segment boundaries and the TB acceleration peaks. Indeed, Fig. 11 displays no alignment and, instead, time lags at V1 onset and V1 offset for all of the target word onsets. At V1 onset, we see negative time lags (meaning that the segment boundary is before peak deceleration) of about 30 ms for a word-initial /m/ and 50 ms for /n/ (Fig. 11). The linear mixed effects model (with random slopes by speaker) predicts a statistically significant difference between places of articulation ($t = -5.5, p < 0.001$; Table IV).

At V1 offset, the acoustic segment boundaries follows the acceleration peaks, instead, as indicated by the 20–30 ms positive time lags (Fig. 11). The difference between places of articulation is statistically significant at V1 offset ($t = 7.5, p < 0.001$) with slightly longer time lags when the following consonant is an alveolar (Table IV). The combination of negative and positive time lags for the two measures reveals that the TB posture interval is shorter than the acoustic vowel segment and even more so when the adjacent consonant is [n] (when the crucial articulator is the TT; Fig. 11).

IV. DISCUSSION

The results confirm the predictions of strong relationships between a posture interval on a crucial articulator and an acoustic segment, and short or no time lags between an acceleration peak of a crucial articulator and an acoustic

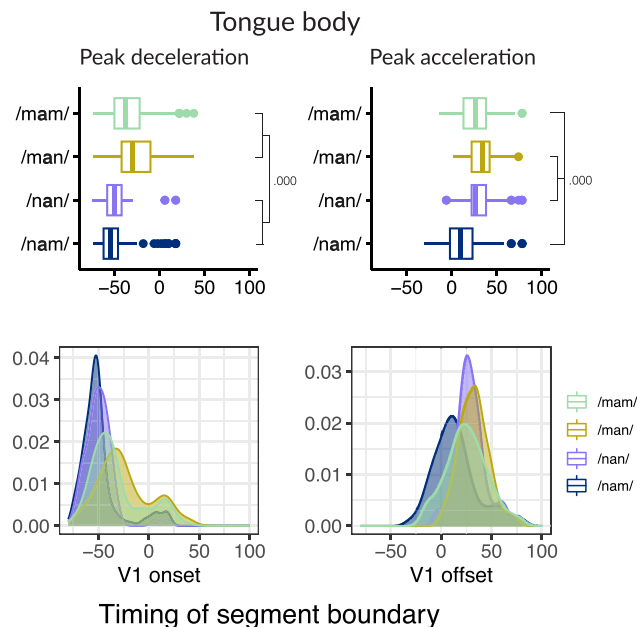


FIG. 11. (Color online) Boxplots and density plots on the time lags between TB acceleration landmarks and acoustic vowel segment boundaries (x axis, in ms). A positive time lag indicates that the segment boundaries (boxes) follow the acceleration peaks (the reference point).

segment boundary. The hypothesis is, therefore, not rejected: acceleration is likely to be responsible for the rapid movements of the segment transitions, that is, the acoustic segment boundaries.

The correlation results reveal that the strength of the relationship is higher between [m] and the lip posture interval (the *crucial* articulator) than between [m] and the TT posture (the *non-crucial* articulator). Additionally, the correlation is stronger between [n] and the TT posture interval (the *crucial* articulator) than between [n] and the lip posture (the *non-crucial* articulator). Furthermore, the time lag results reveal that short or no time lags are found when measured on crucial articulators, whereas long varied time lags are found when measured on non-crucial articulators. However, these results are only applicable for consonant segments. As for the vowel results, a posture interval on the crucial articulator (the TB) does not correspond to the acoustic vowel segment. Instead, the duration of an acoustic vowel segment seems determined by the timing of two adjacent consonantal posture intervals. Very strong relationships are found when both adjacent posture intervals are measured on *crucial* articulators, although the results also reveal no weak relationships on the distance between two posture intervals (whether they are crucial articulators or not). The time lag measurements confirm the correlation results: no alignment to the acceleration peaks of the TB. However, the acceleration peaks of TT and lips are well aligned with C1 offset and C2 onset segment boundaries, which are also, coincidentally, the start and end of the acoustic vowel segment (V1 onset and V1 offset, respectively). To summarize, the acoustic segment boundaries of the consonant and the vowel are determined by the acceleration peak timing of the

crucial consonantal articulators or what can be referred to as the *segmental articulators*.

There are some exceptions to the predictions. First, the word-initial and word-medial segments appear to differ: [n] displays a weaker relation to the TT posture interval in word-initial position than in the word-medial position. There may be several explanations to this finding. On the one hand, the TT may be using a different strategy in onset as opposed to coda position as suggested by the time lag results in which the segment boundaries were more aligned in coda position (Fig. 10). On the other hand, we might find a natural explanation outside of the target word. Because each target word follows a verb that contains the final VCV sequences, /ade/, the tip of the tongue is a crucial articulator in the sequences before the target word onsets. Thus, a nonlocal effect of the crucial role of the TT could have affected the word-initial segments in this study. In this particular context, the TT might stay close to the target, shaping a movement plateau, causing the analyzing script to either not find or misplace the accelerating moment at the start of the interval. This presumed nonlocal effect seems to not affect the distance between two postures as strong relationships are found with the acoustic vowel segments in /nam/ and /nan/ (Fig. 7). This suggests that peak acceleration at the end of the TT posture is unaffected, and the preceding word only affects the start of the interval: the peak deceleration of the crucial articulator.

A. Timing of segment boundary

Related to the above are the unexpected results that acoustic segment boundaries appear before the acceleration peaks of the lips and TT and not after as expected given the hypothesis of a causal relationship. The negative time lags are the longest at the peak deceleration at word onset. Moreover, we see time lag differences between syllabic positions and between deceleration and acceleration of a movement. Although the consonants investigated in this study all follow the same manner of articulation, i.e., they are nasals, the active movement of the velum has not yet been addressed. The differences in time lags between onset and coda could, in fact, be a result of velar lowering differences between syllabic positions in nasals as has been reported elsewhere (Ushijima and Hirose, 1974; Bell-Berti and Krakow, 1991; Krakow, 1993; Byrd *et al.*, 2009): the velar in word-final position lowers already during a preceding vowel, whereas in the word-initial position, it is synchronized with the oral closure, a pattern that has been found in [n] and [m]. As the segment boundary appears before the acceleration landmarks of the lips and TT at word onset, the effect of the velar lowering cannot be ruled out.

Another possible reason for the negative time lags could be a methodological one. The EMA lip sensors are placed at the vermillion border, and it is possible that the placement does not capture the labial movements in their full entity. As the lips are soft anatomical structures, a constriction at the margins of the lips may precede in time a “complete

closure” as captured by the EMA sensors. Also, the TT sensor, placed 1 cm from the TT, may not capture the initial moment of closure at the palate. Moreover, yet another factor explaining the negative time lags could be that the acoustics of the vocal tract changes may be present even before the actual constriction of the crucial articulator, e.g., the upper and lower lip touching each other, or an effect by the velar lowering. However, as the unexpected direction of the time lag is present across all of the segment boundaries for both segmental articulators (crucial consonantal articulators) and, in addition, decreases throughout the word, it could also be related to articulatory effort. Articulatory effort is known to be more present at word onset and, besides, in the case of speakers inserting a phrase boundary, it too would entail even more articulatory effort. The acceleration peak signifies the time that an articulator slows down or speeds up the most. As more articulatory effort would, in theory, correspond to longer distances to travel and faster movement, for any given articulator, articulatory effort could affect when the articulator starts to accelerate or decelerate. Then, as the alignment is clearly there, the negative time lags could also suggest that the segment boundary is not necessarily because of the peak of the acceleration but, instead, related to the *jerk* of the movement. A closer look at segment boundaries and lip constrictions, also across different manners of articulation, which would include comparing segmentation strategies, would help in understanding the phenomena further.

B. Relationships with non-crucial articulators

Another exception to the prediction is the strong relation to non-crucial articulators, for example, between the alveolar consonant, [n], and the non-crucial lip posture, which is unexpectedly moderately positive, in the word-initial and word-medial positions. A strong relationship between lip movement characteristics and the alveolar acoustic segment duration may be related to the fact that apart from being mechanically connected, the lips and jaw work together as a control unit (Gracco, 1988; Kollia *et al.*, 1995; Bose and van Lieshout, 2012). As the lower lip is highly coordinated with the jaw, it can be assumed that the degree of lip opening is affected by the jaw kinematics. Thus, during the jaw opening of a vowel, the lower lip moves as well, irrespective of any status as a crucial or non-crucial articulator. In addition, the jaw displacement has been shown to display different degrees of opening depending on the manner of coronal constrictions (Lindblom, 1983; Mooshammer *et al.*, 2007). Moreover, the amount of jaw displacement correlates strongly with degree of syllable prominence once the intrinsic vowel height effects have been factored out (Erickson and Kawahara, 2016). As such, the jaw is functional in syllable and vowel production and during the different manners of TT articulations. Thus, there are several reasons for why we would see a correlation between the lips and [n] irrespective of the functional role of the jaw.

C. The results on vowels

When the TB is the crucial articulator only weak relationships with the acoustic vowel segments are found. This is opposite of the results found on the consonantal articulators' lips and TT. Instead, the TB posture intervals, which represent the place of constriction of the vowel, appear to be shorter than the acoustic vowel segments. However, the relationship strength between the TB acceleration and vowel segment is affected by the place of articulation of the consonants as the correlation is stronger in /nan/ than in /mam/ (Fig. 8). The TB posture seems to also be shorter in /nan/ (Fig. 8), as is similarly indicated by the longer time lags when the adjacent consonant is [n] in Fig. 11. Within the time window of the acoustic vowel segment, there is even a slightly right-skewed timing in /nan/: the diphthongal vowel seems to end in a TB constriction closer to the syllable coda constriction than the syllable onset (Fig. 11). This pattern of a shorter and right-skewed TB posture interval could be explained by the biomechanical connection between the TT and TB: the TB may need to leave its target sooner for the TT to reach its target in time. It could also be a reflection of the fact that the acoustic intervals for the nasals exclude the transitions, the rapid intervals, and the acoustic vocalic intervals include them. Possibly, the TB acceleration is more coordinated with the rapid movements, peak velocity, of the consonantal articulators and not with the acceleration peaks denoting the acoustic segments.

As for the correlation with the segmental (consonantal) articulators, strong relationships are found with the acoustic vowel segments when the lip and/or TT are crucial articulators. However, when the lips and/or TT are non-crucial, we still see quite a strong correlation with the acoustic vowel segment, as also shown in Fig. 9 on the time lags to the lip acceleration landmarks at C1 offset and C2 onset. This pattern may be explained by the jaw cycle: as the jaw is open during the vowel segment, it affects the movement characteristics of the other active articulators, whether they are crucial or not. For example, at syllable onset, the jaw opening movement entails a longer distance for a target constriction to undertake, which possibly causes the timing of the acceleration to change. Likewise, in syllable coda, during the closing of the jaw, the change in jaw displacement presumably affects when deceleration occurs. The acceleration and deceleration phases of the two articulators involved in the CVC sequence are, thus, affected by the jaw opening as the displacement of the jaw determines the distance to the target and, therefore, also changes the conditions for achieving the target, hence, affecting task difficulty. Task difficulty is related to the stiffness and damping of a mass (i.e., the articulator), a relationship based on Fitt's law, which, in short, describes a linear relationship between speed, distance, and accuracy in a movement (see, e.g., Bootsma *et al.*, 2004). The law is generally robust and holds for most conditions, but in-depth studies of the underlying kinematic processes are required to, e.g., define task difficulty (Bootsma *et al.*, 2004). For instance, Fitt's law has been

shown to be valid only for very fast speech movements beyond an individual speaker's critical speech rate (Kuberski and Gafos, 2021). One aspect to consider for speech may be how an acceleration phase (stiffness) and a deceleration phase (damping) are different in function and may also differ in shape (Bootsma *et al.*, 2004). As stiffness and damping are related to the speed of the articulator, a faster movement equals more stiffness, i.e., a right-skewed shape, whereas more damping results in a slower movement, i.e., a left-skewed shape (Bootsma *et al.*, 2004; Iskarous, 2017). As an acceleration peak is the moment in time when velocity changes the most, the task difficulty of the target could be directly related to the timing of peak acceleration and peak deceleration. Possibly, timing of peak acceleration is correlated more to stiffness while peak deceleration is correlated more to damping. Furthermore, acoustic segment duration depends on a number of linguistic factors (Elert, 1964; Cho and Ladefoged, 1999; Heldner, 2001; Fletcher, 2010; Foulkes *et al.*, 2013; Harrington, 2010; Turco and Braun, 2016; Svensson Lundmark *et al.*, 2017; Svensson Lundmark, 2022b), and any of these factors may be related to task difficulty, affecting either stiffness or damping of the articulator and, as a result, the timing of peak acceleration and peak deceleration. Hence, acoustic segment variation, such as reduction and lengthening, is a direct result of timing of acceleration. Moreover, acoustic segment duration of not only the consonants themselves but also the vowels are a result of the articulatory strategies of the constrictions, which are highly dependent on the segmental context. Specifically, the acoustic vowel segments are conditional on the timing of the constrictions, which, in turn, is coordinated with the height of the TB, jaw displacement, and other factors, also explaining acoustic segment duration differences between vowel types. It is an intricate coordinated articulatory relationship, however; in the midst of it, acceleration is suggested to be systematic and denotes the boundaries of the acoustic segment irrespective of the cause for the timing of its peak. Because of this intricately coordinated relationship, we see strong relations between the resulting vowel segment and consonantal posture intervals, whether they are based on crucial articulators or not. In fact, it may be useful for a continued discussion on this to further separate the crucial articulators and refer to the consonantal articulators as *segmental articulators* and TB as a *non-segmental articulator*. Moreover, the moments of peak acceleration and peak deceleration at the edges of the syllable might be more variable and context-dependent than the peak acceleration and peak deceleration in the proximity of the vowel segment in the syllable nucleus as these display a strong correlation with any of the acoustic vowel segments.

V. CONCLUSION

The present paper reports on a one-to-one relation between articulation and acoustics, where segment boundaries are proposed to be a result of rapid articulatory movements. In the acceleration profile, these are identified as

acceleration peaks. To test the hypothesis, the acceleration peaks are measured and compared to the acoustic segmentation in two ways: by calculating articulatory posture intervals, which are correlated with the acoustic segments, and aligning the acoustic segment boundaries to the acceleration peaks. The results confirm the hypothesis and suggest that (a) rapid articulatory movements at the segment transitions consist of acceleration peaks; (b) acoustic segments, consonants and vowels, are determined by the timing of acceleration of the crucial consonantal articulators: the *segmental articulators*; (c) place of articulation does not matter as the phenomena is present at lip and TT movement; and (d) acceleration of the TB, i.e., the *non-segmental articulator*, does not correlate with the acoustic vowel segment boundaries. Moreover, jaw movements may affect the results, specifically, the acoustic vowel segments which display correlation with acceleration of lips and TT. As the jaw is coordinated with the lips, the lip postures, too, are correlated with most consonant segments, although to a varying degree.

Previous research shows that acoustic changes perceived by listeners are related to even small changes in bodily movements (Iskarous, 2016; Pouw *et al.*, 2020). A next step may be to make use of this knowledge to further test the proposed causal one-to-one relationship between acceleration and the large acoustic changes at segment boundaries and the effect on listeners. In addition, future studies will continue to include different manners and places of articulation and the effect of the jaw. This is ongoing work in which preliminary results show that the proposed one-to-one relationship is, indeed, present across tonal contexts (Svensson Lundmark, 2022b), prosodic levels (Svensson Lundmark and Frid, 2022), and, to a varying degree, across different manners of lip and TT movements (Svensson Lundmark, 2022a).

The present approach makes it possible to examine and predict acoustic segment duration regardless of the cause of segmental variation. It suggests a joint invariance in acoustics and articulation. This approach is based on a strategy where placement of the segment boundary is determined by manual segmentation of the constrictions. Hopefully, this line of work enables a discussion on an alternative method of phoneme analysis that is not based on orthographic segmentation, which is unfortunately a possible source for circular arguments.

ACKNOWLEDGMENTS

This work was supported by an International Postdoc grant from the Swedish Research Council (Grant No. 2021-00334) and has, in part, been funded by an infrastructure grant from the Swedish Research Council (SWE-CLARIN, 2018–2024; Grant No. 2017-00626). The author gratefully acknowledges the Lund University Humanities Laboratory. Helpful comments on a draft of this article were made by Professor Donna Erickson. M.S.L. thanks two anonymous reviewers for insightful suggestions on the submitted

manuscript, with a special thank you to Dr. Man Gao, and Professor Sven Strömqvist, Dr. Gilbert Ambrazaitis, and Dr. Oliver Niebuhr are also thanked for their valuable input on this study during the different phases of research.

Bates, D., Maechler, M., Bolker, B. M., and Walker, S. C. (2015). "Fitting linear mixed-effects models using lme4," *J. Stat. Soft.* **67**(1), 1–48.

Bell-Berti, F., and Krakow, R. A. (1991). "Anticipatory velar lowering: A coproduction account," *J. Acoust. Soc. Am.* **90**, 112–123.

Boersma, P., and Weenink, D. (2018). "Praat: Doing phonetics by computer (version 6.0.37) [computer program]," available at <http://www.praat.org> (Last viewed 3 February 2018).

Bombien, L., Mooshammer, C., and Hoole, P. (2013). "Articulatory coordination in wordinitial clusters of German," *J. Phon.* **41**(6), 546–561.

Bootsma, R. J., Fernandez, L., and Mottet, D. (2004). "Behind Fitts' law: Kinematic patterns in goal-directed movements," *Int. J. Human-Comput. Stud.* **61**(6), 811–821.

Bose, A., and van Lieshout, P. (2012). "Speech-like and non-speech lip kinematics and coordination in aphasia: Movement kinematics and coordination in aphasia," *Int. J. Lang. Commun. Disorders* **47**(6), 654–672.

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

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

Browman, C. P., and Goldstein, L. (2000). "Competing constraints on inter-gestural coordination and self-organization of phonological structures," *Les Cahiers l'ICP, Bull. Commun. Parlée* **5**, 25–34.

Byrd, D. (1996). "Influences on articulatory timing in consonant sequences," *J. Phon.* **24**(2), 209–244.

Byrd, D., Tobin, S., Bresch, E., and Narayanan, S. (2009). "Timing effects of syllable structure and stress on nasals: A real-time MRI examination," *J. Phon.* **37**(1), 97–110.

Cho, T., and Ladefoged, P. (1999). "Variation and universals in VOT: Evidence from 18 languages," *J. Phon.* **27**, 207–229.

Eager, D., Pendrill, A.-M., and Reistad, N. (2016). "Beyond velocity and acceleration: Jerk, snap and higher derivatives," *Eur. J. Phys.* **37**(6), 065008.

Elert, C.-C. (1964). *Phonologic Studies of Quantity in Swedish* (Almqvist and Wiksell, Uppsala, Sweden).

Erickson, D., Kawahara, S., Moore, J., Menezes, C., Suemitsu, A., Kim, J., and Shibuya, Y. (2014). "Calculating articulatory syllable duration and phrase boundaries," in *Proceedings of the 10th International Seminar on Speech Production (ISSP)*, edited by S. Fuchs, M. Grice, A. Hermes, L. Lancia, and D. Mücke, Cologne, Germany, pp. 102–105.

Erickson, D., and Kawahara, S. (2016). "Articulatory correlates of metrical structure: Studying jaw displacement patterns," *Linguist. Vanguard* **2**(2), 102–110.

Fant, G., and Lindblom, B. (1961). "Studies of minimal speech sound units," Dept. for Speech, Music Hear. Quart. Prog. Status Rep. **2**(2), 1–11.

Fletcher, J. (2010). "The prosody of speech: Timing and rhythm," in *The Handbook of Phonetic Sciences*, 2nd ed., edited by W. J. Hardcastle, J. Laver, and F. E. Gibbon (Wiley-Blackwell, Chichester, UK), pp. 523–602.

Foulkes, P., Scobbie, J., and Watt, D. (2013). "Sociophonetics," in *The Handbook of Phonetic Sciences*, 2nd ed., edited by W. J. Hardcastle, J. Laver, and F. E. Gibbon (Wiley-Blackwell, Chichester, UK), pp. 703–754.

Fowler, C. A. (1986). "An event approach to the study of speech perception from a directrealistic perspective," *J. Phon.* **14**, 3–28.

Fowler, C. (1996). "Listeners do hear sounds, not tongues," *J. Acoust. Soc. Am.* **99**, 1730–1741.

Fowler, C. A., and Saltzman, E. (1993). "Coordination and coarticulation in speech production," *Lang. Speech* **36**(2–3), 171–195.

Fujimura, O. (2000). "The C/D model and prosodic control of articulatory behavior," *Phonetica* **57**(2–4), 128–138.

Gårding, E. (1967). *Internal Juncture in Swedish* (Gleerup, Lund, Sweden).

Gracco, V. (1988). "Timing factors in the coordination of speech movements," *J. Neurosci.* **8**(12), 4628–4639.

Harrington, J. (2010). "Acoustic phonetics," in *The Handbook of Phonetic Sciences*, 2nd ed., edited by W. J. Hardcastle, J. Laver, and F. E. Gibbon (Wiley-Blackwell, Chichester, UK), pp. 81–129.

- Heldner, M. (2001). "On the non-linear lengthening of focally accented Swedish words," in *Nordic Prosody: Proceedings of the VIIIth Conference, Trondheim 2000*, edited by W. van Dommelen and T. Fretheim (Peter Lang Publishing Group, Frankfurt am Main, pp. 103–112).
- Iskarous, K. (2016). "Compatible dynamical models of environmental, sensory, and perceptual systems," *Ecol. Psychol.* **28**(4), 295–311.
- Iskarous, K. (2017). "The relation between the continuous and the discrete: A note on the first principles of speech dynamics," *J. Phon.* **64**, 8–20.
- Jakobson, R., Fant, G., and Halle, M. (1969). *Preliminaries to Speech Analysis* (MIT Press, Cambridge, MA).
- Jenkins, J. J., Strange, W., and Trent, S. A. (1999). "Context-independent dynamic information for the perception of coarticulated vowels," *J. Acoust. Soc. Am.* **106**(1), 438–448.
- Kawahara, S., Masuda, H., Erickson, D., Moore, J., Suemitsu, A., and Shibuya, Y. (2014). "Quantifying the effects of vowel quality and preceding consonants on jaw displacement," *J. Phon. Soc. Jpn.* **18**(2), 54–62.
- Kollia, H. B., Gracco, V. L., and Harris, K. S. (1995). "Articulatory organization of mandibular, labial, and velar movements during speech," *J. Acoust. Soc. Am.* **98**(3), 1313–1324.
- Krakow, R. A. (1993). "Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate," in *Nasals, Nasalization, and the Velum (Phonetics and Phonology V)*, edited by M. A. Huffman, and R. A. Krakow (Academic, New York), pp. 87–116.
- Kuberski, S. R., and Gafos, A. I. (2021). "Fitts' law in tongue movements of repetitive speech," *Phonetica* **78**(1), 3–27.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). "lmerTest package: Tests in linear mixed effects models," *J. Stat. Soft.* **82**(13), 1–26.
- Lindblom, B. (1983). "Economy of speech gestures," in *The Production of Speech*, edited by P. MacNeilage (Springer, New York).
- Machač, P., and Skarnitzl, R. (2009). *Principles of Phonetic Segmentation* (Epocha Publishing House, Prague).
- Marin, S. (2013). "The temporal organization of complex onsets and codas in Romanian: A gestural approach," *J. Phon.* **41**(3–4), 211–227.
- Mooshammer, C., Hoole, P., and Geumann, A. (2007). "Jaw and order," *Lang. Speech* **50**(2), 145–176.
- Ohala, J. J. (1992). "The segment: Primitive or derived?," in *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, edited by G. J. Docherty and D. Robert Ladd (Cambridge University Press, Cambridge, UK).
- Öhman, S. E. G. (1966). "Coarticulation in VCV utterances: Spectrographic measurements," *J. Acoust. Soc. Am.* **39**(1), 151–168.
- Poupplier, M. (2012). "The gestural approach to syllable structure: Universal, language- and cluster-specific aspects," in *Speech Planning and Dynamics*, edited by S. Fuchs, M. Weirich, D. Pape, and P. Perrier (Peter Lang Publishing Group, Frankfurt am Main), pp. 63–96.
- Pouw, W., Paxton, A., Harrison, S. J., and Dixon, J. A. (2020). "Acoustic information about upper limb movement in voicing," *Proc. Natl. Acad. Sci. U.S.A.* **117**(21), 11364–11367.
- R Core Team (2015). "R: A language and environment for statistical computing [computer program]" (R Foundation for Statistical Computing, Vienna, Austria), available at <http://www.R-project.org/> (Last viewed 17 February 2023).
- Stevens, K. N., and Blumstein, S. E. (1978). "Invariant cues for place of articulation in stop consonants," *J. Acoust. Soc. Am.* **64**(5), 1358–1368.
- Stevens, K. N., and House, A. S. (1955). "Development of a quantitative description of vowel articulation," *J. Acoust. Soc. Am.* **27**(3), 484–493.
- Strange, W. (1987). "Information for vowels in formant transitions," *J. Mem. Lang.* **26**(5), 550–557.
- Svensson Lundmark, M. (2020). "Articulation in time: Some word-initial segments in Swedish," Lunds Universitet, available at https://portal.research.lu.se/ws/files/84061604/Malin_Svensson_Lundmark_Articulation_ii_time.pdf (Last viewed 17 February 2023).
- Svensson Lundmark, M. (2022a). "Rapid movements at segment boundaries – preliminary reports on manner," in *Proceedings of FONETIK 2022* (Quarterly Progress and Status Report, Dept. for Speech, Music and Hearing, KTH, Stockholm), Vol. 2(2).
- Svensson Lundmark, M. (2022b). "Evidence of segmental articulations: Acceleration determines vowel segment duration in Swedish word accents," in *Proceedings of 1st International Conference of Tone and Intonation (TAI 2021)* SDU, Sønderborg.
- Svensson Lundmark, M., Ambrazaitis, G., and Ewald, O. (2017). "Exploring multidimensionality: Acoustic and articulatory correlates of Swedish word accents," in *Proceedings of Interspeech 2017*, Stockholm, pp. 3236–3240.
- Svensson Lundmark, M., and Frid, J. (2022). "Segmental articulations across prosodic levels," in *Proceedings of the 13th International Conference of Nordic Prosody* (Scienco, Warsaw).
- Svensson Lundmark, M., Frid, J., Schötz, S., and Ambrazaitis, G. (2021). "Word-initial consonant-vowel coordination in a lexical pitch-accent language," *Phonetica* **78**(5–6), 515–569.
- Turco, G., and Braun, B. (2016). "An acoustic study on non-local anticipatory effects of Italian length contrast," *J. Acoust. Soc. Am.* **140**(4), 2247–2256.
- Ushijima, T., and Hirose, H. (1974). "Electromyographic study of the velum during speech," *J. Phon.* **2**(4), 315–326.
- Wieling, M., and Tiede, M. (2017). "Quantitative identification of dialect-specific articulatory settings," *J. Acoust. Soc. Am.* **142**(1), 389–394.
- Wood, S. (1979). "A radiographic analysis of constriction locations for vowels," *J. Phon.* **7**(1), 25–43.
- Xu, Y. (2013). "ProsodyPro—A tool for large-scale systematic prosody analysis. 4," in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, edited by B. Bigi and D. Hirst (Laboratoire Parole et Langage, Aix-en-Provence, France), pp. 7–10, available at http://www2.lpl-aix.fr/~trasp/Proceedings/TRASP2013_proceedings.pdf (Last viewed 17 February 2023).
- Zsiga, E. C. (1994). "Acoustic evidence for gestural overlap in consonant sequences," *J. Phon.* **22**(2), 121–140.