Research Article

Malin Svensson Lundmark, Johan Frid, Gilbert Ambrazaitis*
and Susanne Schötz

Word-initial consonant–vowel coordination in a lexical pitch-accent language

<https://doi.org/10.1515/phon-2021-2014>

Published online November 17, 2021

Abstract: Previous research has acknowledged the effect of prosody on inter-gestural coordination, but specifically the effect of tones is still understudied. This paper has a two-fold purpose. First, it aims to explore effects of the Swedish word accents on word-initial consonant–vowel (CV) coarticulation. Second, it aims to revisit the existing evidence for tonal integration. Based on Articulatory Phonology, it has been suggested that tones – in tone languages – are integrated in the gestural organization of a syllable-initial CV sequence in the same manner as would an additional consonant (CCV), indicated by a time lag between the gestural onsets of the C and the V gesture (CV onset time lag). However, we argue that the existing evidence is inconclusive, because previous cross-linguistic research has used small-scale data sets (one to seven speakers), and we still lack a well-grounded consensus on how gestural onsets are to be measured. This study uses Electromagnetic Articulography (EMA) to investigate word-initial CV coordination in a lexical pitch-accent language (Swedish) with a binary tonal word accent distinction: a tonal fall and a tonal rise, respectively. A selection of 13 spatial, temporal or coordinative measures of bilabial and tongue body data from 19 speakers, and acoustic f_0 data, were examined to study the CV sequence /ma/. Mixed effects regression models revealed a longer tongue body movement in the rising tone context and small but significant differences in tongue body height, in the closing and the opening of the lips, as well as in the CV onset time lag between the two tonal contexts. We argue that these effects are biomechanical in nature, due to the physiological connections between the tongue, the jaw, and the larynx.

***Corresponding author: Gilbert Ambrazaitis**, Department of Swedish, Linnaeus University, Växjö, Sweden, E-mail: Gilbert.Ambrazaitis@lnu.se

Malin Svensson Lundmark, Centre for Languages and Literature, Lund University, Lund, Sweden
Johan Frid, Lund University Humanities Lab, Lund University, Lund, Sweden

Susanne Schötz, Logopedics, Phoniatrics and Audiology, Clinical Sciences, Lund University, Lund, Sweden

In addition, our results suggest either synchronized CV onsets or a CV onset time lag (as in tone languages), depending on the timing landmarks used. In order to evaluate such results as evidence for or against the integration of tone in CV coarticulation, we argue that future research needs to compare data from a variety of languages using a considerable number of speakers. The present study provides new reference values for such comparisons.

Keywords: articulation; articulatory phonology; c-center; coarticulation; EMA; lexical pitch accent; lexical tone; speech production; tone gesture; tongue-pull hypothesis

1 Introduction

This study investigates how the articulation of the word-initial consonant–vowel (CV) sequence /ma/ is affected by the tonal context, using Electromagnetic Articulography (EMA). To this end, the two tonal conditions offered by the pitch-accent language Swedish (Accent 1 and Accent 2) are compared. In Swedish, the binary tonal word-accent distinction is morphophonological, and – in the southern dialect chosen for this study – the tones are executed on the stressed syllable by means of a fall in the fundamental frequency (f_o)¹ in Accent 1 and a rising f_o movement in Accent 2.

Previous cross-linguistic articulatory work has aimed to understand how consonant–vowel coarticulation varies as a function of suprasegmental features and phonotactics. For instance, it has been shown that CV sequences exhibit a greater degree of coarticulation, and more phonetic change, than VC sequences (e.g., Fant 1969; Löfqvist and Gracco 1999; Macneilage and Declerk 1969; Recasens 2002; Stevens and Blumstein 1981), and articulatory coordination varies depending on the types of consonants and vowels involved (e.g., Fowler and Saltzman 1993; Kawahara et al. 2014; Löfqvist and Gracco 1999; Mooshammer et al. 2006; Recasens 2002).

Effects of prosody, both concerning its grouping and its prominence functions, have also been studied within this line of research (e.g., Byrd et al. 2005; Cho 2002; de Jong et al. 1993; Lindblom et al. 2007; Mooshammer et al. 2013). Recent

¹ We have adopted the nomenclature suggested by Titze et al. (2015), where the fundamental frequency is represented by a small letter f and a small letter “o” (for oscillation) as a subscript. This notation is widely used already and, they argue, more meaningful than the also widely used subscript “0”, as there is no such thing as a zeroth harmonic (see Titze et al. 2015, for a full explanation of their suggested nomenclature, also in relation to the notation of resonances and formants).

contributions to the development of this field include work on lexical tones in tone languages, suggesting effects of tone on lip and tongue body coordination (Gao 2008; Hu 2016; Karlin and Tilsen 2015; Zhang et al. 2019), as well as differences in tongue body height between the tones (Erickson et al. 2004; Hoole and Hu 2004; Shaw et al. 2016).

The few existing relevant studies on the relation between articulatory movements and prosody have so far been based on small-scale data sets involving one to seven speakers (e.g., Byrd et al. 2005; Cho 2002; Erickson et al. 2004; Erickson et al. 2012; Gao 2008; Hoole and Hu 2004; Hu 2016; Karlin and Tilsen 2015; Lindblom et al. 2007; Mooshammer et al. 2013; Mücke et al. 2012; Niemann et al. 2011; Shaw et al. 2016; Zhang et al. 2019). Small data sets, however, might constitute a serious limitation, since inter-gestural coordination has been shown to be speaker-dependent (e.g., Gao 2008; Löfqvist and Gracco 1999; Zhang et al. 2019). In addition, previous studies have used a variety of measures (see 1.5 below), making cross-linguistic comparisons, or comparisons between studies, difficult.

The overall goal of the present study is therefore to further explore the effect of prosody on inter-gestural coordination by (i) including a considerable number of speakers, (ii) replicating a range of measures that have been used in previous studies, and (iii) extending the field to a previously neglected type of prosodic condition (tones in a pitch-accent language). To this end, we are investigating the effect of the Swedish word accent as a binary categorical distinction on lip and tongue body coordination in the word-initial CV sequence /ma/. The falling and rising tones provide us with two divergent tonal environments, hence two tonal categories. Our analysis is based on kinematic data on bilabial and tongue body movements and acoustic f_0 data from 19 speakers.

The study has a two-fold purpose. One aim is to explore articulatory differences between the two tonal categories. The other aim is to revisit the idea of tonal integration in CV coordination (see 1.2), as suggested by Gao (2008). To this end, we discuss the existing evidence for tonal integration in the light of our own results, with a particular focus on methodological questions. Our data, we claim, provide new reference values for a continued discussion. To the best of our knowledge, the study is the first large-scale articulatory study (in terms of measures and participants) that uses EMA to examine CV coordination in a lexical pitch-accent language.

1.1 Sources of variation in CV coordination

The production of consonants and vowels can be understood as a process in which phonetic gestures are implemented through a coordination of articulators (see

Fowler and Saltzmann 1993, for a more detailed discussion). In the production of a sequence of phonemes, this coordination involves a temporal overlap (or coproduction) of phonetic gestures, which results in anticipatory and carryover coarticulatory effects.

An influential study by Öhman from 1966 revealed how the coarticulation of stop consonants and vowels is the result of overlapping articulatory gestures. In a VCV sequence, the two vowels can be understood as one basic diphthongal gesture, on which the consonantal gesture is superimposed (Öhman 1966). This overlapping of consonants and vowels has later been shown to differ depending on syllabic position. A CVC structure, for instance, displays more coarticulation in the CV than in the VC sequence (Löfqvist and Gracco 1999; Macneilage and Declerk 1969; Recasens 2002). Löfqvist and Gracco (1999) were able to demonstrate that the consonant and the second vowel in a VCV sequence are roughly synchronous with each other, with the vowel gesture onset often even preceding the onset of the consonantal gesture, depending on the vowels involved. Hence, the timing of the gestures in the CV sequence is also sensitive to phonotactic information. Different consonants do indeed exhibit different overlapping patterns with the vowel (Fowler and Saltzman 1993; Mooshammer et al. 2006; Recasens 2002), in addition to being dependent on syllable structure (Byrd 1996; Byrd et al. 2005; Nittrouer et al. 1988). In consonant clusters, timing differences with the vocalic gesture have also been observed between different consonant combinations (Bombien et al. 2013; Marin 2013; Pouplier 2012). Finally, tendencies towards language-specific patterns of CV coordination have also been discussed (e.g., Fowler and Saltzman 1993; Marin 2013; Smith 1995).

1.2 The effect of tone on CV coordination

While the effect of phonotactics on CV coordination is well documented, an area that is still understudied is the effect of tone on the articulators' movements. A couple of studies have, however, revealed both spatial and temporal effects of tone (e.g., Erickson et al. 2004; Gao 2008; Honda et al. 1999; Shaw et al. 2016).

A major contribution to this line of work can be found within the Articulatory Phonology (AP) framework. The framework proposes articulatory gestures as phonological units (Browman and Goldstein 1992), and since the pioneering work of Gao (2008) *tone gestures* are assumed as a phonological representation of tone and of the coordinated articulatory actions to achieve the tonal task goal. Tone gestures are comparable to consonantal and vocalic gestures, in that they seem to be integrated in the “global organization” (Browman and Goldstein 1988) that had previously been observed for consonantal gestures in syllable-initial position: As

shown by Gao (2008), the presence of a tone gesture may have a similar effect on the coordination of the CV pattern as the presence of an additional consonant (as in CCV) would have, as detailed in the following paragraphs. Notably, however, such effects have so far not been observed for non-lexical tones (i.e. intonational tones, see Mücke et al. 2012; Niemann et al. 2011).

The assumption of a global organization of gestures is based on articulatory and acoustic evidences from syllable-initial consonant clusters (i.e. syllable onsets in CCV or CCCV) (Browman and Goldstein 1988; Byrd 1995), showing, for instance, that acoustic vowel duration depends on the number of consonants in the syllable onset, but not on the number of consonants in coda position. This was explained in terms of a *global* organization in syllable onset, where all consonants are coordinated with the subsequent vowel (instead of only with each other – which could be regarded as a *local* organization). It was further assumed that it is the temporal center of a consonant's constriction (the so-called *c-center*) that is coordinated with the vowel. In a cluster, then, a “c-center effect” occurs, meaning that the consonants are shifted away from each other and that the c-center of the entire cluster (i.e. the mean of the individual c-centers) is coordinated with the vowel. This c-center effect has subsequently been modeled with reference to gestural coupling relationships dependent on syllable position (Nam et al. 2009). According to this account, the consonantal gesture in a CV sequence in the syllable onset is *in-phase* coupled with the vocalic gesture. When an additional consonantal gesture is present (as in C₁C₂V), however, the two consonantal gestures are competitively coupled, as they are both *in-phase* coupled with the V and *anti-phase* coupled with each other. This competitive coupling forces the consonants away from each other, giving rise to the fact that the consonants are recoverable (i.e. that they appear in sequence) and that the subsequent vowel is shortened, since the overlap with the rightmost C increases (Browman and Goldstein 1988; see also Browman and Goldstein 2000; Gafos 2002). Consequently, the c-center effect also predicts, for all consonants in a cluster, that the time lag between the consonantal gesture and the vocalic gesture (CV time lag) will differ from the CV time lag in a simple CV syllable onset.

Within this theoretical framework Gao (2008) conducted laboratory work on the four tones of Mandarin (Tone 1: high tone, HH; Tone 2: rising tone, LH; Tone 3: low tone, LL; Tone 4: falling tone, HL), measuring the time lag between the gestural onsets of the consonant (C), the vowel (V) and the tone (T). Gao found that the time lag between the onset of the bilabial closing (C) and the onset of the tongue body movement (V) in a /ma/ sequence was approximately 45 ms, and likewise, the time lag was approximately 45 ms (30–60 ms for Tone 1, 2, and 3) between the tongue body and the tone gesture onset (VT time lag), where T was measured in the f_0 curve (although there was no conclusive result for Tone 4). In other words, the onsets of the consonantal gesture and the tone gesture were timed roughly

symmetrically around the onset of the vocalic gesture (C starting 45 ms before, T starting 45 ms after the V onset), which is different from the established pattern attested for single-C syllable onsets (CV) in non-tonal languages, but rather comparable to the attested pattern for consonant clusters (CCV) (e.g., Browman and Goldstein 1988; Marin 2013). The time lags were interpreted as the tones being subject to the c-center effect just as consonants are (Gao 2008).

Following Gao's (2008) example, studies on additional tone languages have been conducted, arriving at very similar conclusions (Karlin and Tilsen 2015, on Thai; Hu 2016, on Lhasa Tibetan). Gao's (2008) study has also been replicated for Catalan and German pitch accents with no time lag between C and V onset found (Mücke et al. 2012). Although the onset of the tone gesture was substantially delayed (~100 ms) for the German speaker, the CV gestures were synchronous with each other. A similar study on three German (and two Italian) speakers found similar patterns (Niemann et al. 2011).

The synchronized gestural CV onsets in non-tonal languages (German, Catalan, Italian) have been interpreted by the authors as a result of the absence of lexical tones in these languages (Niemann et al. 2011; Mücke et al. 2012). However, in order to clarify whether the synchronization differences between tonal and non-tonal languages are due to the presence versus absence of tone, rather than other language-specific parameters, a baseline in the experimental designs would be required, such as a non-tonal condition in the Mandarin data. Zhang et al. (2019) recently made a first attempt to solve this issue by comparing a full tonal condition with a reduced tonal and a non-tonal condition in Mandarin. They found CV time lags in the tonal condition, but no CV time lags in the non-tonal condition. However, they measured gestural targets, rather than onsets. For gestural onsets, they observed, "a high degree of temporal variability" (Zhang et al. 2019).

Moreover, although the proposition of different gestural patterns for tonal and non-tonal languages might be theoretically plausible (CV time lags for lexical tones, no CV time lags for the non-tonal condition), the studies mentioned have used different timing landmarks as well as a small number of speakers (seven in Gao 2008, one from each language in Mücke et al. 2012, two or three, respectively, for the two languages in Niemann et al. 2011, four in Karlin and Tilsen 2015, and three in Hu 2016) which might be insufficient considering speaker variability (see, e.g., Löfqvist and Gracco 1999; Gao 2008; Zhang et al. 2019). As an intermediate conclusion, we still lack convincing evidence for tone gestures that would be integrated in the global organization of CV articulation.

Other articulatory studies on Mandarin tones have observed differences in tongue body movements between the tones. For instance, Shaw et al. (2016), studying spatial tongue data, found a lower tongue body in /a/ for the tones with initial low tone context (Tone 2 and Tone 3) than in the initial high tone contexts

(Tone 1 and Tone 4). A similar effect of tone was also observed in another study on Mandarin comparing Tone 1 (a high tone) and Tone 3 (a low tone) (Erickson et al. 2004). The tongue was more retracted for Tone 3, and for one of the speakers, also lower. In addition, there was also an effect of tone height on the jaw movement in both of these studies (Erickson et al. 2004; Shaw et al. 2016). As has been shown in a study using Magnetic Resonance Imaging (MRI) by Honda et al. (1999), low tones encompass lower larynx movements, with the jaw moving downwards as well, while higher tones display the jaw moving backward as the high tone falls.

The cited studies on tongue and jaw positions or movement patterns suggest a link between f_0 and the articulators, reminiscent of the well documented case of intrinsic f_0 in vowels (e.g., Lehiste and Peterson 1961; Peterson and Barney 1952; Reinholt Petersen 1978): low-tongue vowels have a lower intrinsic f_0 than high vowels; and likewise, low tones seem to trigger a lower jaw/tongue height than high tones (e.g., Erickson et al. 2004; Shaw et al. 2016). It has been suggested that this connection between f_0 and the articulators can be explained by the physiological (muscular) connections between the tongue (the hyoid bone), the larynx (more specifically, the cricothyroid joint) and the jaw (Erickson et al. 2017). The connection between the larynx and the hyoid had already been discussed in connection with the tongue-pull hypothesis in the early studies on vowel-intrinsic f_0 (e.g., Lehiste 1970; for an overview see Ohala and Eukel 1987), and more recent studies suggest an active role of the jaw, too (Chen et al. 2019; Erickson 2011), which we further elaborate on in the discussion (4.1.1).

To conclude, previous studies suggest that the anticipatory movements of CV coarticulation, as well as their spatial and temporal coordination, may be affected by the presence of phonological tones, and differently by different tones. These effects might be due to a phonologically determined coupling of the tone with the consonant and the vowel, or through physiological mechanisms. The present study extends these lines of research to a pitch-accent language, that is, another type of usage of phonological tone, exemplified by Swedish.

1.3 Tone in Swedish

Swedish has a binary tonal word accent distinction (Accent 1, Accent 2, henceforth A1 and A2) with tonal phonological representations and phonetic realizations that differ largely between dialects (Bruce 2005; Bruce and Gårding 1978). In this study, we examine South Swedish, where A1 is represented by an initial high tone at vowel onset (a falling tone, H*L, in the stressed syllable), and A2 by an initial low tone at vowel onset (a rising tone, L*H) (Figure 1). Swedish word accent is

morphophonological: Words with final lexical stress (which includes all monosyllabic words) always bear A1, while in polysyllabic words with non-final stress, both accents can appear. In these words, they can have either lexical or morphological uses, as they are, for instance, associated with different suffixes. Phonologically, however, the falling and the rising tone are both associated with the stressed syllable (located in the stem, not in the suffix), which results in an early f_0 peak in A1, and a late f_0 peak in A2. Previous studies have shown that the early and the late timing of the f_0 peaks are consistent, and perceptually important (Ambrazaitis and Bruce 2006; Ambrazaitis et al. 2012; Bruce 2005; Bruce and Gårding 1978; Svensson 2014).

Details of the co-production of the Swedish tones and the CV sequence have been hardly addressed in the past, and the few exceptions have been based on acoustic data. Early studies by Öhman (1965) and Löfqvist (1975), have, for instance, studied f_0 contours in varying segmental conditions aiming to establish invariant f_0 properties of the word accents (Öhman 1965), or, in other words, to disentangle intrinsic from extrinsic f_0 variations (Löfqvist 1975). Other acoustic studies have revealed durational effects of the Swedish word accents (Elert 1964; Svensson Lundmark et al. 2017). Another more recent example is a study on tonal-segmental alignment in South Swedish (Svensson 2014). Articulatory measurements in connection with the Swedish word accents have, to the best of our knowledge, been restricted to investigations of laryngeal control (e.g., Gårding et al. 1975; Öhman et al. 1967). Little is known about whether or how the articulatory movements of consonants and vowels in Swedish are affected by the tones of the word accents.

Figure 1: Mean f_0 contours of South Swedish speakers in Accent 1 and Accent 2 target words: short vowels (a); long vowels (b). The normalized time scale indicates the points of measurement (10 temporally equidistant measurements per segment); the vertical lines indicate segment boundaries.

1.4 Scope, limitations and predictions

Previous articulatory studies on tones suggest that tones are integrated with the articulation of the consonant and the vowels in a CV syllable onset (see 1.2). This integration is evidenced by time lags between the gestural onsets of C and V in tonal languages, while previous results on non-tonal languages suggest a synchronicity of the C and V onsets. Within Articulatory Phonology, such findings have been interpreted as evidence of a tone gesture that is coupled with the consonantal and the vocalic gesture in much the same way as the two consonantal gestures and the vocalic gesture are coupled in a CCV cluster. Since the lexical pitch accents of Swedish are understood as consisting of tones,² we would expect that their hypothetical tone gestures are coupled with the consonantal and vocalic gesture in a CV syllable onset, just as in a tone language. Hence, for the present study, we would predict time lags between the consonantal and the vocalic gesture onsets in a CV syllable onset. However, as discussed in 1.2, we cannot consider the existing evidence on tonal and non-tonal languages as entirely conclusive, which means that the basis for our prediction is somewhat unstable. Consequently, the present study will not be entirely conclusive either as to whether or not Swedish tones behave more like Mandarin lexical or German post-lexical tones. The main purpose of our coordinative measurements is, therefore to produce new reference values for time lags as measured with a variety of metrics, based on a considerable number of speakers.

Furthermore, we might expect to see differences in tongue body movements between A1 and A2, in line with the findings for Mandarin tones by Erickson et al. (2004), Hoole and Hu (2004), and Shaw et al. (2016). Therefore, in this study we focus on the effect of the tonal context on the lips and the tongue body and present spatio-temporal data on 1) the bilabial closing and opening during the word-initial consonant, 2) the tongue body during the lowering and retraction of the vocalic gesture, and 3) the time lags between these separate articulator movements. To this end, we compare the measures just mentioned in two conditions: /ma/ word onsets in words with A1 versus words with A2.

² In fact, the lexical status of the Swedish tones has been debated for decades (for an overview, see Ambrazaitis 2009). The most common view has been that only Accent 2 is “marked” and involves a (lexical) tone, but the distinction has also been conceived of as equipollent (Bruce 1977), with phonological (although not necessarily lexical) tones for both A1 and A2, and finally, it has been argued that Accent 1, rather than Accent 2, is the marked case (Lahiri et al. 2005). Because of this persistent disagreement, we cannot confidently claim that Swedish would provide a reliable case for comparing a tonal and a non-tonal condition.

The scope of this study is to take this first step of testing whether CV coordination is affected by the tone (T). A subsequent step, not taken in this study, could then be to establish a measure of the onset of a tonal gesture and to include it in the study of temporal coordination, measuring time lags between C, V, and T.

Initial attempts of including T have been made in the studies by Gao (2008) and Mücke et al. (2012) (but see the discussion in 4.4 below). The present material, however, would not allow us to present results on V and T time lags in comparison to the studies by Gao (2008) and Mücke et al. (2012). The initial low tone gesture of A2 presupposes a preceding fall from a high tone, a constellation that is not possible to produce in South Swedish without inserting a phrase boundary.

Therefore, instead of measuring tonal alignment as has been done in Gao (2008) and Mücke et al. (2012), we undertake a different approach to the f_0 data. As already mentioned, we treat the word accents as two tonal categories (as has been similarly done for Mandarin in Shaw et al. 2016) and present a visual representation of the overall f_0 pattern as well as an analysis of f_0 differences between the word accents in the most relevant region of interest, that is during the word-initial consonant /m/.

1.5 Methodological considerations

Previous related studies cited above have made use of a variety of methodological choices, which are not always compatible and to some extent inhibit direct comparisons. In the present study, we aim to avoid this problem by opting to include a variety of alternative measures enabling such comparisons. In presenting our choice of measures, we give motivations for why we chose these measures, with Figures 2 and 3 providing visual explanations of them.

The EMA data was obtained from speakers pronouncing A1 and A2 words with onsets containing a nasal bilabial consonant and an open vowel (/ma/). Most previous articulatory studies have included this specific CV sequence, which makes it useful for comparisons. For example, the dataset in the dissertation by Gao (2008) included bilabial and alveolar nasal consonantal onsets, as well as voiceless onsets, followed by an open vowel. Mücke et al. (2012), in their study of Catalan and German, compared bilabial and alveolar onsets followed by open vowels, in both open and closed syllables.

Central for this study is the measurement of gestural onset-to-onset (CV) time lags, that is, the temporal distance (in ms) between the onset of the bilabial closing and the onset of the tongue body movement (henceforth also the “CV (onset) time lag”). Both Gao (2008) and Mücke et al. (2012) based their measurements of CV time lags on the vertical movements of the tongue body (for V) and on the calculated lip

Figure 2: An illustration of the 13 spatial, temporal, or coordinative measurements on the bilabial (Lip aperture) and tongue body data of the target CV sequence /ma/. The figure is a stylized visualization of the approximate temporal location of each measurement landmark. Some of the measurements occur already during the previous vowel. (* significant differences between the word accents found after the analysis). See Figure 3 for more detail.

aperture (distance between lip sensors) for C. However, while Gao (2008) labelled the C and V landmarks on the velocity curve at a 20% threshold from the zero-crossing to peak velocity, in Mücke et al. (2012) the corresponding landmarks were instead placed at the zero-crossings (also on the velocity curve), that is, when the

Figure 3: A detailed visualization of the 13 spatial, temporal, or coordinative measurements. Lip landmarks are based on the first derivative (velocity) and second derivative (acceleration) of the lip aperture calculation. Tongue body landmarks are based on the first derivative of the actual TB sensor movement (vertical and tangential velocity).

vertical movement was minimal. It is unclear how or if the type of landmark affects the duration of the time lags. However, a threshold mark of 20% from zero crossing velocity might be preferable, since, around the zero-crossing in the velocity curve,

a plateau may occur in the vertical tongue body trace, creating several possible landmarks (see, e.g., Blackwood Ximenes et al. 2017). In addition, in previous studies, the perhaps more stable 20% threshold mark has been found to be advantageous for capturing an on-going movement (Kroos et al. 1997; Mooshammer et al. 2006). We replicated and compared these two measurements in our study of Swedish tones.

Even though Löfqvist and Gracco (1999) did not include tone in their study on the coarticulation of VCV sequences, the methodology is of potential interest for our study. They captured the dynamic tongue body movements using data on both tongue height and tongue frontness (i.e. both the vertical and the horizontal dimension) by calculating the tangential velocity. Tangential velocity, $[v = \sqrt{(\dot{x}^2 + \dot{y}^2)}]$, where x and y indicate the horizontal and vertical velocities, respectively, captures the dynamic tongue body movements and combines the front-back and the high-low dimensions, and has been used to define measurement landmarks in several studies (e.g., Blackwood Ximenes et al. 2017; Gafos et al. 2020; Löfqvist and Gracco 1999; Shaw et al. 2009). Furthermore, since they investigated several vowel types, they found that a lip aperture landmark based on acceleration, that is, the second derivative of the lip movement, was the most stable across the vowels. We believe that their methodology could add to the comparison with Gao (2008) and Mücke et al. (2012) since it will provide us with a wider account of the dynamics of the lips and tongue body movement.

Erickson et al. (2004), in their study on Tone 1 and Tone 3, examined spatial dimensions of vowels using the vertical and the horizontal position (measured separately) of a sensor on the tongue dorsum at a time when the jaw was at its lowest point. This landmark reflects the target of the tongue body gesture. In order to be able to compare our spatial tongue body results on the Swedish word accents to the study by Erickson et al. (2004), we include a measure that aims to capture tongue height at the vowel target. However, since we do not include jaw data in the present study, we cannot replicate Erickson et al.'s (2004) jaw-based landmark. Instead, we base our measure of tongue height at vowel target on the tangential velocity curve, which considers both vertical and horizontal tongue position, as Erickson et al. (2004) saw an effect of tone on both spatial dimensions. Using the tangential velocity curve for capturing the vowel *target* is also motivated for reasons of consistency, as we already use it for capturing the vocalic gestural *onset* in the CV onset time lag measure by Löfqvist and Gracco (1999).

In addition, to gain a more comprehensive picture of the tongue movement, we measure tongue height not only at the gestural target, but also at the gestural onset, and at a landmark in between onset and target. All three landmarks were defined on the tangential velocity curve. The latter landmark was the maximum

tangential velocity intervening between the two minima at the onset and the target (Figures 2 and 3). We found this landmark to be a stable one reflecting a point in time during the lowering and retraction of the tongue. Note that the first minimum, i.e. the onset landmark, is the same landmark that was used to measure the CV onset time lag according to Löfqvist and Gracco (1999).

Since we have previously noticed that the tongue body movement differs between A1 and A2 (Svensson Lundmark et al. 2017), we also include a measurement of the duration of this movement: the *palatal wide interval*. This term refers to the time it takes for the tongue body to reach its target (i.e. the temporal distance from onset to target as defined above).

We also add bilabial interval measurements of the lips in order to gain a more comprehensive picture of the CV coordination to support the evaluation of the CV time lag measurements. Accordingly, “bilabial interval” refers to measured intervals for the time it takes the lips to close and to open again. Both intervals are measured in three ways: Since we are including three landmarks to measure the bilabial closing onsets, based on previous studies (see above), we also use the same landmarks to measure the closing target, and for the sake of consistency, we use the same three types of landmarks for the bilabial opening interval. The bilabial intervals are thus either based on when the movement of the lips is minimal (zero-crossing in the velocity curve), based on when the movement is on the way or has slowed down (the 20% peak velocity threshold), and based on when the lips accelerate and decelerate most (maxima and minima in the acceleration curve).

These considerations result in a selection of 13 measures to be included in the present study. An overview is presented in Tables 1 and 2.

2 Materials and methods

2.1 Speakers

Twenty speakers (12 female) of the South Swedish dialect of Scania (the southernmost region of Sweden) participated in the EMA recordings. The speakers' ages ranged from 23 to 75 ($\bar{x} = 40$ years, $sd = 12.3$ years). The participants were unaware of the purpose of the study. All speakers grew up with South Swedish speaking parents. However, four of the speakers (20%) grew up with one parent with a first language other than Swedish, and one speaker with one parent with another Swedish dialect.

Table 1: The different types of measurements and their relation to gestural landmarks in Articulatory Phonology (Browman and Goldstein 1988, 1992). Italicized terms according to Gafos (2002).

Type of measurement	Gestural landmarks or temporal relations estimated	Articulators involved	Measurement label
Temporal	Consonantal <i>onset</i> to <i>target</i>	Lips (lip aperture)	Bilabial closing interval
	Consonantal <i>release</i> to <i>release offset</i>	Lips (lip aperture)	Bilabial opening interval
	Vocalic <i>onset</i> to <i>target</i>	Tongue body	Palatal wide interval
Spatial	Vocalic <i>onset</i> , vocalic <i>target</i> , and a point in between onset and target	Tongue body	Tongue body height
Coordinative	<i>Onset</i> phasing of consonantal and vocalic gestures	Lips (lip aperture) and tongue body	CV onset time lag

Note that in AP, the term *landmark* is used at a theoretical level, denoting the structurally relevant points in time during the course of a gesture: the *onset*, the *target*, the *c-center*, the *release*, and the *release offset* (Gafos 2002). Landmarks in this sense are thus abstract points in time. In this study, we are estimating some of these abstract landmarks using a variety of concrete points in time, which we, in turn, also refer to as *landmarks*. In other words, we estimate a number of abstract landmarks using various concrete *measurement landmarks*, or henceforth simply *landmarks*.

2.2 Speech material

The original data set consisted of 18 target words, divided into nine word accent pairs with identical stress pattern, that is, stress on the initial syllable (e.g., *bilen-bilar* or *manen-manar*). For this study, we only used the four word accent pairs that shared the same word-initial CV sequence /ma/. Hence, the material consists of eight disyllabic target words, four with A1 and four with A2, with phonologically similar word onsets, namely either the target CV sequence /ma/ ([ma]) or /ma:/ ([ma:]). Both comprise a nasal bilabial closure and a palatal wide tongue body movement, but the vowels differ in quality and quantity, since in Swedish, a difference in vowel quantity also implies a difference in vowel quality, especially in the case of open vowels. Thus, in the long vowel [a:] the tongue is significantly retracted compared to the short vowel [a] (Bruce and Engstrand 2006). This is also clearly seen in our data (see Figure 9b below). Moreover, in the case of the short vowel, the target CV sequence is followed by a coda: either a nasal alveolar, /'CVn.nV(C)/, or a nasal bilabial, /'CVm.mV(C)/. If the vowel is long it is followed by a nasal alveolar, /'CV:.nVC/, or a lateral alveolar, /'CV:.lVC/, in post-syllabic position (Table 3). Hence, the short condition implies a closed syllable, while the long condition signifies an open syllable.

Table 2: The different measures and their calculations based on the assortment of articulatory landmarks.¹ = the local *minimum* in the acceleration curve (see Footnote 3 for an explanation).

Measurement	Calculation	Lip aperture landmark	Tongue body landmark	Dependent variable
Bilabial closing interval	LA target – LA onset	zero velocity	–	C_close_Ovel
		20% zero velocity	–	C_close_20vel
		max. acceleration ¹ /deceleration	–	C_close_acc
		zero velocity	–	C_open_Ovel
		20% zero velocity	–	C_open_20vel
Bilabial opening interval	LA offset – LA target	max. acceleration ¹ /deceleration	–	C_open_acc
		–	min. tangential velocity	V_interval
		–	min. tangential velocity (onset)	V_height_onset
		–	max. tangential velocity	V_height_peak
		–	min. tangential velocity (target)	V_height_target
Palatal wide interval	TB target – TB onset	zero velocity	zero vertical velocity	CV_timelag_Mückeetal
		20% zero velocity	20% zero vertical velocity	CV_timelag_Gao
		max. acceleration ¹	min. tangential velocity	CV_timelag_L&G
		–	–	–
		–	–	–
Tongue body height	y-score at landmark	–	–	–
		–	–	–
		–	–	–
		–	–	–
		–	–	–
CV onset time lag	TB onset – LA onset	zero velocity	zero vertical velocity	CV_timelag_Mückeetal
		20% zero velocity	20% zero vertical velocity	CV_timelag_Gao
		max. acceleration ¹	min. tangential velocity	CV_timelag_L&G
		–	–	–
		–	–	–

Table 3: Stimuli material arranged according to the conditions *vowel length* and *word accent* (A1, A2). Target words in SMALL CAPS; note that main sentence prominence is elicited on the final word.

		Leading question and target sentence
Short vowel (V)	A1	"Hur rammade Eva mannen? Eva rammade MANNEN med cykeln." <i>How did Eva hit the man?</i> <i>Eva hit the man with her bike.</i>
	A2	"Hur skummade Inger manna? Inger skummade MANNA med vispen." <i>How did Inger foam semolina?</i> <i>Inger foamed semolina with the whip.</i>
	A1	"Med vem kollade Sanna filmen 'Mammut'? Sanna kollade 'MAMMUT' med Nora." <i>With whom did Sanna watch the movie 'Mammoth'?</i> <i>Sanna watched 'Mammoth' with Nora.</i>
	A2	"Var lämnade morfar mamma? Morfar lämnade MAMMA med doktorn." <i>Where did grandpa leave mom?</i> <i>Grandpa left mom with the doctor.</i>
Long vowel (V:)	A1	"Vilken häst kammade Jonas manen på? Jonas kammade MANEN på Ronja." <i>Which horse did Jonas comb the mane of?</i> <i>Jonas combed the mane of Ronja.</i>
	A2	"Vad rimmade Isak 'manar' med? Isak rimmade 'MANAR' med 'svanar'. <i>What did Isak rhyme 'urge' with?</i> <i>Isak rhymed 'urge' with 'swans'.</i>
	A1	"Var lämnade Inger malen hon fångat? Inger lämnade MALEN till kocken." <i>Where did Inger leave the catfish she caught?</i> <i>Inger left the catfish to the chef.</i>
	A2	"Var lämnade Åsa alla malar hon fångat? Åsa lämnade MALAR vid båten. <i>Where did Åsa leave all the catfish she caught?</i> <i>Åsa left catfishes by the boat.</i>

The eight disyllabic target words were embedded in a set of similarly structured target sentences. The VCV sequence preceding the target words was identical across all sentences (/ade/). Each target sentence was preceded by a leading question eliciting a contrastive focus on the last element in the target sentence (Table 3). This left the target word in a low-prominence inducing context, ensuring it would not be associated with a sentence-level prominence. In Central Standard Swedish, sentence-level prominence is typically associated with an additional *f*₀

peak (Bruce 1977). This kind of categorical distinction between higher and lower prominence is absent from the South Swedish dialect (Bruce 2005; Bruce and Gårding 1978). However, higher-level prominence has nevertheless been shown to gradually affect the accentual f_0 pattern even in South Swedish (Ambrazaitis et al. 2012) and, under certain pragmatic conditions, a double-peaked pattern can be observed in this dialect, too (Svensson Lundmark et al. 2017). Therefore, effects of sentence intonation were controlled for in this study by means of the low-prominence inducing context.

2.3 Recording procedure and data preprocessing

Speakers were recorded with the Carstens AG501 Electromagnetic Articulograph (EMA). Data was collected from sensors on the midline of the upper and lower lips, at the vermillion border, and from one sensor placed on the midline of the tongue body. The tongue body sensor, corresponding roughly to the tongue dorsum, was placed on the tongue where the participant made a mark with the upper incisors after having stretched out the tongue as far as possible. The movement of the jaw is integrated with the data on the lips and the tongue and was not treated separately, although we do acknowledge the effect of the jaw on the lip aperture and on the tongue body movements. In order to correct for head movements, we used three additional reference sensors: one behind each ear, and one on the nose ridge. The occlusion plane was not controlled for other than by observing and correcting the speaker's position during the recording. While the vertical tongue body measurements could be affected by the head rotation angle, this was handled in terms of speaker variation in the statistical analysis, where we assumed random intercepts for subjects in a mixed effects model, as well as random slopes when warranted.

The eight sentence pairs (leading questions + target sentences) were read from a computer screen in a random order, each set appearing eight times. Articulatory data was recorded at 250 Hz, and audio was recorded simultaneously using an external condenser microphone (a t. bone EM 9600) at a sampling rate of 48 kHz.

Before analysis, the data were corrected for head movements with the Carstens software, using the three reference sensors mentioned above, and transferred to R (R Core Team 2015). The articulatory data was smoothed using locally weighted regression by the R function *loess*. We used a low span (0.1) so the smoothing should not cause any distortion. This value was determined by testing different span values on a subset of the data and visually inspecting the result.

The first author segmented the acoustic data manually in PRAAT (Boersma and Weenink 2018) into consonants and vowels (C_1 , V_1 , C_2 , V_2 , C_3) using the PRAAT

script ProsodyPro (Xu 2013). In order to avoid unnecessary f_0 analysis errors, f_0 calculation was performed in the time domain based on “pulses” automatically determined by PRAAT, which we manually corrected using ProsodyPro. The script also applies a smoothing algorithm removing minor spikes from f_0 curves (Xu 1999). For the visualization of the tonal patterns of the two word accents (Figure 1), f_0 was time normalized by taking 10 temporally equidistant f_0 measurements for each segment. Finally, these corrected, trimmed, and time normalized f_0 data (in Hz) were converted into semitones (using 100 Hz as a reference value). For the purpose of visual inspection, these pre-processed f_0 data were then used to plot mean f_0 curves across repetitions and speakers, but separated for the eight target words (Figure 1).

For statistical analysis of the f_0 differences during the word-initial /m/, the individual words were time-normalized by z-transforming the time points (i.e. subtracting the mean and dividing by the standard deviation; the mean and standard deviation were obtained on the basis of all time points in the word). Furthermore, f_0 measurements were normalized for each word by z-transforming the measurements per word on the basis of all measurements in the utterance it occurred in.

All speakers but one (who had one parent with another Swedish dialect), displayed the typical South Swedish f_0 pattern of the word accents, with an f_0 fall in the stressed vowel for A1, and a rise in the stressed vowel for A2 (Figure 1). The speaker with the deviant f_0 pattern was excluded from subsequent analysis. With repetitions, and some omissions due to interferences in the signal during the recordings, the data collection ended up consisting of 1191 tokens.

2.4 Measurements

The PRAAT TextGrids were further used in R (R Core Team 2015) to automatically extract articulatory data on lip and tongue body movements from selected time frames of each target word. These time frames were based on acoustic segment boundaries and adjusted manually for each speaker. A constant time frame for all speakers (starting, say, 30 ms before the acoustic segmental consonant onset) would be problematic, because inter-gestural coordination and speech rate vary between speakers. After a visual inspection of the position, velocity and acceleration curves, the first author adjusted the time frames in the script for each individual speaker.

Our 13 articulatory measurements were either temporal, coordinative or spatial (see Tables 1 and 2). The six temporal measurements of the consonantal gestures were based on lip aperture (LA). Lip aperture was calculated in R as the

three-dimensional Euclidian distance between the sensors on the upper and the lower lip. The temporal landmarks on lip aperture were automatically extracted in R from when the lips had minimal movement (zero-crossing in the first derivative, i.e. velocity curve, see Figure 3), when the movement was on its way or when it had slowed down (20% threshold from zero to peak velocity, i.e. to local minimum/maximum in the first derivative, i.e. velocity curve), and when the movement accelerated or decelerated the most (local minima/maxima in the second derivative, i.e. acceleration curve). First and second derivatives were calculated by the R `diff` function with `lag = 2` (R Core Team 2015).

For the three coordinative measurements of consonantal (C) and vocalic (V) gestures, landmarks at the onset of the bilabial closing interval (C) and at the onset of the lowering (and the retraction) of the tongue (V) were used to calculate CV time lags. We have replicated the measurements of CV time lags from Mücke et al. (2012), which are based on the landmarks of zero-crossing in the velocity curve for lip aperture and the vertical tongue body movement; from Gao (2008), based on the 20% threshold from zero-crossing to peak velocity for lip aperture and the vertical tongue body movement; and from Löfqvist and Gracco (1999), based on maximal acceleration (the local minimum in the second derivative, i.e. the acceleration curve)³ of lip aperture and the minimum tangential velocity of the tongue body (Figures 2 and 3).

The tangential velocity of the tongue body was further used to also collect one temporal and three spatial measurements of the vocalic gesture. In order to measure tongue body height while considering both the lowering and the retraction of the tongue, as has been reported in Erickson et al. (2004), landmarks were placed at minimum tangential velocity (onset and target) and at peak tangential velocity of the tongue body movement (Figures 2 and 3). The collected landmarks at the onset and the target of the vocalic gesture (the tongue body movement) were also used to extract temporal data in order to calculate the palatal wide interval.

2.5 Analysis

First, we tested whether f_o differed between the word accents during the word-initial consonant by performing a Welch Two Sample t-test on the f_o data. Secondly, we used linear mixed effects regression models to investigate the effect of

³ Note that we refer to this point in time as the maximal acceleration (of the lip closing movement), despite the fact that it corresponds to a local minimum in the acceleration curve. The polarity of the extremum is negative, because the lip aperture curve goes down.

the main predictor *word accent* on each of the dependent variables (the articulatory measurements).

A mixed effects model is justified since individual speakers' articulation as well as sensor placement on the articulators might vary. Hence, the random effect *speaker* (random intercepts) was included in all models fitted on the measurements (random slopes for speakers were only added when the complexity was warranted, as explained in the next paragraph). Moreover, a qualitative f_0 analysis revealed differences between target words (see the f_0 analysis in 3.1). Some of the target words might have attracted more prominence by some speakers, possibly because the information structure of the corresponding target sentences enabled inserting phrase boundaries or in other ways emphasizing the target words. Consequently, we added *word* as a random effect (random intercept) to avoid target sentence information structure or other factors (e.g., word frequency) increasing the risk of reporting p-values that are too low (a Type I error). Adding *word* as a random effect also accounts for the different post-vocalic consonants included in our material (Table 3; see also the discussion in 4.1.2).

Furthermore, although our main interest was the effect of word accent, the covariate *vowel length* (which in our material corresponds to syllable type), and its interaction with word accent, needed to be controlled for. We performed a likelihood-ratio test for each measurement to test whether the additional complexity of vowel length was warranted. The complexity was only 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, and the difference between the values was at least 2 (following Wieling and Tiede 2017). The same procedure was applied in order to test whether the additional complexity of random *slopes* for *speakers* was warranted (slopes that vary by the accent condition, i.e. the speaker-specific size and direction of the effect of word accent), instead of random *intercepts* only. Hence, the mixed effects regression models included the predictor *word accent* as fixed effect, with the random effects *speaker* (random intercepts) and *word* (random intercepts), and with the optionally added complexity of the fixed effect *vowel length* and/or the random slopes of the random effect *speaker*.

Outliers were excluded by the use of time frames during the automatic data collection in R. In case of negative values for the intervals (the bilabial intervals and the palatal wide interval), these were replaced by missing values and excluded from the regression analysis. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. The mixed 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).

3 Results

3.1 Acoustic results

A qualitative inspection of the mean f_0 contours (Figure 1) shows that A1 starts with a high tone at vowel onset, followed by a fall during the vowel, while A2 starts with a low tone at vowel onset, followed by a rise during the vowel and an f_0 peak at around the syllable boundary. The absolute average f_0 peak can be seen to differ in height between the target words, which is most clearly seen in Figure 1b, where f_0 reaches, on average, slightly higher values in *manen* and *manar*. This might at least partially be related to the observation that individual target words have attracted a higher level of prominence (see 2.5 above).

Next, a statistical analysis was performed to find whether f_0 differed between the word accents already during the word-initial C_1 segment. The mean f_0 of the first segment of the word accents, /m/, differed significantly according to Welch's t -test, $t(12877) = 113.08$, $p < 0.001$. On average, mean f_0 for A1 was 1.16 semitones higher than for A2 (Figure 4). The 95% confidence interval for the effect of word accent on mean f_0 was between 1.14 and 1.18 semitones. In other words, although the falling and the rising tone of the Swedish word accents are executed mostly during the vowel segment, there is a significant difference between the f_0 environments of the word accents already when the tongue body is lowering, that is during the span that includes the majority of our articulatory measurements (see Figure 2 for reference).

3.2 Articulatory results

In this section, we present the results for the 13 articulatory measures (see again Tables 1 and 2; Figures 2 and 3) grouped by the dimension measured, starting with temporal measurements concerning a single gesture at a time (vocalic or consonantal, seven measures; 3.2.1), followed by spatial measurements (only applicable to the vocalic gesture, three measures; 3.2.2.), and finally, temporal measurements of the coordination of vocalic and consonantal gestures (CV time lags, three measures, 3.2.3).

3.2.1 Temporal measurements

The linear mixed effects regression models revealed significant effects of word accent on three of our seven temporal measurements: the duration of the tongue

Figure 4: f_0 in the first consonant, grouped by vowel length and accent. In the left panel, each line shows a polynomial fit through all data points in each subgroup. For esthetic reasons, only a random subsample of the f_0 points are shown (left). The right panel shows the polynomial regression lines of the normalized f_0 data during the word-initial consonant, grouped by word accent and vowel length.

body movement (the palatal wide interval, see 3.2.1.3), as well as one of the three bilabial closing (3.2.1.1), and one of the three bilabial opening intervals (3.2.1.2). The palatal wide interval was found to be significantly longer in A2 than in A1. The same holds for the bilabial opening, while the closing was significantly longer for A1. However, the two significant effects for the bilabial measurements stem from different types of landmarks (velocity-based for the opening, acceleration-based for the closing). In addition, both bilabial effects display a small order of magnitude as measured in ms (<10 ms). These temporal results are presented in greater detail in the following subsections.

3.2.1.1 Bilabial closing intervals

After a model comparison for the bilabial closing measurements, a mixed effects regression model with only *word accent* as fixed effect was fit on the basis of the temporal landmarks from when the lip movement was minimal (C_close_0vel) and from when the movement was on its way or was slowing down (C_close_20vel).

The models did not show any significant differences between the word accents (Table 4). In contrast, for the acceleration-based landmarks (C_close_acc), a more complex model was warranted with the added fixed effect *vowel length*, which showed a significant difference between the word accents ($t = -2.6$, $p < 0.01$), as well as between the vowel lengths ($t = -3.2$, $p < 0.01$): A1 and the long vowels displayed slightly longer intervals than A2 and the short vowels (see Table 4).

Figure 5a visualizes the small but statistically significant differences obtained with the de-/acceleration landmarks (C_close_acc) compared to the non-significant differences with velocity landmarks (middle and left panel in Figure 5a). The figure also displays the fact that the intervals calculated from when the lips de-/accelerated (C_close_acc) and when the movement was on its way/slowed down (C_close_20vel) are much shorter but also less varied than the intervals from when the lip movement is minimal (C_close_0vel). Finally, note that the means obtained for the velocity landmarks suggest an interaction between word accent and vowel length (Figure 5a), which was, however, not supported by the modeling (as vowel length was not even justified as a fixed factor).

3.2.1.2 Bilabial opening intervals

The bilabial opening intervals differ in some regards from the bilabial closing intervals. First, as opposed to the bilabial closing, the opening intervals calculated from the velocity landmarks (C_open_0vel and C_open_20vel) warranted added complexity to the mixed effects regression models. The models with the added *vowel length* as fixed effect showed a clear significant difference between the vowels (C_open_0vel: $t = -3.1$, $p < 0.05$; C_open_20vel: $t = -4.5$, $p < 0.01$): Long vowels displayed longer bilabial opening intervals than short vowels (see Table 4 and Figure 5b, C_open_0vel and C_open_20vel). In addition, the model for C_open_20vel revealed slightly longer intervals for A2 than for A1 ($t = 2.7$, $p < 0.05$).

Second, the model fitted for the de-/acceleration-based interval (C_open_acc) did not warrant adding *vowel length* as fixed effect, and showed no significant difference between the word accents, in contrast with the result on the bilabial closing interval. Note that only one of the bilabial opening measurements showed a significant difference between the word accents (C_open_20vel), although the panels in Figure 5b generally suggest a slight tendency towards longer opening intervals for A2.

3.2.1.3 Palatal wide interval

For the calculated interval of the tongue body lowering and retraction (palatal wide interval, Figure 6) the model comparison warranted the added complexity of *vowel length* as fixed effect. The model fit on the palatal wide interval (V_interval) showed a significant difference between word accents ($t = 2.4$, $p < 0.05$), as well as

Table 4: Results from the mixed effects regression models, with estimates on the effect by word accent and by the addition of interaction with vowel length. * Significant effects. ^R = Random slopes by speaker added to the model (these were only warranted for C_close_Ovel; see text for explanations). The estimates for the intercepts correspond to the mean of the reference conditions (A1 for Word accent; long vowels for Vowel length); the estimates for the effects corresponds to the difference between the conditions; example: the estimated effect of “Word Accent (A2)” is “-0.430” for V_height_target, meaning that the tongue body height is on average 0.4 mm lower (due to the minus sign) in A2 than in A1 (which is the reference condition).

Measurement	Dependent variable	Estimate	SE	df	t-Value	p-Value
Bilabial closing interval (ms)	C_close_Ovel ^R	Intercept	139.671	9.216	22.172	15.155
		Word accent (A2)	-3.581	8.932	7.640	-0.401
	C_close_20vel	Intercept	72.900	6.033	22.740	12.084
		Word accent (A2)	-1.104	5.720	7.494	-0.193
	C_close_acc	Intercept	43.967	1.364	31.407	32.223
		Word accent (A2)	-2.788	1.056	1170.069	-2.641
Bilabial opening interval (ms)	C_open_Ovel	Vowel length (V)	-3.410	1.062	1170.118	-3.211
		Interaction	2.365	1.501	1170.123	1.576
	C_open_20vel	Intercept	132.614	6.629	15.947	20.004
		Word accent (A2)	15.826	7.492	7.463	2.112
		Vowel length (V)	-23.174	7.499	7.490	-3.090
		Interaction	-16.249	10.607	7.494	-1.532
Palatal wide interval (ms)	C_open_20acc	Intercept	69.208	2.453	22.125	28.208
		Word accent (A2)	6.216	2.255	6.917	2.756
	C_open_acc	Vowel length (V)	-10.248	2.274	7.138	-4.507
		Interaction	-5.989	3.214	7.126	-1.863
	V_interval	Intercept	30.187	2.202	19.707	13.712
		Word accent (A2)	3.225	2.257	7.562	1.429
Palatal wide interval (ms)	V_interval	Intercept	204.091	6.628	18.104	30.791
		Word accent (A2)	16.712	7.047	7.339	2.371
	V_interval	Vowel length (V)	-32.583	7.066	7.418	-4.611
		Interaction	-5.884	10.008	7.460	-0.588

Table 4: (continued)

Measurement	Dependent variable	Estimate	SE	df	t-Value	p-Value
Tongue body height (mm)	Intercept	11.178	4.338	19,078	2.557	0.018
	Word accent (A2)	0.081	0.284	7,006	0.285	0.784
	Intercept	9.371	4.222	19,014	2.220	0.039
	Word accent (A2)	-0.717	0.135	1043,000	-5.302	0.000*
	Vowel length (V)	-0.732	0.138	1043,000	-5.310	0.000*
CV time lag (ms)	Interaction	0.973	0.195	1043,001	4.979	0.000*
	Intercept	6.735	4.170	19,026	1.615	0.123
	Word accent (A2)	-0.430	0.173	6,941	-2.483	0.042*
	Intercept	63.421	10.063	19,925	6.303	0.000
	Word accent (A2)	-6.367	10.404	7,593	-0.612	0.558
CV timelag_Gao	Intercept	70.172	8.547	21,801	8.210	0.000
	Word accent (A2)	-4.807	8.358	7,527	-0.575	0.582
CV timelag_L&G	Intercept	2.271	3.399	28,087	0.668	0.510
	Word accent (A2)	-7.908	2.357	1140,109	-3.355	0.001*
	Vowel length (V)	6.460	2.392	1140,342	2.700	0.007*
	Interaction	-0.497	3.377	1140,238	-0.147	0.883

Figure 5: Results of the temporal measurements (in ms) of the consonantal gesture (bilabial intervals): Bilabial closing (a) and bilabial opening (b) as an effect of word accent and vowel length, measured when the lip movement is minimal (C_close_0vel and C_open_0vel), when the movement is on its way or slowing down (C_close_20vel, and C_open_20vel), and when the lips de-/accelerate (C_close_acc and C_open_acc). Error bars represent standard deviations.

between vowel lengths ($t = -4.6$, $p < 0.01$) (see Table 4). The estimated average duration of A1 words with long vowels was 204 ms and the A2 intervals were approximately 17 ms longer than the A1 intervals (in the long vowel condition). Moreover, the interval was 33 ms longer for long vowels than for short vowels (in the A1 condition). There was no significant interaction, and thus, the effect by word accent on the palatal wide interval is found to be independent of vowel length.

3.2.3 Spatial measurements on the tongue body

Figure 7 displays the results for our tongue body height measures. Adding *vowel length* as a fixed effect was justified only for the measurements taken from

Figure 6: Results of the temporal measurements (in ms) of the vocalic gesture: Palatal wide interval (V_interval) as an effect of word accent and vowel length. Error bars represent standard deviations.

Figure 7: Results of the spatial measurements (in mm) of the vocalic gesture: Tongue body height (mean-centered for reasons of clarity) as an effect of word accent and vowel length: at minimum tangential velocity in the gestural onset (V_height_onset), at peak tangential velocity (V_height_peak), and at minimum tangential velocity at the gestural target (V_height_target). The ranges of the y-axes have been manipulated as a means of visual support, such that relative tongue position is roughly indicated by the position of 0 on the y-axis. Error bars represent standard deviations.

landmarks during the lowering and retraction (peak tangential velocity, V_height_peak).

Significant differences between the word accents were found for the measurements taken at the peak tangential velocity landmark (V_height_peak; $t = -5.3$, $p < 0.001$) and at the tongue body target landmark (V_height_target;

$t = -2.5$, $p < 0.05$), but not for the measurement at the onset of the tongue body movement (V_height_onset, see Table 4). In addition, the model fit for the measurements at peak velocity revealed an interaction with vowel length ($t = 5.0$, $p < 0.001$), as well as a significant difference between the vowel lengths ($t = -5.3$, $p < 0.001$). In order to resolve the interaction, we tested for simple main effects of *word accent* per vowel length category. To this end, two new models (with only word accent as a fixed effect, and again, random intercepts for speaker and word; random slopes were not warranted) were fit for V_height_peak based on the subsets of words with long or short vowels, respectively. These models revealed a significant effect of word accent for the long vowels ($t = -5.5$, $p < 0.001$), but not for the short vowels ($t = 1.7$, $p = 0.089$).

The models for V_height_peak and V_height_target suggest a slightly lower tongue body for A2, but the estimated average differences between the word accents are slim: 0.7 mm for the variable V_height_peak, and 0.4 mm for V_height_target, which is beneath the reliable resolution of the EMA. Figure 7 indicates a lot of variation, which we would interpret as a sign of speaker variability. Figure 8 displays results for V_height_peak and V_height_target (where a significant effect of word accent was attested) for each individual speaker, indeed suggesting a certain degree of speaker variability. In particular, the attested overall direction of the effect of word accent (a lower tongue body for A2) is clearly observable for only about half of the speakers, and either absent or even reversed for the other half of the speakers (Figure 8).

In order to better understand the significance of the tongue body height measurements given the observed variability, Figure 9a visualizes the vertical tongue body movement, pooled across all speakers, but speaker-normalized, by means of regression lines per word accent and vowel length category. To this end, the acoustic consonant and vowel intervals were first time-normalized by z -transforming the time points (for details see 2.3), and then vertical positions of the tongue body sensor were normalized for each of the two segments by z -transforming the positions on the basis of all positions collected for the individual speaker.⁴ The figure shows that the regression lines for three of the four conditions hardly differ, while A1 words with long vowels display an overall higher tongue body throughout the CV sequence. The figure suggests, in other words, an interaction between word accent and vowel length, with a main effect of word

⁴ That is, a mean and standard deviation of the vertical tongue body position data was obtained for each individual speaker. For reasons of feasibility, however, these calculations were not based on all available recordings per speaker, but only on the material that was segmented for the purpose of the present study, i.e. the eight target sentences (excluding the leading questions) according to Table 3 (with all repetitions).

Tongue body height

Figure 8: Results (per speaker) of the spatial measurements (in mm) of the vocalic gesture: Tongue body height (mean-centered) as an effect of word accent (A1, A2) and vowel length (long V: grey, short V yellow): at peak tangential velocity (V_height_peak) and at minimum tangential velocity at the gestural target (V_height_target).

accent for long vowels only, namely a slightly higher tongue body for A1. This analysis is, basically, well in line with our modeling results, although the interaction was confirmed by the model for V_height_peak only.

Although this study does not include any horizontal tongue body measurements, Figure 9b displays the horizontal tongue body position over time in order to provide a more complete picture of how tongue body movements might differ between the accent conditions. The figure shows, unsurprisingly, that long vowels reach a far more retracted tongue body position than short vowels (hence, the established transcriptions [a] vs. [ɑ:]). In addition, the figure suggests that for long vowels, A2 displays not only a slightly lower (as seen in Figure 9a), but also slightly more retracted tongue body than A1.

To sum up the main result of the tongue body height measurements, there is a small but significant effect of word accent on tongue body height in interaction with vowel length, displaying a tendency for a higher tongue body for A1 words with long vowels. However, this tendency is observable for about a half of the speakers only (Table 4; Figures 7–9).

Figure 9: Polynomial regression lines for the normalized vertical (a) and horizontal (b) tongue body position data during the word-initial consonant (C₁) and the vowel (V₁), as an effect of word accent and vowel length.

3.2.4 Coordinative measurements: CV time lag

Finally, Figure 10 displays the results for the time lag measurements between the consonantal and the vocalic gestural onsets: the onset of the bilabial closing interval (C) and the onset of the tongue body movement (V). For the two CV time lags that were measured based on lip and vertical tongue body velocity (CV_timelag_Mücketal and CV_timelag_Gao), *vowel length* as a fixed factor was not warranted. The models showed no significant difference between the two word

Figure 10: Results of the coordinative measurements (in ms) of the consonantal and the vocalic gesture: CV time lags between bilabial closing onset (0) and tongue body onset (boxes) as an effect of word accent, based on: zero-crossing in the velocity curve for lip aperture and the vertical tongue body movement (CV_timelag_Mückeetal), 20% threshold from zero-crossing to peak velocity for lip aperture and the vertical tongue body movement (CV_timelag_Gao), and lip acceleration and minimum tangential velocity of the tongue body (CV_timelag_L&G).

accents (see Table 4). On the other hand, *vowel length* as fixed effect did contribute to the model fit on the CV time lag that was measured based on the acceleration of the lips and the tangential velocity of the tongue (CV_timelag_L&G). The results for this CV time lag measure revealed a significant difference between the word accents ($t = -3.4, p < 0.01$), as well as between the vowel lengths ($t = 2.7, p < 0.01$), but no interaction between word accent and vowel length. The estimated difference in time lag between the word accents (for long vowels) was 8 ms (Table 4, CV_timelag_L&G).

Furthermore, in contrast to all other measurements in this study, for the time lags it is also of interest for the discussion to report overall means, pooled across word accents (Table 5). The overall mean was 60⁵ ms based on when the lip and the vertical tongue movements were minimal (CV_timelag_Mückeetal), compared to 68 ms for when the movements were in progress (CV_timelag_Gao). In contrast, there was no time lag to speak of (1 ms) when measured based on the acceleration of the lips and the tangential velocity of the tongue (CV_timelag_L&G), thus suggesting synchronized CV onsets.

To conclude, although Figure 10 suggests slight differences between the word accents in all three panels, it was only the synchronized CV onsets (CV_timelag_L&G) (bottom panel) that showed a significant difference between the word

⁵ Note that these overall means are not included in Table 4, but have been calculated separately. In Table 4, the Estimates for the Intercepts correspond to the mean of the reference categories; for instance, for CV_timelag_Gao, the Intercept Estimate is 70 ms according to Table 4, which corresponds to the mean for Accent 1 words only.

Table 5: CV onset-to-onset time lags obtained with three alternative measures for Swedish (present study), compared to Mandarin (Gao 2008), and German and Catalan (Mücke et al. 2012). All measurements presented in the table refer to a /ma/-sequence. Our results are pooled across vowel lengths and word accents (based on 19 speakers). Likewise, the result for Mandarin represents a grand mean across the four tones (pooled across /ma/ and /man/, 7 speakers). Data for German and Catalan are means from different conditions (open vs. closed syllables; broad and contrastive focus; based on 1 speaker per language).

Measure	Present results: Swedish		Previous results		Source
CV_timelag_Gao	68 ms		44 ms	(means for individual tones vary between 43–46 ms, n.s.)	Gao (2008)
CV_timelag_Mücketal	60 ms		≈5 ms	(3 ms ≤ x ≤ 7 ms)	Mücke et al. (2012)
			≈−2 ms	(−6 ms ≤ x ≤ 3 ms)	Mücke et al. (2012)
CV_timelag_L&G	1 ms	(difference between A1 and A2 ~8 ms ***)		(no data for /ma/ included in Löfqvist and Gracco 1999)	

accents: in A1 the movement towards bilabial closing accelerates before the tongue body movement starts, while in A2 the lips accelerate after the tongue has started to move. On the other hand, an overall CV time lag (60–68 ms) was observed only for the other two measures (CV_timelag_Mücketal and CV_timelag_Gao).

4 Discussion and conclusions

The present study has revealed small but significant spatial, temporal and coordinative differences in the articulation of a CV syllable onset between the two tonal contexts provided by the Swedish word accents, A1 and A2. In Figure 2, articulatory measurements that showed statistically significant differences between the word accents are marked with an asterisk (*); this applies to one of the three bilabial closing intervals, one bilabial opening interval, the palatal wide interval, the tongue body height during tongue lowering and at the vocalic target, and the CV time lag as measured at maximal acceleration (lips) and tangential velocity minimum (tongue body). In addition, the other two CV time lag measurements (using velocity-based landmarks) revealed an overall CV time lag (but no distinction between the word accents), comparable to the time lags reported for Mandarin and other tone languages (Gao 2008; Hu 2016; Karlin and Tilsen 2015).

In the following sections, the results are discussed with respect to each articulator and the coordination between them, starting with the tongue body, as we deem the effect of word accent on the palatal wide interval as the most straightforwardly explained result of this study. It is also relatively salient (in terms of milliseconds) and robust (i.e. independent of vowel length).

To outline our conclusions, we suggest that the observed spatial, temporal and coordinative differences between the word accent conditions are most likely explained in terms of biomechanical effects of the physiological connections between the articulators. Furthermore, although our results seem to suggest that Swedish CV coordination is affected by the presence of a tone gesture, as has been suggested for tone languages, we argue that more research on tonal and non-tonal languages is required, before such conclusions can be drawn.

4.1 The tongue body movements

Our results suggest a robust effect of the tonal context on the temporal measurement of the tongue movement towards the hypothetical tongue body target (the low and retracted tongue position; the palatal wide interval), but a less robust effect on the spatial measurements of tongue body height.

The interval of the lowering and retraction of the tongue body movement (the palatal wide interval) differed between the word accents (A2 having the longer intervals) and between vowel lengths (long vowels having longer intervals, unsurprisingly). There was no interaction between these two factors on the duration of the interval, which means that the vocalic gesture interval is shorter in A1 than in A2 no matter whether the vowel is a short and more fronted [a] or a long [ɑ:].

Our spatial measurements were less robust and less conclusive, as an overall effect of word accent on tongue body position could only be observed for long vowels, where the measured effect on tongue body height was small (<1 mm on average). It was also highly dependent on the individual speaker (3.2.3).

In what follows, we first argue that our temporal results are straightforwardly explained as a biomechanical effect of the physiological connections between the tongue, the jaw and the larynx (4.1.1), and then discuss how the present results may have been affected by the post-vocalic consonant (4.1.2).

4.1.1 A biomechanical account

Previous studies have observed effects of lexical tones on tongue body height (Erickson et al. 2004; Hoole and Hu 2004; Shaw et al. 2016), and it has been argued that such relations can be explained as a result of the physiological connections between the tongue, the jaw and the larynx (see 1.2 above; henceforth, the tongue- f_0 effect). We argue that our primary finding concerning the tongue movements – i.e. the longer palatal wide interval in A2 – can be explained with reference to these physiological connections.

The tongue- f_0 effect can be said to favor a lower tongue position when f_0 is low, and a higher tongue position when f_0 is high. In extension, a lowering of the tongue with a simultaneous fall in f_0 could be considered “in line” with the tongue- f_0 effect, and hence probably articulatorily favorable (as in /ma/ with A1 in the present study), compared to when f_0 is simultaneously rising. In other words, a tongue body lowering and a simultaneous f_0 rise (as in /ma/ with A2) can be characterized as antagonistic, involving a more complex articulation. This complexity, we suggest, might explain the longer duration of the palatal wide interval in A2 than in A1. This explanation might likewise be formulated in terms of an advantage of the non-antagonistic movements, that is the less complex articulation: The lowering of f_0 in A1 can be assumed to support the lowering of the tongue, resulting in the shorter palatal wide interval in A1.

The account just sketched assumes that antagonistic movements represent a more complex articulation which takes more time to execute than a combination of non-antagonistic movements. This durational effect might simply be because the antagonistic movement is executed slower. However, it may also be the case that

the antagonism is what conditions the tongue to take a slightly different route, or vary its velocity differently over the course of time, for A1 and A2, and thus resulting in a more complex movement. In the remainder of this section, we first further elaborate on the biomechanics behind the tongue- f_0 effect, and second, present some additional, although preliminary, evidence in favor of it. We also briefly address the spatial results and a possible role of vowel length on tongue body height.

First, in order to provide a more solid basis for our tongue- f_0 explanation, we should elaborate on the role of the jaw. The jaw has been shown to have a larger displacement in connection with low tones than with high tones (Erickson et al. 2017). While the cricothyroid (CT) muscle has been shown to be most important in f_0 control, being responsible for the tension and relaxation of the vocal folds, for a low tone, there is also the possibility of lowering the entire larynx, using other extrinsic strap muscles, such as the sternohyoid (SH), that are also active when lowering the jaw (Erickson 2011; Erickson et al. 1983; Honda et al. 1999). Thus, the jaw and its connections with the tongue on the one hand and the larynx on the other might also exert an additional support, at least indirectly, to the lowering of the tongue when f_0 is lowered, in addition to the support provided directly through the muscular connection between the tongue and the larynx (via the hyoid bone). Hence, the tongue- f_0 effect can be assumed to be the result of a combination of at least two biomechanical connections, supporting its explanatory power in the present context.

Additional, external evidence for our tongue- f_0 explanation is also provided by the results of our pilot study published in Svensson Lundmark et al. (2017). This study adds relevant evidence since the test words used involved the high vowel /i/ instead of the low vowel /a/. Interestingly, the results for the *palatal narrow interval* obtained for the /i/-words were opposite to the results for the *palatal wide interval* in the present study: a longer interval for A1 than for A2. These results are, however, plausible and in line with our present results under the assumption of our tongue- f_0 explanation. In both the 2017 study and the current one, we measured a longer palatal (wide/narrow) interval when articulation was more complex, that is when the tongue body movement was antagonistic to the direction of the f_0 movement, which was the case when a falling f_0 was combined with tongue raising (A1 in /i/ in Svensson Lundmark et al. 2017) or when a rising f_0 was combined with tongue lowering (A2 in /a/ in the present study). Likewise, shorter intervals were found when there were non-antagonistic movements, that is a less complex articulation, in the case of a rising tone combined with high vowel in A2 (Svensson Lundmark et al. 2017) and a falling tone combined with an open vowel in A1 (this study). However, since our pilot study (Svensson Lundmark et al. 2017) included EMA data from two speakers only, future studies should try to further corroborate the relation between tongue body movement duration (i.e. the palatal

wide/narrow interval) and the articulatory antagonism of f_0 and tongue body movement direction by means of testing a variety of open and close vowels using a larger number of speakers.

Turning to our small and less conclusive spatial effects of word accent on tongue body height, these results might to some degree be affected by inherent quantity-related differences in tongue height for the Swedish open /a/ vowel. As mentioned, in the long condition, the vowel is significantly retracted, pronounced near Cardinal 5 ([ɑ:]), while it is realized rather central, roughly between Cardinal 4 and 5, in the short condition ([a]). One could hypothesize that these allophones are not only distinguished in the front-back dimension, but also in openness. In other words, the long [ɑ:] possibly does not reach the same low tongue position as the short [a], because the speaker puts more effort on the retraction, which might be more important for the characteristic distinction between the allophones. It would seem, though, that our results are only partly in line with this explanation, as a higher tongue body for long vowels is, overall, only observed for A1 (Figure 9a). However, a quantity-effect of this kind may be part of an explanation for the observed effect of word accents, if we assume that in A2, for some other reason, the tongue body is forced to a lower (and more retracted) position in the long-vowel condition (Figure 9); then, the quantity would practically only affect A1 words with long vowels, explaining the observed interaction concerning tongue body height (Figure 9a). This “other reason” just mentioned might be the more complex movement behavior in A2 conditioned by the antagonistic articulation pattern discussed above. This complexity possibly involves an adjustment of the target position, such that the tongue body reaches a lower and more retracted position for A2 than for A1. We cannot fully solve this puzzle in the present study.

However, another piece in this puzzle is likely to be related to speaker variability, which we have informally observed in this study (Figure 8). While A1 displays a trend for a higher tongue body than A2 in some speakers, in other speakers this effect seems to be absent or even reversed. What we probably can conclude from this inconsistency is that the effect of word accent on tongue height is most likely also conditioned by mechanical factors, just as the temporal effect. These factors should be addressed in future research, starting with a closer look at speaker variability.

To conclude, in this section we have reflected upon possible explanations for our spatial results (tongue body height), and first and foremost, suggested a tentative biomechanical explanation for the observed effect of word accent on the palatal wide interval, by means of reference to the tongue- f_0 effect. Thus, we suggest that the tongue- f_0 effect, typically observed when studying spatial effects of tone on tongue movements (e.g., Erickson et al. 2004; Hoole and Hu 2004; Shaw et al. 2016), can also account for temporal effects.

4.1.2 Effects of the post-vocalic consonant

We need to raise a concern to be considered when evaluating our discussion in 4.1.1. Our material consists of words with different post-vocalic consonants (two alveolars, /n/ and /l/, and bilabial /m/). Although our analyses focus on the tongue body movement towards the target (corresponding roughly to the first half of the acoustic vowel interval), we cannot exclude the possibility that post-vocalic consonantal gestures might affect both the jaw position and the tongue body movement of the open stressed vowel. The tongue body movement may be affected, since it will either need to adjust to a following tongue tip constriction (in the case of an alveolar) or not (in the case of a bilabial). Furthermore, for a post-vocalic tongue tip constriction, the jaw displays different degrees of displacement depending on place and manner (Kawahara et al. 2014; Lindblom 1983; Mooshammer et al. 2007).

The choice of post-vocalic consonant could have affected our results especially because it was not perfectly balanced with respect to the vowel length distinction in the material. In the speech material the long-vowel words are followed by dentoalveolar consonants only, while the short-vowel words contain either dento-alveolar or bilabial post-vocalic gestures. When pooled across all words with long and short vowels respectively, we might therefore expect to observe a stronger tongue body movement upwards away from the vowel target in long vowels (where all words contain a post-vocal tongue-tip constriction), compared to short vowels (where only half of the words contain a tongue-tip constriction). Such a tendency is indeed seen in Figure 9a, where the dotted lines representing long vowels slightly diverge upwards towards the offset of V1. However, this effect seems to commence after the vowel target is reached and has, therefore, probably not seriously affected our measurements.

We have tried to account for possible effects of the post-vocalic consonant by means of letting our mixed models consider the target word as a source of random effects. However, future analyses of the interplay between the tongue body, the jaw and the laryngeal structures should probably control for subsequent phonemes in order to compare a variety of post-vocalic conditions.

4.2 The lip movements

The results from the bilabial interval measurements suggest only small effects of the tonal context. For the closing interval, the results did not provide strong support for any robust effects of word accent (Figure 5a), as we only observed a small, yet significant, tendency for longer intervals in A1 when measured with the

acceleration landmarks (Table 4). For the opening intervals, instead, the descriptive results suggest a more robust effect, seen for all three measures (although significant only for C_open_20vel), with a tendency for longer intervals in A2 than in A1 – i.e. in the opposite direction as for the closing (Figure 5b, Table 4). Both effects, we suggest, can be reasonably explained with reference to biomechanics, just as in the case of the tongue body.

4.2.1 The bilabial opening

Beginning with the more robust effect on the opening interval (Figure 5b), we propose a biomechanical explanation along the same lines as for the tongue body movement (see 4.1.1), that is, in terms of articulatory complexity due to antagonistic movements. For both the palatal wide interval and the lip opening, the longer interval was observed for A2. In 4.1.1 we have argued that the lowering of the tongue can benefit from the falling f_0 in A1, due to the tongue- f_0 effect, while the rising f_0 in A2 is antagonistic to tongue lowering, resulting in a more complex articulation and hence a longer duration of the tongue lowering and retraction movement. The jaw, we have argued, plays an important role in this process, too, because it is usually lowered when the tongue is lowered, while it is also directly involved in an f_0 lowering mechanism. That is, just as for the tongue, a jaw lowering can be conceived of as antagonistic to an f_0 rise, but non-antagonistic to, and even supportive for, an f_0 lowering.

Turning to our lip movements again, it is obvious that lip opening and closing movements are strongly related to jaw movements.⁶ Thus, we can infer that a lip opening (= jaw lowering) is antagonistic to the rising f_0 in A2, but beneficial for the falling f_0 in A1. Hence, assuming that in A2, f_0 is rising during the interval of the lip opening, we could predict a longer lip opening in A2 than in A1 due to the articulatory complexity resulting from the antagonistic f_0 raising and jaw lowering. Likewise, if in A1, f_0 is falling during this interval, we could predict a shorter lip opening interval for A1, because the jaw might be lowered faster in support of the lowering of f_0 , resulting in a shorter lip opening for A1. It is difficult to determine which of these two effects plays the more important role – the antagonistic jaw movement in A2, or the supportive jaw movement in A1 – but as they both predict a longer lip opening in A2 than in A1, it may be that both effects have a part in the observed durational difference.

⁶ Note that we disregard degrees of freedom of the lips in relation to the jaw, and instead simply assume that lip opening and closing is mostly controlled by jaw movements. This assumption needs to be tested with future experiments.

A necessary pre-requisite for the account discussed here is that f_o is indeed falling in A1, or rising in A2, during the interval of the lip opening. The lip opening occurs, obviously, around the acoustic onset of the vowel, and we can see in Figure 1 that, for A2, f_o has started rising around this time, while for A1, f_o is about to start falling. This would mean that the faster jaw lowering in A1 is probably executed right before f_o starts falling, or in preparation for it, which would seem to be in line with our biomechanical account. The jaw lowering taking place in order to open the bilabial closure during the CV transition might thus, we suggest, be amplified in order to support the upcoming lowering of f_o . To conclude, the same biomechanical reasoning can account for the effect of word accent on the bilabial opening and the palatal wide interval (where in both cases, the interval is shorter in A1 than in A2): Jaw (and tongue) lowering is antagonistic to a rising f_o , but beneficial for a lowering of f_o .

The bilabial opening interval differed also between the two vowel lengths, but only when measured with velocity landmarks, the opening occurring faster in the short vowel condition. To anticipate the results for the bilabial closing, we also found a significantly shorter closing for short vowels, although only when measured with acceleration-landmarks. A possible interpretation of these results is that the production of the initial C is speeded up in preparation for an earlier upcoming post-vocalic coda consonant in the short vowel condition. However, since in Swedish, the quantity contrast also signifies different vowel qualities, the effect on the lip opening could instead also be related to the quality of the vowel, although we cannot offer any explanation for why the lips should open faster in connection with the more front vowel.

4.2.2 The bilabial closing

For the results for the bilabial closing, however, we cannot develop an equivalent explanation of the effect of word accent, because at about the time when the lips are closing (i.e. prior to the acoustic onset of /m/), neither A1 nor A2 involve an f_o movement that would be antagonistic to the lip closing (i.e. jaw raising). This is indicated in Figure 1, where we can see for both A1 and A2 that f_o is either rather flat or about to rise when /m/ sets in.

An alternative explanation of the longer lip closing interval observed for A1 can, instead, be derived from the small, yet significant, difference seen in the acceleration-based landmark between A1 and A2, perhaps indicating a difference in lip openness, prior to the closing movement, between the two tonal contexts of A1 and A2. The logic is as follows. From the second order equation we know that acceleration, velocity and position are dependent of each other (see Iskarous 2017, for an account on a Dynamical Systems Theory for speech). If the timing of

velocity-based landmarks for the lip movements does not vary while the timing of acceleration-based landmarks does, it is most likely that the lips are positioned differently in A1 compared to A2. We have already seen a tendency for this in a study we are still working on, where preliminary results suggest that at around the bilabial closing onset, A1 tends to display a larger lip aperture than A2. At present, we cannot offer a full explanation for such effects of word accent on lip aperture or lip closing. However, we can speculate that these effects may be linked to the head and jaw movements, which we have already observed to differ between the word accents (Frid et al. 2019; Svensson Lundmark and Frid 2019). Since the articulators are physiologically connected with each other, these word-accent effects on, in particular, jaw movements might in turn be explained as a secondary effect of the word-accent specific coordinated movements with the larynx and the tongue (see 4.1.1 above). Hence, the effect of the tonal context on the lip closing might in fact be related to the effect on the opening.

To conclude, an effect of the tonal context was found for the closing interval of the lips (longer for A1), when measured with acceleration-based landmarks, and for the opening interval (longer for A2), when measured with velocity-based landmarks. Both effects, we have argued, can be accounted for with reference to biomechanical processes.

4.3 Lip and tongue body coordination

The purpose with our coordinative timing measurements, the CV time lags, was two-fold. We were, on the one hand, interested in time lag differences conditioned by the word accents, and on the other hand, in the overall size of the time lags obtained irrespective of the word accent. Focusing on the latter, we aim to contribute to the discussion on the usefulness of a CV time lag as evidence for a tone gesture that is integrated in the global CV organization (see 1.2).

At first glance our time lag results seem somewhat inconsistent. We found time lags as well as (roughly) synchronized C and V onset, depending on the choice of measure. Furthermore, a small but significant synchronization difference between the word accents was only observed for one measure (CV_timelag_L&G), for which the overall result was a roughly synchronized CV coordination. The finding of different CV time lags for A1 and A2 was not predicted, but perhaps can be explained as a result of the biomechanical effects of f_0 on the tongue body and lip movements, as further discussed in 4.3.1.

An overall time lag, however, was predicted – and found – with the velocity-based measures. Table 5 summarizes the results for the CV time lags, obtained with the three alternative measures used in this study, and compares them to results

reported in previous studies. The table shows that our results are overall similar to the results obtained by Gao (2008) for Mandarin, suggesting that, in Swedish, tone gestures are integrated in the global organization at word onset (see 1.2), just as it has been suggested for tone languages. However, recall that the existing evidence on tonal and non-tonal languages is somewhat inconclusive, as we have argued in 1.2. We still lack convincing evidence supporting the claim that lexical tone systematically yields a CV time lag, while non-lexical tone does not, which we will continue to discuss in 4.3.2.

Therefore, the major contribution our results add to the discussion is that they provide a new reference for future studies, based on a relatively large number of speakers (see 4.3.2). That said, our results show that the presence or absence of time lags between C and V gestural onsets seems to depend on the choice of measure. To the best of our knowledge, it is an unsettled question how gestural onsets are most adequately estimated, and which out of several alternative metrics should be selected for the purpose of CV time lag calculations (to be further discussed in 4.3.3).

4.3.1 The asynchrony between A1 and A2

Using Löfqvist and Gracco's (1999) measure (based on lip acceleration and tongue tangential velocity), we found overall synchronized gestural C- and V-onsets (Table 5), but at the same time small but significant differences in CV time lag between A1 and A2. The synchronization difference between A1 and A2 was not predicted; at least we had no reason to assume different coupling relations for A1 and A2, even under the possible assumption of a tone gesture in Swedish. We believe that, again, a biomechanical account provides a reasonable explanation of the A1-A2 asynchrony.

To start with, the small but significant difference (the tongue body lowering starting later in A1, in relation to the lip movement) could be a secondary effect caused by the differences in the lowering and the retraction of the tongue body and the jaw between A1 and A2 (as discussed in 4.1.1). In Figure 11, we can see an earlier tongue body target in A1, reached through a faster lowering of the tongue than in A2, a lowering which also tends to start slightly later in A1 than in A2. Possibly, for A2, the tongue body starts moving earlier in order to compensate, although only partially, for the longer time required to execute the antagonistic movements (tongue body lowering with simultaneous f_0 raising).

In addition, the CV time lag difference between A1 and A2 might also to some degree be caused by the difference found for the bilabial closing movements. The bilabial landmark involved in the Löfqvist and Gracco (1999) measure (for which the A1/A2 CV asynchrony was found) is identical to the left boundary of the acceleration-based lip-closing interval. This interval was slightly but significantly longer for A1 than for A2. Hence, it could be that the earlier start of the lip closing in

Figure 11: A representative illustration of the timing between the vertical tongue body movement and the tangential landmarks used for the measurement of tongue body height at onset, peak and target. The four trajectories represent the y -trace of TB sensors over time in four target words (A1: *man:en*, *ma:nen*; A2: *man:a*, *ma:nar*) spoken by one of the speakers. The trajectories are aligned at acoustic C onset. The dotted lines mark the approximate position of the tangential landmarks from which the height measurements are taken, that is, these landmarks also include the horizontal movement.

A1 is related to its slightly longer duration. To conclude, we believe it is reasonable to assume that the observed effect of word accent on the CV time lag (as measured with the Löfqvist and Gracco, 1999, measure) is a result of the biomechanical effects of f_0 on tongue body movements, lip movements, or both.

4.3.2 CV time lag: evidence of tonal integration?

If we were to explain the observed CV time lags (obtained with the velocity-based measures; see Table 5) in terms of Articulatory Phonology and the c-center effect, this would lean towards the lexical pitch-accent language Swedish behaving more like a tone language (Gao 2008; Hu 2016; Karlin and Tilsen 2015) rather than an intonation language (Mücke et al. 2012). Furthermore, like Gao (2008) and Hu (2016), we did not find significant differences in CV time lags between tones (A1 and A2 in our case), when using Gao's (2008) measure. This might indicate that the Swedish word accents exhibit identical coupling relations with the consonant and the vowel⁷ (see also our discussion in 4.3.1 above).

⁷ Notably, the model that Gao (2008) derives from her results predicts a difference between tone 4 in Mandarin on the one hand, and tone 1, 2, and 3 on the other, which is, however, not confirmed by her data.

A possible conclusion from our CV time lag results is, thus, that Swedish tone is integrated in the global organization of articulatory gestures in the syllable onset (Browman and Goldstein 1988), just as lexical tones in Mandarin are (Gao 2008), which would in turn also provide further evidence for the assumption of a tone gesture in AP. However, as we have argued in 1.2. above, the evidence in favor of the integration of tone gestures is still weak, for a number of reasons.

For instance, CV coordination is to some degree language specific (see our discussion in 1.2. and, e.g., Fowler and Saltzman 1993), meaning that differences in CV time lags found between languages cannot unambiguously be attributed to the presence or absence of lexical tone in these languages. Furthermore, previous results on CV time lags based on only one (as in Mücke et al. 2012) or a few participants (as in Gao 2008) cannot reliably be regarded as representative of the language in question, as a high degree of speaker variability has been attested in previous studies. Löfqvist and Gracco (1999) reported speaker (and context) dependent patterns of inter-gestural coordination: In their analysis of CV time lags in a VCV sequence, they observed a consistent negative time lag (final V-onset preceding the C-onset) for one of their three American English speakers, while the other two displayed both positive and negative time lags depending on vocalic context, with slightly different effects from vocalic context for the two speakers. Speaker variability can also be observed in our data. Figure 12 displays the distribution of values obtained for our three CV time lag measures per speaker. The plots suggest a considerable variability within, but also between speakers, where, for many speaker-speaker comparisons, there is not even an overlap of the quartiles. For the Löfqvist and Gracco (1999) measure, for instance, mean values (not displayed in the figure) range from –22 ms (speakers M7 and F9) to +25 ms (for F12), thus displaying a sizeable degree of variability, similar to that reported by Löfqvist and Gracco (1999) for American English. Some speakers deviate more than others. In particular, when measured as in Mücke et al. (2012), most speakers display a positive time lag of at least 40 ms on average, while for some, the mean is closer to zero or even negative (16 ms for M4, –16 ms for F9). This illustrates how individual speakers may be able to bias the results, especially if only one or two speakers are included in a study.

Hence, it is possible that with larger datasets, encompassing a large number of speakers and a wide variety of tonal and phonological contrasts, we would find evidence for similar patterns of CV time lags in tonal and non-tonal languages. In particular, when the labial landmark is measured with velocity, we might observe CV time lags between the lips and the tongue body of approximately 40–60 ms on average for /ma/, as in Gao (2008), as well as in the present study. But, when the labial landmark is measured with acceleration, synchronized CV might be seen, for both tonal and non-tonal languages (as observed in the present study, and roughly

Figure 12: Results (per speaker) of the coordinative measurements (in ms) of the consonantal and the vocalic gesture: CV time lags pooled across word accent and vowel length conditions. For a description of the landmarks, see Figure 10 (or Tables 1 and 2; Figures 2 and 3).

observed in Löfqvist and Gracco 1999, although their data are not entirely comparable to ours, since they did not include a /ma/-sequence). Such hypothetical results – presence or absence of time lag depending on the way it is measured – cast doubt on the idea that a tone gesture is integrated in the global organization at word onset. Future studies should therefore collect data from a variety of tonal and non-tonal languages, and through this cumulative evidence gradually build up a valid picture of the effect of tone on CV coordination. Such comparisons should always involve a sufficient number of speakers.

The results of the present study, we believe, could serve as an adequate initial reference for such future comparisons. That is, future studies should aim to match

(but reach beyond) the, admittedly, limited set of conditions provided by this study: /ma/ with a long or short vowel (whatever is applicable), in a low-prominent utterance position. However, if such future results are to be used as evidence for evaluating the hypothesis that lexical tone leads to a CV time lag, we also need to arrive at a consensus about how CV time lag is to be measured. We discuss this matter in the next section.

4.3.3 How should gestural onset be measured?

This study has shown that different measures for consonantal and vocalic gestural onsets provide different CV time lags, but which are most valid? In our study, the shortest time lags (around 0 ms) were obtained with the measure used by Löfqvist and Gracco (1999), where the consonantal gestural onset was measured at an acceleration-based landmark. Maximal acceleration appears later in the kinematic movement than the velocity zero-crossing, as it marks the maximal change in the velocity of the object being moving (see also Figure 3). This explains why the acceleration-based CV time lag is smaller than the others. One might argue that the acceleration-based landmark is too late to capture the gestural onset of the lip movement, and the velocity-based landmarks are thus more adequate. One might also suggest that the consonantal and the vocalic gestural onsets should always be measured using equivalent landmarks in the sense that they are defined using equivalent signals (e.g., a velocity curve, an acceleration curve etc.). If this would seem reasonable, then the Löfqvist and Gracco (1999) time lag measure would be ruled out, because it uses different types of landmarks for the lips and for the tongue body (the TB minimum tangential velocity and the LA maximal acceleration), while Gao (2008) and Mücke et al. (2012) both use similar landmarks for the two articulators.

We argue, though, that different types of landmarks should only be a concern if consonantal and vocalic gestures were executed by equivalent kinematic control mechanisms. However, vocalic gestures are made with the constantly and relatively slowly moving tongue body (e.g., Öhman 1966), while the consonantal gestures are instant in nature, with rather sharp starts and ends overlapping the vocalic gesture. In other words, consonantal gestures are more often characterized by larger changes in velocity, and hence by rather well identifiable peaks in the acceleration curve. We therefore propose acceleration-based landmarks for measuring consonantal gesture onsets, but not necessarily for vocalic gestures. We suggest that gestural onsets be measured with reference to different signals, such as velocity or acceleration, depending on the articulator. Hence, the Löfqvist and Gracco (1999) measure might well be most adequate for capturing CV onset time lags. However, further research should aim at a consensus on where an

articulatory movement begins and ends, and how it should be measured, as well as what can be counted as timing between different gestural movements.

4.4 Limitations and future studies

The discussion has revealed various important directions for future research. First and foremost, we have argued that the idea of CV time lags as evidence for the integration of a tone gesture in the global CV coordination needs to be revisited. We need a well-grounded consensus on how gestural onsets should be measured, possibly depending on the type of articulator involved (see 4.3.3). Furthermore, more speakers, compared to previous standards, must be included, and a variety of languages with and without lexical tone need to be considered.

It is also important to include different vowels, different initial and post-vocalic consonants, and different phonotactic and prosodic conditions, as some of these factors may have a great influence on CV coordination. Including vowels of varying openness and frontness is especially important in order to evaluate our biomechanical account developed in 4.1.1.

In this study, we have focused on articulatory movements under two prosodic conditions, rather than integrating measurements of tonal events in the articulatory analysis, as done in some previous studies (e.g., Gao 2008; Mücke et al. 2012). Future work should aim to include measurements of f_0 landmarks again, in order to further examine the tone gesture hypothesis and the c-center effect (e.g., Gao 2008). However, the representation of f_0 by means of minima and maxima, as traditionally used in segment-to-pitch analyses, including the cited studies on tonal-articulatory alignment, has been questioned (e.g., D'Imperio 2000; Knight 2003; Niebuhr 2007). More valid means of representing f_0 have been proposed (e.g., Barnes et al. 2008, 2012; Cangemi et al. 2019) which could be used in future studies on tonal-articulatory alignment.

Given the comprehensiveness of the present study, we have excluded jaw movements from the analysis and refer to a parallel study (Svensson Lundmark and Frid 2019). We have also excluded lip position data but referred to ongoing work in a subsequent study. We should mention that we have so far not used peak velocity of the lips as a landmark, which is a viable method of measurement used in many articulatory studies on consonants (Byrd et al. 2005; Cho 2002; Erickson et al. 2014; Mooshammer et al. 2006).

Speaker variability was accounted for in our mixed effects models and exemplified for some selected measurements, but a detailed analysis of the phenomenon *per se* was deemed outside the scope of the present study. However, the

number of speakers included in this study would enable a future exploration of individual speaker strategies as a possible source of variability.

The differences between the word accents found for the tongue body, the lips, and their coordination most likely also affect the acoustic space beyond f_o , such as the formant transitions from the consonant to the vowel. This presumed acoustic difference might possibly render the word accents decodable even before the tones are clearly distinguished by virtue of f_o . There is indeed recent evidence suggesting that tones can affect vowel articulation in a way that renders perceivable differences (Shaw and Tylor 2020). The strong relationship between the tone and the suffix in Swedish word accents suggests that any anticipatory cue in word onsets helps the listener in decoding the morphology, as has been shown for the tonal pattern realized in the initial (stressed) syllable (e.g., Roll 2015; Roll et al. 2010, 2017; Söderström et al. 2017). The present study suggests that further anticipatory acoustic cues beyond f_o might come into play, possibly even before the tonal pattern is clearly distinguished.

4.5 Conclusions

Previous studies have shown that coarticulatory anticipatory movements are sensitive to phonotactics as well as suprasegmental information, but articulatory studies on the coordination with f_o have been sparse. This study critically adds to this line of research, as it is the first comprehensive study investigating the articulatory movements of the lips and the tongue body as a function of the Swedish word accents.

We can derive two main conclusions from the present study. First, the two Swedish word accents, A1 and A2, portray different tongue body movements, as well as slightly different lip movements and coordinated patterns between the lips and the tongue body. Hence, both the consonantal and the vocalic gestures are affected by the tonal context, which we explain as a biomechanical effect of the physiological connections between supra-laryngeal and laryngeal movements, in combination with articulatory complexity: a combination of antagonistic movements (e.g., a tongue body lowering with a rising f_o) takes longer time to execute than a combination of non-antagonistic movements (e.g., tongue body lowering with falling f_o).

Second, our results constitute a novel reference for future studies of CV onset time lags in a variety of languages. This novel reference is more reliable and potentially more valid than time lag values presented in previous studies, because our study is based on a considerably larger number of speakers than the previous related studies (19, as compared to between one and seven per language in the

previous studies), and because we have used three means to capture the CV time lag. We have initiated a discussion of the validity of these measures, which we hope will be continued. Meanwhile, we suggest that a range of measures should be used, as in the present study, and discussed critically. The measurement idea put forward by Gao (2008) is appealing. Lexical tone, as a part of a word's phonological architecture, might well be integrated in the production of a word in much the same way as consonants and vowels are. However, in order to ultimately be able to verify this hypothesis, we need to develop means to measure gestural onsets in a reliable and valid manner and to control for language-specific factors.

Acknowledgements: The authors gratefully acknowledge Lund University Humanities Lab. Thank you to Prof. Merle Horne for fruitful discussions during the initial steps of the outline of this study. Parts of this work were originally presented by the first author at the Speech Motor Control conference in Groningen 2017, and she would like to express her gratitude for all the comments from conference delegates. Special thanks to Prof. Anders Löfqvist for valuable comments. Thank you, Dr. Joost van de Weijer, for advice on statistics in R. We greatly appreciate the very insightful and constructive input from three anonymous reviewers. We also thank Prof. David House for an additional, informal review of a previous draft. And finally, many thanks to all the participants who graciously worked the EMA.

Author contributions: The study was designed by PhD candidate Malin Svensson Lundmark in collaboration with her supervisors Dr. Susanne Schötz and Dr. Gilbert Ambrazaitis. Data were collected by Svensson Lundmark assisted by Dr. Schötz who also instructed Svensson Lundmark in the data collection procedure using EMA. Data analysis was performed by Svensson Lundmark assisted by Dr. Johan Frid and Dr. Ambrazaitis. Data interpretation and discussion was drafted by Svensson Lundmark and further developed in collaboration with Dr. Ambrazaitis. The manuscript was primarily drafted by Svensson Lundmark, in places with assistance by Dr. Ambrazaitis. All authors contributed with critical revisions.

Competing interests: The authors have no conflicts of interest to declare.

Research funding: This study was supported by a Lund University PhD position awarded to the first author. The contribution by Dr. Frid was supported by an infrastructure grant from the Swedish Research Council (Swe-Clarín, 2019–2024; contract no. 2017-00626).

Statement of ethics: All participants read and signed an informed consent. Consent form and methodology were borrowed from a related research project granted to Dr. Susanne Schötz (the fourth author) by the Swedish Research Council (contract no. 2010-1599) and approved by the Regional Ethical Review Board, Lund, Sweden (reference no. 2011/430).

References

- Ambrazaitis, Gilbert. 2009. Nuclear intonation in Swedish: Evidence from experimental-phonetic studies and a comparison with German. Doctoral dissertation. *Travaux de l'institut de linguistique de Lund*, vol. 49. Centre for Languages and Literature, Lund University.
- Ambrazaitis, Gilbert & Gösta Bruce. 2006. *Perception of South Swedish word accents*. Working papers, vol. 52, 5–8. Lund University, Department of Linguistics and Phonetics Available at <https://journals.lub.lu.se/LWPL/issue/archive>.
- Ambrazaitis, Gilbert, Johan Frid & Gösta Bruce. 2012. Revisiting South and Central Swedish intonation from a comparative and functional perspective. In Oliver Niebuhr (ed.), *Understanding prosody – The role of context, function, and communication*, 138–158. DeGruyter.
- Barnes, Jonathan, Nanette Veilleux, Alejna Brugos & Stefanie Shattuck-Hufnagel. 2008. Alternatives to f0 turning points in American English intonation. *The Journal of the Acoustical Society of America* 124. 2497.
- Barnes, Jonathan, Nanette Veilleux, Alejna Brugos & Stefanie Shattuck-Hufnagel. 2012. Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology* 3(2). 337–383.
- Bates, Douglas, Martin Maechler, Benjamin M. Bolker & Steven C. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.
- Blackwood Ximenes, Arwen, Jason A. Shaw & Christopher Carignan. 2017. A comparison of acoustic and articulatory methods for analyzing vowel differences across American and Australian dialects of English. *The Journal of Acoustical Society of America* 142(2). 363–377.
- Boersma, Paul & David Weenink. 2018. *Praat: Doing phonetics by computer [Computer software]*. Version 6.0.37. Retrieved 3 February 2018 from <http://www.praat.org/>.
- Bombien, Lasse, Christine Mooshammer & Philip Hoole. 2013. Articulatory coordination in word-initial clusters of German. *Journal of Phonetics* 41(6). 546–561.
- Browman, Catherine P. & Louis Goldstein. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45(2–4). 140–155.
- Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: an overview. *Phonetica* 49(3–4). 155–180.
- Browman, Catherine P. & Louis Goldstein. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée* 5. 25–34.
- Bruce, Gösta. 1977. *Swedish word accents in sentence perspective*. *Travaux de l'institut de linguistique de Lund*, vol. 12. Lund: Gleerup Doctoral dissertation.
- Bruce, Gösta. 2005. Intonational prominence in varieties of Swedish revisited. In Sun-Ah Jun (ed.), *The phonology of intonation and phrasing*, 410–429. Oxford: Oxford University Press.
- Bruce, Gösta & Olle Engstrand. 2006. The phonetic profile of Swedish. *Sprachtypologie und Universalienforschung (STUF)* 59(1). 12–35.
- Bruce, Gösta & Eva Gårding. 1978. A prosodic typology for Swedish dialects. In Eva Gårding, Gösta Bruce & Robert Bannert (eds.), *Nordic Prosody—Papers from a symposium*, 219–228. Lund: Gleerup.
- Byrd, Dani. 1995. C-Centers revisited. *Phonetica* 52. 285–306.
- Byrd, Dani. 1996. Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24(2). 209–244.

- Byrd, Dani, Sungbok Lee, Daylen Riggs & Jason Adams. 2005. Interacting effects of syllable and phrase position on consonant articulation. *The Journal of the Acoustical Society of America* 118(6). 3860–3873.
- Cangemi, Francesco, Aviad Albert & Martine Grice. 2019. Modelling intonation: Beyond segments and tonal targets. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 572–576. Canberra, Australia: Australasian Speech Science and Technology Association Inc. Retrieved from <http://intro2psycholing.net/ICPhS/>.
- Chen, Wei-Rong, Doug H. Whalen & Mark K. Tiede. 2019. Mandibular contribution to vowel-intrinsic F0. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 152–156. Canberra, Australia: Australasian Speech Science and Technology Association Inc. Retrieved from <http://intro2psycholing.net/ICPhS/>.
- Cho, Taehong. 2002. *The effects of prosody on articulation in English*. New York: Routledge.
- de Jong, Kenneth, Mary E. Beckman & Jan Edwards. 1993. The interplay between prosodic structure and coarticulation. *Language and Speech* 36(2–3). 197–212.
- D'Imperio, Mariapaola. 2000. *The role of perception in defining tonal targets and their alignment*. The Ohio State University Doctoral dissertation.
- Elert, Claes-Christian. 1964. *Phonologic studies of quantity in Swedish*. Stockholm: Almqvist & Wicksell.
- Erickson, Donna. 2011. Thai tones revisited. *Journal of the Phonetic Society of Japan* 15(2). 1–9.
- Erickson, Donna, Thomas Baer & Katherine S. Harris. 1983. The role of the strap muscles in pitch lowering. In Diane M. Bless & James H. Abbs (eds.), *Vocal fold physiology: Contemporary research and clinical issue*, 279–285. San Diego, CA: College-Hill Press.
- Erickson, Donna, Kiyoshi Honda & Shigeto Kawahara. 2017. Interaction of jaw displacement and F0 peak in syllables produced with contrastive emphasis. *Acoustical Science and Technology* 38(3). 137–146.
- Erickson, Donna, Ray Iwata, Mitsuaki Endo & Akinori Fujino. 2004. Effect of tone height on jaw and tongue articulation in Mandarin Chinese. In *Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing, China 2004*, 53–56. International Speech Communication Association. Retrieved from <http://www.isca-speech.org/archive/tal2004>.
- Erickson, Donna, Shigeto Kawahara, Jeff Moore, Caroline Menezes, Atsuo Suemitsu, Jangwon Kim & Yoshiho Shibuya. 2014. Calculating articulatory syllable duration and phrase boundaries. In Susanne Fuchs, Martine Grice, Anne Hermes, Leonardo Lancia & Doris Mücke (eds.), *Proceedings of the 10th International Seminar on Speech Production (ISSP), Cologne, Germany 2014*, 102–105. Retrieved from http://www.issp2014.uni-koeln.de/wp-content/uploads/2014/Proceedings_ISSP_revised.pdf.
- Erickson, Donna, Atsuo Suemitsu, Yoshiho Shibuya & Mark Tiede. 2012. Metrical structure and production of English rhythm. *Phonetica* 69. 180–190.
- Fant, Gunnar. 1969. Distinctive features and phonetic dimensions. *Quarterly Progress and Status Report, Dept. for Speech, Music and Hearing, KTH Stockholm*, 2-3, 1-18. Retrieved from <https://www.speech.kth.se/qpsr/>.
- Fowler, Carol A. & Elliot Saltzman. 1993. Coordination and coarticulation in speech production. *Language and Speech* 36(2–3). 171–195.
- Frid, Johan, Malin Svensson Lundmark, Gilbert Ambrazaitis & David House. 2019. EMA-based head movements, word accent and vowel length. In *Book of Abstracts MMSYM 2019, the 6th*

- European and 9th Nordic Symposium on Multimodal Communication, Leuven, Belgium 2019*, 11. Leuven, Belgium: University of Leuven. Retrieved from <http://mmsym.org/wp-content/uploads/2016/09/MMSYM2019-book-of-abstracts-0905.pdf>.
- Gafos, Adamantios I. 2002. A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20. 269–337.
- Gafos, Adamantios I., Jens Roeser, Stavroula Sotiropoulou, Philip Hoole & Chakir Zeroual. 2020. Structure in mind, structure in vocal tract. *Natural Language & Linguistic Theory* 38(1). 43–75.
- Gao, Man. 2008. *Tonal alignment in Mandarin Chinese: An articulatory phonology account*. New Haven: Yale University Doctoral dissertation.
- Gårding, Eva, Osamu Fujimura, Hajime Hirose & Zyunici Simada. 1975. *Laryngeal control of Swedish word accents*. Working papers, vol. 10, 53–82. Lund University, Department of Linguistics and Phonetics. Retrieved from <https://journals.lub.lu.se/LWPL/issue/archive>.
- Honda, Kiyoshi, Hiroyuki Hirai, Shinobu Masaki & Yasuhiro Shimada. 1999. Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech* 42(4). 401–411.
- Hoole, Philip & Fang Hu. 2004. Tone-vowel interaction in standard Chinese. In *Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing, China 2004*, 89–92. International Speech Communication Association. Retrieved from <http://www.isca-speech.org/archive/tal2004>.
- Hu, Fang. 2016. Tones are not abstract autosegmentals. In Jonathan Barnes, Alejna Brugos, Stefanie Shattuck-Hufnagel & Nanette Veilleux (eds.), *Proceedings of the 8th International Conference on Speech Prosody, Boston, USA 2016*, 302–306. International Speech Communication Association. <https://doi.org/10.21437/SpeechProsody.2016>.
- Iskarous, Khalil. 2017. The relation between the continuous and the discrete: A note on the first principles of speech dynamics. *Journal of Phonetics* 64. 8–20.
- Karlin, Robin & Sam Tilsen. 2015. The articulatory tone-bearing unit: Gestural coordination of lexical tone in Thai. *Proceedings of Meetings on Acoustics* 22(060006). 1–9.
- Kawahara, Shigeto, Hinako Masuda, Donna Erickson, Jeff Moore, Atsuo Suemitsu & Yoshiho Shibuya. 2014. Quantifying the effects of vowel quality and preceding consonants on jaw displacement: Japanese data. *Journal of the Phonetic Society of Japan* 18(2). 54–62.
- Knight, Rachael-Anne. 2003. *Peaks and Plateaux: The production and perception of high intonational targets in English*. University of Cambridge Doctoral Dissertation.
- Kroos, Christian, Philip Hoole, Barbara Kühnert & Hans G. Tillmann. 1997. Phonetic evidence for the phonological status of the tense-lax distinction in German. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 35. 17–25. Retrieved from <https://www.phonetik.uni-muenchen.de/forschung/publikationen/fipkm/>.
- Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26.
- Lahiri, Aditi, Allison Wetterlin & Elisabet Jönsson-Steiner. 2005. Lexical specification of tone in North Germanic. *Nordic Journal of Linguistics* 28(1). 61–96.
- Lehiste, Ilse 1970. *Suprasegmentals*. Cambridge, Massachusetts: M.I.T. Press.
- Lehiste, Ilse & Gordon E. Peterson. 1961. Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America* 33(4). 419–425.
- Lindblom, Björn. 1983. Economy of speech gestures. In Peter MacNeilage (ed.), *The production of speech*, 217–245. New York: Springer.
- Lindblom, Björn, Augustine Agwuele, Harvey M. Sussman & Elisabet Eir Cortes. 2007. The effect of emphatic stress on consonant vowel coarticulation. *The Journal of the Acoustical Society of America* 121(6). 3802–3813.

- Löfqvist, Anders. 1975. Intrinsic and extrinsic f₀ variations in Swedish tonal accents. *Phonetica* 31. 228–247.
- Löfqvist, Anders & Vincent L. Gracco. 1999. Interarticulator programming in VCV sequences: Lip and tongue movements. *The Journal of the Acoustical Society of America* 105(3). 1864–1876.
- Macneilage, Peter F. & Joseph L. Declerk. 1969. On motor control of coarticulation in CVC monosyllables. *The Journal of the Acoustical Society of America* 45(5). 1217–1213.
- Marin, Stefania. 2013. The temporal organization of complex onsets and codas in Romanian: A gestural approach. *Journal of Phonetics* 41(3–4). 211–227.
- Mooshammer, Christine, Lasse Bombien & Jelena Krivokapic. 2013. Prosodic effects on speech gestures: A shape analysis based on functional data analysis. *The Journal of the Acoustical Society of America* 133. 3565.
- Mooshammer, Christine, Philip Hoole & Anja Geumann. 2006. Interarticulator cohesion within coronal consonant production. *The Journal of the Acoustical Society of America* 120(2). 1028–1039.
- Mooshammer, Christine, Philip Hoole & Anja Geumann. 2007. Jaw and order. *Language and Speech* 50(2). 145–176.
- Mücke, Doris, Hosung Nam, Anne Hermes & Louis M. Goldstein. 2012. Coupling of tone and constriction gestures in pitch accents. In Philip Hoole, Lasse Bombien, Marianne Pouplier, Christine Mooshammer & Barbara Kühnert (eds.), *Consonant clusters and structural complexity*, 205–230. Berlin: Mouton de Gruyter.
- Nam, Hosung, Louis Goldstein & Elliot Saltzman. 2009. Self-organization of syllable structure: A coupled oscillator model. In François Pellegrino, Egidio Marsico, Iona Chitoran & Christophe Coupé (eds.), *Approaches to phonological complexity*, 297–328. Berlin: Walter de Gruyter.
- Niebuhr, Oliver. 2007. The signalling of German rising-falling intonation categories—The interplay of synchronization, shape, and height. *Phonetica* 64(2–3). 174–193.
- Niemann, Henrik, Doris Mücke, Hosung Nam, Louis Goldstein & Martine Grice. 2011. Tones as gestures: The case of Italian and German. In Wai Sum Lee & Eric Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong 2011*, 1486–1489.
- Nittrouer, Susan, Kevin Munhall, J. A. Scott Kelso, Betty Tuller & Katherine S. Harris. 1988. Patterns of interarticulator phasing and their relation to linguistic structure. *The Journal of the Acoustical Society of America* 84(5). 1653–1661.
- Ohala, John J. & Brian W. Eukel. 1987. Explaining the intrinsic pitch of vowels. In Robert Channon & Linda Shockey (eds.), *In honor of Ilse Lehiste*, 207–215. Dordrecht: Foris.
- Öhman, Sven E. G. 1965. On the coordination of articulatory and phonatory activity in the production of Swedish tonal accents. *Quarterly Progress and Status Report, Dept. for Speech, Music and Hearing, KTH Stockholm*, 6, 14–19. Retrieved from <https://www.speech.kth.se/qpsr/>.
- Öhman, Sven E. G. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America* 39(1). 151–168.
- Öhman, Sven E. G., B. Mårtensson, R. Leanderson & A. Persson. 1967. Cricothyroid and vocalis muscle activity in the production of Swedish tonal accents: A pilot study. *Quarterly Progress and Status Report, Dept. for Speech, Music and Hearing, KTH Stockholm*, 8(2–3), 55–57. Retrieved from <https://www.speech.kth.se/qpsr/>.
- Peterson, Gordon E. & Harold L. Barney. 1952. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America* 24(2). 175–184.

- Poupplier, Marianne. 2012. The gestural approach to syllable structure: Universal, language- and cluster-specific aspects. In Susanne Fuchs, Melanie Weirich, Daniel Pape & Pascale Perrier (eds.), *Speech planning and dynamics*, 63–96. New York, Oxford, Wien: Peter Lang.
- R Core Team. 2015. *R: A language and environment for statistical computing [Computer software]*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>.
- Recasens, Daniel. 2002. An EMA study of VCV coarticulatory direction. *The Journal of the Acoustical Society of America* 111(6). 2828–2841.
- Reinholt Petersen, Niels. 1978. Intrinsic fundamental frequency of Danish vowels. *Journal of Phonetics* 6. 177–189.
- Roll, Mikael. 2015. A neurolinguistic study of South Swedish word accents: Electrical brain potentials in nouns and verbs. *Nordic Journal of Linguistics* 38(2). 149–162.
- Roll, Mikael, Merle Horne & Magnus Lindgren. 2010. Word accents and morphology—ERPs of Swedish word processing. *Brain Research* 1330. 114–123.
- Roll, Mikael, Pelle Söderström, Johan Frid, Peter Mannfolk & Merle Horne. 2017. Forehearing words: Pre-activation of word endings at word onset. *Neuroscience Letters* 658. 57–61.
- Shaw, Jason A., Wei-rong Chen, Michael I. Proctor & Donald Derrick. 2016. Influences of tone on vowel articulation in Mandarin Chinese. *Journal of Speech, Language, and Hearing Research* 59(6). 1566–1574.
- Shaw, Jason A., Adamantios I. Gafos, Philip Hoole & Chakir Zeroual. 2009. Syllabification in Moroccan Arabic: Evidence from patterns of temporal stability in articulation. *Phonology* 26. 187–215.
- Shaw, Jason A. & Michael D. Tyler. 2020. Effects of vowel coproduction on the time course of tone recognition. *The Journal of the Acoustical Society of America* 147(4). 2511–2524.
- Smith, Caroline L. 1995. Prosodic patterns in the coordination of vowel and consonant gestures. In Bruce Connell & Amalia Arvaniti (eds.), *Laboratory phonology IV: Phonology and phonetic evidence*, 205–222. Cambridge: Cambridge University Press.
- Söderström, Pelle, Merle Horne & Mikael Roll. 2017. Stem tones pre-activate suffixes in the brain. *Journal of Psycholinguistic Research* 46(2). 271–280.
- Stevens, Kenneth N. & Sheila E. Blumstein. 1981. The search for invariant acoustic correlates of phonetic features. In Peter D. Eimas & Joanne L. Miller (eds.), *Perspectives on the study of speech*, 1–38. Hillsdale, NJ: Erlbaum.
- Svensson, Malin. 2014. Constant tonal alignment in Swedish word accent II. In NickCampbell, Dafydd Gibbon & Daniel Hirst (eds.), *Proceedings of the 7th International Conference on Speech Prosody, Dublin, Ireland 2014*, 987–991 Retrieved from <http://fastnet.netsoc.ie/sp7/sp7book.pdf>.
- Svensson Lundmark, Malin, Gilbert Ambrazaitis & Otto Ewald. 2017. Exploring multidimensionality: Acoustic and articulatory correlates of Swedish word accents. In *Proceedings of Interspeech, Stockholm, Sweden 2017*, 3236–3240. International Speech Communication Association. <https://doi.org/10.21437/Interspeech.2017>.
- Svensson Lundmark, Malin & Johan Frid. 2019. Jaw movements in two tonal contexts. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 1843–1847. Canberra, Australia: Australasian Speech Science and Technology Association Inc. Retrieved from <http://intro2psycholing.net/ICPhS/>.
- Titze, Ingo R., Ronald J. Baken, Kenneth W. Bozeman, Svante Granqvist, Nathalie Henrich, Christian T. Herbst, David M. Howard, Eric J. Hunter, Dean Kaelin, Raymond D. Kent,

- Jody Kreiman, Malte Kob, Anders Löfqvist, Scott McCoy, Donald G. Miller, Hubert Noé, Ronald C. Scherer, John R. Smith, Brad H. Story, Jam G. Švec, Sten Ternström & Joe Wolfe. 2015. Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *The Journal of the Acoustical Society of America* 137(5). 3005–3007.
- Wieling, Martijn & Mark Tiede. 2017. Quantitative identification of dialect-specific articulatory settings. *The Journal of the Acoustical Society of America* 142(1). 389–394.
- Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics* 27(1). 55–105.
- Xu, Yi. 2013. ProsodyPro — A tool for large-scale systematic prosody analysis. In Brigitte Bigi & Daniel Hirst (eds.), *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France 2013*, 7–10. Aix-en-Provence, France: Laboratoire Parole et Langage. Retrieved from http://www2.lpl-aix.fr/~trasp/Proceedings/TRASP2013_proceedings.pdf.
- Zhang, Muye, Christopher Geissler & Jason Shaw. 2019. Gestural representations of tone in Mandarin: Evidence from timing alternations. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 1803–1807. Canberra, Australia: Australasian Speech Science and Technology Association Inc. Retrieved from <http://intro2psycholing.net/ICPhS>.