The Coordination of Pitch Accents with Articulatory Gestures: A Dynamical Approach

Inaugural-Dissertation zur Erlangung des Doktorgrades der
Philosophischen Fakultät der Universität zu Köln
im Fach Phonetik

vorgelegt von

Henrik Hubert Niemann
aus Coesfeld (Westf.)

Köln, Oktober 2016

Erste Referentin: PD Dr. Doris Mücke

Zweite Referentin: Prof. Dr. Martine Grice

Datum der letzten Prüfung: 11. Januar 2017

Contents

List of Figures

List of Tables

1 Introduction

The present thesis investigates the coordination of nuclear and prenuclear rising pitch accents in the acoustic and articulatory domain in German. In the past thirty years, there have been many studies investigating the alignment of intonational events relative to segmental structure. Specifically, researchers have looked at the synchronization of low and high turning points of rising pitch accents with landmarks in the acoustic domain such as the beginning or end of the accented syllable with which the pitch accent is phonologically associated. Most of these studies have found, however, that the beginning of an accentual rise is stably aligned with the beginning of the accented syllable, while the end of the accentual rise is highly sensitive to, for example, syllable structure, word length and phrasal position (see, for example, Ladd 2008, Prieto 2011, D'Imperio 2012).

The work reported on here focuses on the coordination of pitch accents with articulatory gestures in the framework of Articulatory Phonology, which views dynamically-defined gestures as phonological primitives (Browman & Goldstein 1989, 1990). In this framework, phonological structure is directly reflected in the timing between articulatory gestures. This intergestural timing successfully can be modelled by associating articulatory gestures with non-linear planning oscillators, or clocks, that are coupled with each other (Saltzman & Kelso 1987, Saltzman et al. 2006, Nam, Goldstein & Saltzman 2009). The organisation of consonants and vowels in a syllable arise from specific coupling structures underlying the interarticulatory timing. More recently, the concept of gestures has been applied to the laryngeal system by integrating (lexical) tones (Gao 2009, Hsieh 2011), (post-lexical) pitch accents and boundary tones (Mücke et al. 2012, Niemann et al. 2011, Katsika et al. 2014) into the syllable coupling network.

A major goal of this study is thus to contribute to the growing body of evidence of a tight link between the laryngeal system (forming the tune of an utterance) and the supralaryngeal system (forming articulatory movements and ultimately producing the sounds) of speech. For rising nuclear accents, for example, we have found evidence for a stable, or at least a more stable, coordination between the accentual rise of the pitch accent with landmarks in the articulatory domain as compared to landmarks in the acoustic domain (cf. Niemann, Grice & Mücke 2014, Niemann & Mücke 2015). This thesis provides an in-depth exploration of the timing of both nuclear and prenuclear pitch accent relative to segmental and articulatory landmarks. The analysis not only focuses on timing patterns found across speakers, but also provides evidence for speaker-specific strategies. More specifically, a dynamical model for text-tune coordination is provided by employing prosodic gestures, the π-gesture and the μ-gesture, in the coupled oscillator model.

This study employs electromagnetic articulography (EMA) to examine the temporal relationship between accentual rises and oral constriction gestures in German in order to shed light on their coordination as a function of phrasal position, syllable structure and word boundary. This thesis is structured as follows: In Chapter 2, the basic concepts of the prevailing model of intonation, the Autosegmental-Metrical model, will be introduced. In particular, the *Segmental Anchoring Hypothesis* will be presented, which posits that tones are attached to the segmental string in a regular and predictable way. The following section reviews studies on different languages investigating the numerous factors that have been found to effect the alignment of pitch accents.

Chapter 3 introduces the framework of Articulatory Phonology and focuses on patterns of intergestural coordination. Therefore, the coupled oscillator model of syllable structure rooted in Articulatory Phonology is introduced. Recently, this model has been applied to higher prosodic units such as the intonation phrase involving boundary- and stress-related effects on gestural timing. The chapter concludes with a review of studies that have adapted the concept of gestures to tonal events such as pitch accents or lexical tones. Chapter 4 gives an overview of the methods of the current study by presenting details on the speakers, speech material, the recording procedure, data processing and the annotation scheme.

Results are presented in Chapter 5 (nuclear accents) and Chapter 6 (prenuclear accents). The analyses include both the alignment of the rising pitch accents relative to segmental boundaries and the coordination of the rising pitch accents with consonantal and vocalic gestures produced by the lips and the tongue body, respectively. Different stability patterns are suggested. Chapter 7 provides a dynamical model for the articulatory data in terms of different coupling structures. For this purpose, different sources of variability caused by the dynamical behaviour of the articulatory system are taken into account. Chapter 8 summarises the findings of this study and suggests possible future directions for research.

2 Intonation

2.1 The Autosegmental-Metrical model of intonation

When we speak, we do not utter each sound in isolation and one after another like pearls on a necklace. We also do not talk in monotonous way. Rather, we integrate speech units into chunks and produce them with a tune. Generally speaking, speech can be broken down into a segmental and a suprasegmental part. The segmental part refers to the sounds produced by our articulators such as a voiced labial nasal [m] produced by closing the lips while lowering the velum, allowing the airflow to pass through the nasal cavity at the same time. Voicing is provided by continuous vocal fold oscillations. The suprasegmental part is said to be superimposed on the segmental part and refers to the prosody involving, on the one hand, the intonation and the rhythm of an utterance and, on the other hand, the stress and accentuation of words and, more specifically, syllables. Physiologically, the intonation, or tune, of an utterance is generated by rapid changes of the frequency, the vocal fold oscillations (fundamental frequency, F0). The perceptual counterpart is pitch; fast vocal fold oscillations (high F0) lead to the perception of a high tone, while slow vocal fold oscillations (low F0) lead to the perception of a low tone.

In general, there are two different approaches to the phonology of intonation (see Arvaniti 2011 for an extensive overview). On the one hand there are configurational models that treat intonation as holistic F0 contours with a specific meaning (e.g. Bolinger 1951, Jones 1972, Hirst & Di Cristo 1998). On the other hand, there are compositional models that claim that an F0 contour can be deconstructed into smaller units. However, the nature of these smaller units is controversial. While some researchers have worked with local F0 movements ('t Hart, Collier & Cohen 1990, Halliday 1970, O'Connor & Arnold 1973, Crystal 1972) others have claimed tonal targets in F0 space to be tonal primitives (e.g. Trager & Smith 1951, Pike 1945, Hockett 1955, Bruce 1977). One of today's prevailing approaches to the intonational analysis of the world's languages is provided by the Autosegmental-Metrical model, which views intonation as a sequence of single tonal targets (Pierrehumbert 1980, Beckman & Pierrehumbert 1986, Pierrehumbert & Beckman 1988, Ladd 2008).

The Autosegmental-Metrical model (henceforth: AM model) claims that the intonation of an utterance rests on a different tier, or level, than the segmental part. Both tiers itself host their own items or segments. Originally describing tone languages, Goldsmith (1976) proposes the two tiers to be one tonal and one phonemic tier, the former containing a "H" (high tone) or "L" (low tone) indicating changes in pitch height, and the latter containing "C" or "V" indicating consonantal or vocalic phonemes. The cohesion between the elements on the segmental tier and the elements on the tonal tier is indicated by association lines that "represent simultaneity in time" (ibid.: 10). Figure 2.1 illustrates this association. "X" denotes an autosegment.

As indicated by Goldsmith's (1976) use of high and low tones, the AM model is a target-based model in that rising or falling F0 contours are modeled as a sequence of low

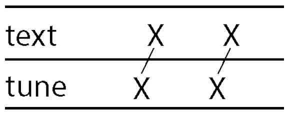


Figure 2.1: Autosegmental representation.

and high tonal targets corresponding to local minima (L) and local maxima (H) in the F0 trace. Thus, it appears that the intonation of a given utterance is underspecified:

> "[...] these L and H do not exhaustively represent the course of F0. Phonetically, the reflexes of L and H tones are tonal targets (typically, though not necessarily, F0 minima and and maxima respectively), with the pitch between them being generated by interpolation" (Arvaniti 2011: 267).

Another central part of the AM model is a hierachical organisation of speech units. That is, segments are integrated into syllables which themselves are dominated by higher prosodic units such as the foot or the phonological word. Figure 2.2 shows the prosodic organisation of the utterance <Two many cooks spoil the broth>.

Segments are grouped into syllables (σ). Syllabes are dominated by feet, (F) which themselves are grouped into phonological words (ω) and phonological phrases (φ). Phonological phrases, in turn, are integrated into Intonational Phrases (ι), ultimately constituting the utterance (υ). However, while the existence of very low levels (syllable, foot) and very high levels (utterance, Intonational Phrase) of the prosodic hierarchy are widely accepted, the intermediate levels and the criteria determining which units on one level are dominated by a higher-level unit is highly theory- and language-specific (see Shattuck-Hufnagel & Turk (1996) for an extensive discussion on this). Phonetic evidence for a hierarchical organisation of speech comes from a number of studies investigating acoustic and articulatory charateristics at the boundaries of prosodic constituents. In general, segments display longer durations before and after boundaries, caused by a slowdown

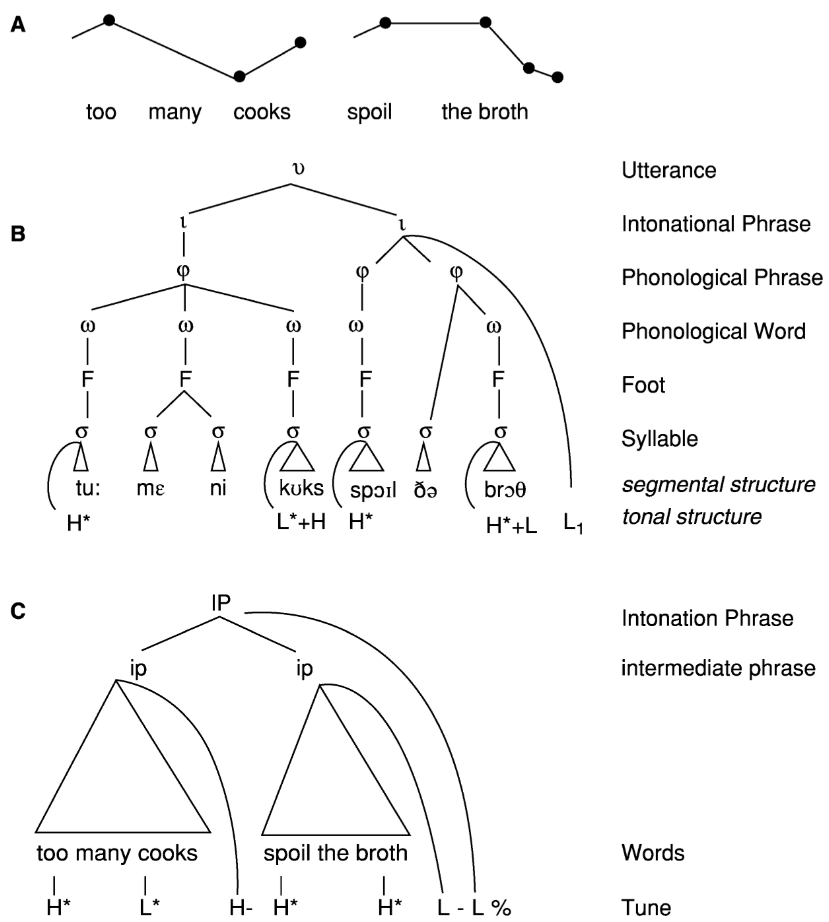


Figure 2.2: Prosodic organisation of the utterance <Too many cooks spoil the broth>. A) Schematized F0 contour, B) analysis from Gussenhoven 2002, C) analysis based on Pierrehumbert 1990 and Pierrehumbert & Beckman 1988 (reprinted from Grice 2006: 781).

of the articulators involved in processes known as final lengthening and domain initial strengthening (see Cho 2016 for an overview).

As for the tonal structure of an utterance, the AM model basically claims two different types of tones: pitch accents and edge tones, the latter often called boundary tones. While pitch accents are associated with a metrically stressed syllable, the tone bearing unit, boundary tones are associated with phrasal edges. In the both analyses given in Figure 2.2 (B and C), the intonation(al) phrase is marked by an obligatory boundary tone

(L_1 and L%, respectively). In addition, the analysis in (C) posits additional boundary tones associated with the edge of the intermediate phrase (H- and L-).

Pitch accents can either be mono- or bitonal and are marked by an asterisk. In addition, pitch accents are specified according to their position in the phrase. The last fully-fledged, and usually the most prominent, pitch accent is called the *nuclear accent*, while preceding pitch accents are called *prenuclear accents*.

The analysis given in Figure 2.2 (B) makes use of both mono- and bitonal pitch accents. Specifically, the syllables <too> and <spoil> are associated with a H* pitch accent, while the syllables <cooks> and <broth> are associated with L*+H and H*+L pitch accents, respectively. The analysis given in Figure 2.2 (C) uses only monotonal accents, i.e. H* and L*. The rise on the accented syllable <cooks> and the fall on the accented syllable <broth> is the result of the interpolation between the monotonal pitch accents and the following high and low boundary tones, respectively, attached to the intermediate phrases.

Next to pitch accents and boundary tones, there is also a phrase accent (often called phrase tone), that "is primarily an edge tone, but is realized on stressed syllables some distance from the edge of the phrase" (Grice 2006: 783). The phrase accent is associated with a phrasal edge but has a (secondary) association with the tone bearing unit.

Figure 2.3 presents the prosodic organisation of the Japanese utterance <Ane-no akai se'etaa-wa do'ko desu ka>. For Japanese, Pierrehumbert & Beckman (1988) posit an additional phrase below the intermediate phrase, namely the accentual phrase. Both the left and the right edge of the accentual phrase are associated with tones (H or L). However, these tones *can* have a secondary association with the tone bearing unit, i.e.

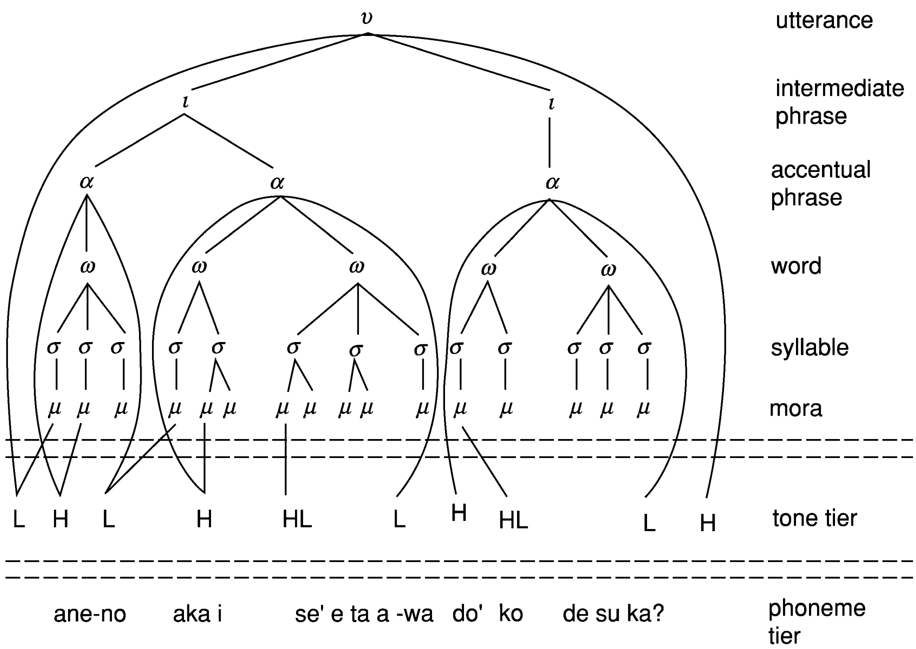


Figure 2.3: Prosodic organisation of the Japanese utterance <Ane-no akai se'etaa-wa do'ko desu ka> (reprinted with permission from Pierrehumbert & Beckman 1988: 21).

the mora in the case of Japanese.

Evidence for the phrase accent in European languages comes from a study conducted by Grice, Ladd & Arvaniti (2000). The authors closely examined what they called the "Eastern European Question Tune", which they analysed as "a low nuclear accent (L*) followed by a final rising-falling pitch movement [. . .] analysed as a sequence of a phrase accent (H-) and low boundary tone (L%)" (Grice, Ladd & Arvaniti 2000: 148). The occurrence of the H- phrase accent, however, is variable and depends on its proximity to the preceding nuclear L* accent. More specifically, the H- phrase accent is realised on the phrase-final syllable when the preceding nuclear L* accent is associated with a syllable in the same (phrase-final) word, but it can be realised earlier when the nuclear L* accent is associated with a syllable in a word preceding the phrase-final word. Due to the variable occurrence of the H- phrase accent, the authors conclude that

> “ [. . .] phrase accents are edge tones with a secondary association to an ordinary tone-bearing unit. These secondary associations can be to syllables which are at or near the periphery of the phrase (e.g. to the penultimate or final syllable). However, they can also be to syllables which are considerably distance [sic!] from the phrase edge, in which case the docking site is a stressed syllable” (Grice, Ladd & Arvaniti 2000: 180).

In a nutshell, the Autosegmental-Metrical model posits two different tonal primitives, namely L and H. These tones, or combinations of them, are the basis for pitch accents and edge tones. Pitch accents are associated with tone bearing units, for example, stressed syllables. Edge tones are associated with phrasal edges such as the intonational phrases. In addition, there are phrase accents, which have an association with both phrasal edges and tone bearing units. The next section will elaborate on the phonological concept of tonal association and its phonetic counterpart, tonal alignment, in more detail.

2.2 Tonal association and tonal alignment

Tonal association is concerned with the connection between units on the tonal tier and units on the phonemic, or CV, tier. Tonal association is a purely phonological concept and can be seen as the "abstract [...] property of 'belonging together' in some way" (Ladd 2008: 179). In contrast, tonal alignment refers to the actual temporal occurence of tonal events with respect to acoustically defined landmarks in the segmental string. Following the idea of a level-based understanding of intonation, these tonal targets correspond to low and high turning points in the F0 contour. Most of the studies conducted in the last decades have focused on the tonal alignment of pitch accents with their (phonologically) associated accented syllables, that is, the alignment of F0 peaks and/or valleys relative to segmental landmarks.

However, it must be noted that the relation between phonological association and phonetic alignment is not straightforward. For example, Arvaniti, Ladd & Mennen (1998) have shown that in Greek both the beginning and the end of a prenuclear accentual rise tends to align outside the accented syllable with which the pitch accent is associated. Instead, the beginning and end of the rise align at fixed distances before and after the accented syllable, respectively. This raises the question as to how to incorporate these findings into the phonological representation in terms of the association of pitch accents. However, Ladd (2008) points out that differences in the alignment of pitch accents do not necessarily entail different phonological associations and states that

> " [a]ssociation should be kept as a phonological concept, and detailed differences of alignment should generally not be given a direct phonological representation" (Ladd 2008: 179).

Rather, differences in tonal alignment of a pitch accent should be treated like gradual realisations of the same phonological concept. There are, however, studies showing that the alignment of pitch accents is indeed systematically affected by both phonological and phonetic factors. The majority of these studies address questions such as to what extent the alignment of pitch accents is affected by focus, syllable structure, speech rate, position within the intonation phrase and eventually the language under investigation. The following sections will explicitly focus on effects of syllable structure, word boundary and phrasal position. The goal of this section is twofold: It will be shown that, first, the end of an accentual rise, the F0 peak, is highly variable in its alignment with the segmental string and, second, that articulatory landmarks provide more adequate anchor points than segmental ones.

2.3 Tonal alignment with segments

This section reviews studies investigating the alignment of pitch accents with landmarks in the segmental string. It starts with an illustration of the *Segmental Anchoring Hypothesis* first mentioned in Ladd et al.'s (1999) seminal paper on prenuclear rising pitch accents in British English. It will be shown, firstly, that the duration of a pitch accent, i.e. the time interval between two phonologically specified tonal targets, is not fixed and, secondly, that tonal targets are attached to anchors in the segmental string such as the acoustic boundary between two segments. A number of studies will be reviewed that have challenged the segmental anchoring hypothesis by showing that syllable structure, vicinity to a word or phrasal boundary and accent status have an effect on the temporal alignment of tonal targets.

2.3.1 The Segmental Anchoring Hypothesis

The Segmental Anchoring Hypothesis (see also Arvaniti 2012, D'Imperio 2012, and Prieto 2011) states that a pitch accent's tonal targets are tied to landmarks in the segmental string, where

> " [...] the beginning and end of a linguistically significant pitch movement are anchored to specific locations in segmental structure, which means that the slope and duration of the pitch movement vary according to the segmental material with which it is associated" (Ladd 2006: 19).

These anchors can be segmental boundaries such as the acoustic boundary between the consonant and the vowel within a syllable or a time point within a segment such as the acoustic midpoint of a vowel. The phenomenon of *segmental anchoring* dates back to a study on prenuclear rising pitch accents in Greek. Based on previously published data (Arvaniti & Ladd 1995), Arvaniti, Ladd & Mennen (1998) found that low and high F0 turning points (L and H) in a prenuclear rising pitch accent were aligned at fixed distances to segmental boundaries irrespective of the acoustic duration of the accented and postaccented syllable. More specifically, L was aligned shortly before the acoustic onset of the accented syllable, whereas H was aligned shortly after the onset of the postaccented vowel following the accented syllable, "[...] even in the presence of large differences in the combined duration of the accented syllable and the immediately following consonant" (Arvaniti, Ladd & Mennen 1998: 24).

These findings led to another study in which the term *segmental anchoring* was introduced: For British English, Ladd et al. (1999) showed that the shape of a pitch accent is determined by speech rate and, as a consequence of the rate, by the acoustic duration of the segments of the target word on which the accent is produced. One of the aims of this study was to argue against what the authors call the “constant duration hypothesis” and the “constant slope hypothesis”. Specifically, they argue against considerations by Fujisaki (1983) who posits the F0 rise duration to have a fixed length (“constant duration”) and Ashby (1978) who argued in favor of a strong correlation between rise duration and rise excursion (“constant slope”). Another motivation for Ladd et al.'s study was to reinforce the assumption of a pitch accent as being composed of tonal targets whose alignment with the segmental string determines the shape of the pitch accent. Before reviewing Ladd et al's (1999) study in detail, the predictions of the “constant slope” and “constant duration” hypothesis will be presented.

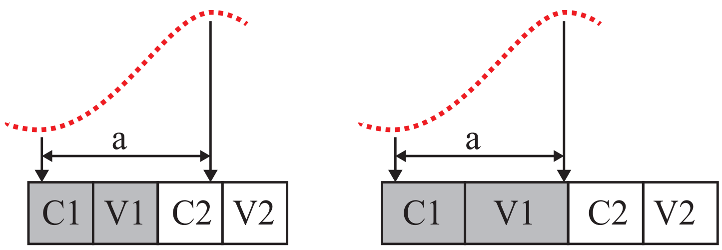


Figure 2.4: Predictions of the “constant duration hypothesis”: alignment of a rising pitch accent at normal speech rate (left-hand side) and slow speech rate (right-hand side), following Ladd et al. (1999).

Figure 2.4 gives an example of a rising pitch accent aligning with a CVCV target word at a normal speech rate (left-hand side) and at a slow speech rate (right-hand side), with longer acoustic segments. The accented syllable is shaded (C1 and V1). The rise duration is indicated by the letter “a” and, according to the “constant duration hypothesis”, remains constant between the two speech rates. In both examples the beginning of the accentual rise aligns shortly after the syllable onset. However, the end of the accentual rise aligns later at a normal speech rate than at a slow speech rate, though the rise duration remains constant. More specifically, the end of the rise aligns at the end of the postaccented consonant in the second syllable at a normal speech rate, whereas it aligns at the end of the accented vowel at the slow speech rate. A different outcome is predicted by the "constant slope hypothesis".

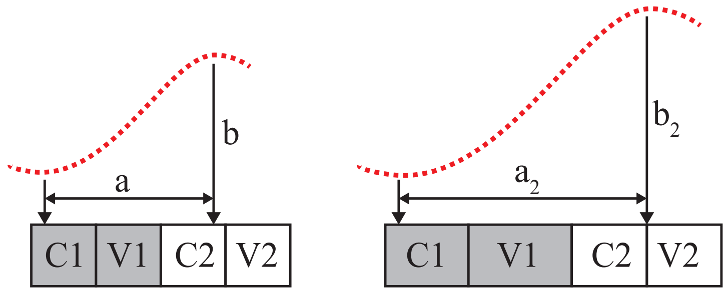


Figure 2.5: Predictions of the "constant slope hypothesis": alignment of a rising pitch accent at normal speech rate (left-hand side) and slow speech rate (right-hand side), following Ladd et al. (1999).

Figure 2.5 displays the predictions of the "constant slope hypothesis" at a normal speech rate (left-hand side) and a slow speech rate (right-hand side). The "constant slope hypothesis" posits that the ratio between the duration of the accentual rise and the pitch excursion of the accent (as indicated by letter "b") remains constant. In this way, a higher pitch excursion is predicted at a slower speech rate such that the ratios a:b and a_2:b_2 remain constant.

What Ladd et al. (1999) found was neither a constant duration nor a constant slope of the pitch accent. Instead, they report on fixed distances, or lags, between the tonal targets and segmental anchors in English prenuclear accents. Target words in the study displayed simple C1 V1 C2 V2 structures, where C1 and V1 denoted the accented syllable. First, they showed a strong positive correlation between the speech rate and the duration between the F0 turning points, thus rejecting the constant duration hypothesis. Second, they found no effect of speech rate on the pitch excursion, thus rejecting the constant slope hypothesis. Third, they found no effect of speech rate on the alignment of the low F0 turning point relative to the onset of C1, providing evidence that it is robustly aligned shortly before the beginning of the accented syllable. However, the authors do report on an effect of speech rate on the alignment of the *high* F0 turning point relative to the onset of the postaccented syllable C2 and relative to the onset of the postaccented vowel V1. Three out of six speakers aligned the F0 max later in the slow speaking rate condition. The authors then presumed that a segmental boundary might not be the anchor point for the F0 max. Instead, they compared the alignment of F0 max relative to the onset of the postaccented consonant C2 as a proportion of the duration of this consonant. This re-calculation of the data and the dismissal of a speaker from the set pointed to a stable, proportional alignment of F0 max. The authors conclude that

> " [...] there was clear evidence that both the beginning and the end of the accentual rise investigated are anchored to specific locations in the segmental string: the beginning (L) of the rise is timed to co-occur with the beginning of the onset consonant of the stressed syllable, and the end (H) is timed to occur somewhere late in the consonant following the stressed vowel" (Ladd et al. 1999: 1553).

The basic concept of tonal targets as anchored to segmental landmarks has gained a lot of attention in the last decades. While some studies on different languages have shown comparable results, there are a number of studies challenging the *Segmental Anchoring Hypothesis*, instead showing that the alignment of tonal targets is subject to a great deal of variation. These studies will be reviewed in the next sections.

2.3.2 Prosodic effects on tonal alignment

A seminal study on F0 peak alignment is Silverman & Pierrehumbert (1990), which investigated the alignment of prenuclear H* accents in American English. The speech material consisted of mono-, di- and trisyllabic target words, each with the prenuclear H* accent on the first syllable (<Ma>, <Mom>, <Mama>, <Mamalie>). On this basis, the authors varied the distance between the prenuclear H* accent on the target syllable <Ma> (or <Mom>) and an upcoming nuclear H+L* accent. This nuclear accent was produced either on the first syllable of the next word or after the injection of an another unstressed syllable. One example is the utterance <Ma Lemm> where the prenuclear H* accent on <Ma> was followed by a nuclear H+L* accent on <Lemm>. Another example was the utterance <Mama Le Mann> where prenuclear H* accent on <Ma> was followed by the nuclear H+L* accent on the syllable <Mann>. Utterances were categorised according to whether there was a word boundary after the prenuclear syllable <Ma> (<Ma> vs. <Mama> and <Mamalie>) and whether there was a stress clash in terms of the nuclear syllable adjacently following the prenuclear one (<Ma Lemm> and <Mom Lemm> vs. the other utterances).

Two speakers were asked to read the utterances at a slow, normal and fast speaking rate. The F0 peak alignment of the prenuclear H* accent was investigated by relating the F0 peak to the onset of the accented vowel /a/ as a function of the rhyme duration, i.e. the duration of /a/ in <Ma> or <Mamalie> and <om> in <Mom>, respectively. Essentially, the authors made three important findings. First, they found a correlation between F0 peak alignment and rhyme duration in that the F0 peak aligns later when the rhyme is longer. Second, the F0 peak was found to align earlier in monosyllables than in di- or trisyllables, even though the rhyme duration was longer, and third, the F0 peak was found to align earlier when the syllable bearing the nuclear accent adjacently followed the prenuclear one:

> "When a syllable is lengthened from being spoken more slowly, the peak will occur corresponding [sic!] later. In contrast, when the lengthening is induced

> by the right-hand prosodic context, the later part of the syllable undergoes disproportionately more lengthening and at the same time the peak will occur earlier in the syllable rhyme. In addition to this length-related effect, for one of the speakers a leftward push on the prenuclear peak is exerted by the upcoming nuclear pitch accent" (Silverman & Pierrehumbert 1990: 95).

Although the F0 peak alignment was affected by these prosodic factors, the authors point to a stable alignment of the F0 peak, when it is measured from the the vowel onset as a proportion of the rhyme duration. They conclude that "[...] it is not the absolute peak delay, but rather the peak placement in proportion to the syllable rhyme length, that exhibits the most regular pattern" and "that the effects [...] on the proportional peak alignment are independent of speaking rate" (Silverman & Pierrehumbert 1990: 87).

The finding that an accentual F0 peak is affected by so called "right-hand prosodic context" is corroborated by a study from Prieto, van Santen & Hirschberg (1995) on nuclear rising accents in Mexican Spanish. The authors investigated the alignment of both the beginning and the end of a nuclear H* rise in trisyllabic target words with stress on one of the three syllables. These target words were embedded in three different prosodic positions: in the middle of a phrase, before the end of an intermediate phrase and before the end of an intonational phrase. Furthermore, the distance between the nuclear syllable under investigation and the following stressed syllable was varied by changing the stress on the following word. Both the beginning and the end of the rise were measured relative to the onset of the accented syllable.

As a rule, the onset of the rise was not affected by the factors under investigation, indicating a stable alignment with the onset of the accented syllable. It "[...] was generally located precisely at the syllabic onset or just a few milliseconds into the onset"

(Prieto et al. 1995: 446). The F0 peak, however, was highly sensitive to the factors under investigation. More specifically, H aligned earlier before an intonational phrase boundary than in the middle of a phrase. It also aligned earlier in target words with stress on the final syllable as compared to target words with stress on the initial or middle syllable. The F0 peak also aligned earlier in syllables that were immediately followed by another stressed syllable leading to a stress clash.

Ladd, Mennen & Schepman (2000) investigated the effect of syllable structure on the alignment of F0 peaks in prenuclear rising accents in Dutch. In both of their experiments they measured the beginning of the prenuclear rise relative to the onset of the stressed syllable. The end of the rise, the F0 peak, was measured relative to the end of the stressed vowel. Speech material included target words with either phonologically long, or tense, vowels and phonologically short, or lax, vowels. As in other languages such as German, phonological vowel length correlates with syllable structure. In trochaic disyllables with a long stressed vowel, the consonant following the vowel is the onset of the second syllable, while in disyllables with a phonologically short stressed vowel, the following consonant is analysed as an ambisyllabic segment (i.e., it fills both the coda position of the first and the onset position of the second syllable).

Figure 2.6 illustrates the syllabification of the trochaic disyllabic words <Miete> (/mi:tə/) and <Mitte> (/mɪtə/), respectively. In /mi:te/ (left-hand side) the postvocalic consonant /t/ is syllabified as the onset (O) of the second syllable whereas in /mɪtə/ this consonant fills both the position of the coda consonant of the first syllable (C) and the onset of the second syllable. The reason for the consonant being analysed as ambisyllabic is that, in German, as rule, either the nucleus (N) or the rhyme (R) must branch (Hall 2011).[1]

[1]It must be noted that reduced vowels in unstressed syllables such as schwa are an exception to this rule.

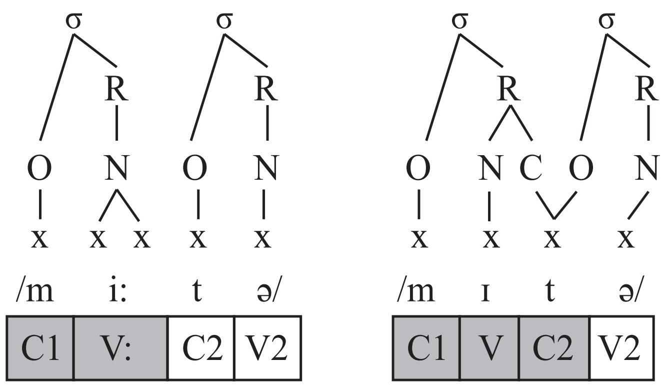


Figure 2.6: Syllabification of syllables with a phonologically long stressed vowel (left-hand side) and a phonologically short stress vowel (right-hand side)

In /mi:tə/ the long vowel /i:/ fills two positions (x x) on the skeletal tier thus enabling the nucleus to branch. In /mɪtə/ the short vowel /ɪ/ fills only one position (x) on the skeletal tier. As a result, the rhyme branches and incorporates the postvocalic consonant /t/ as a coda consonant.

Usually, the difference between phonologically long and short vowels involves different phonetic vowel durations, with phonologically short vowels exhibiting a shorter intrinsic vowel duration. In Dutch, however, long /i:/ and short /ɪ/ have the same phonetic duration. This fact was one of reasons behind Ladd et al's (2000) study in that the authors wanted to show that a difference in F0 peak alignment was not due to a phonetic adjustment of the vowel duration, but rather to the phonological distinction between long and short vowels. They hypothesized that the anchor point of the F0 peak is the syllable edge.

Their first experiment with different vowel qualities such as /y:/, /ʏ/, /o:/ and /ɔ/ confirmed that the beginning of the prenuclear rise was aligned shortly before the onset of the stressed syllable and that phonological length difference did not influence its alignment. More specifically, L aligned 3 ms before the syllable onset in syllables with a long vowel, and 1 ms before it with a short vowel. In contrast, H was affected by the phonological vowel length.

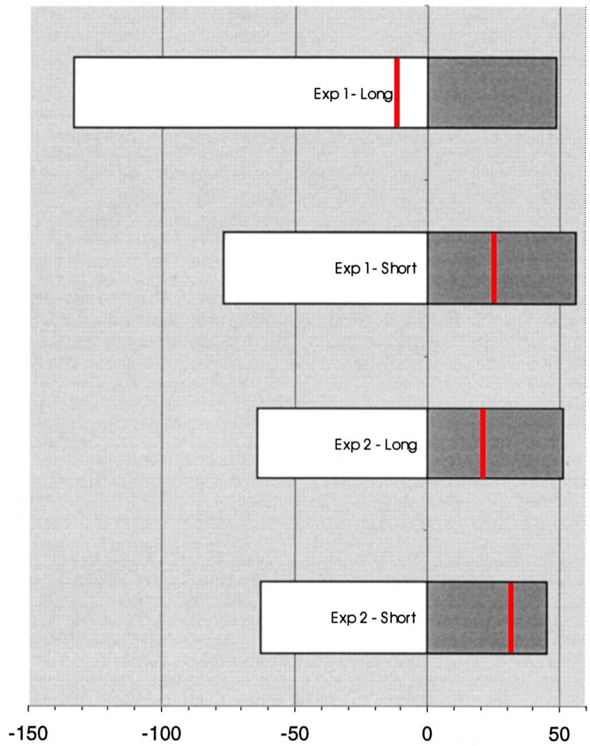


Figure 2.7: F0 peak alignment in syllables with phonologically long and short vowels (reprinted with permission from Ladd, Mennen & Schepman 2000: 2692).

Figure 2.7 displays the alignment of H (red line) in relation to the acoustic offset of the accented vowel (white box). In long vowels, H aligned at the end of the stressed vowel (12 ms before the end of the vowel) and in short vowels, it aligned within the following consonant (25 ms after the end of the vowel). This result led to a second experiment in which the authors only used target words involving /iː/ and /ɪ/ whose phonetic durations were not significantly different. In contrast to the expectation, in both conditions, the F0 peak aligned *after* the vowel. Crucially, the F0 peak aligned later in the short vowel /ɪ/ than in the long vowel /iː/. More specifically, H aligned 32 ms after the onset of the postaccented consonant in /ɪ/ and 21 ms after this landmark in /iː/. These findings were against the assumption that in phonologically long vowels, the F0 peak aligns at the syllable edge, i.e. the end of the vowel. The explanation that the authors offer is that the phonologically long vowel /i:/ does not provide enough time for the prenuclear rise to be completed. They state:

> "[W]hen the actual duration of the 'long' vowel is relatively short [...], the end of the rise is aligned later than the end of the vowel, though still earlier than the end of the rise accompanying a short vowel" (Ladd et al. 2000: 2693).

In addition to the study by Ladd, Mennen & Schepman (2000), Wichmann, House & Rietveld (2000) provided evidence for the F0 peak being strongly affected by the position of the syllable in the phrase. The authors investigated rising pitch accents produced by speakers of Southern Standard British English. Speakers had to read a text in which each of the four target words[2] were placed in "paragraph-initial", "sentence-initial" and "sentence-final" position. In paragraph-initial position, the target word was also utterance-initial. In sentence-initial position, the target word was preceded by an unstressed syllable. In sentence-final position, the target word was utterance-final.

The authors measured the proportional alignment of the F0 peak relative to the onset of the accented syllable. In both paragraph-initial and sentence-initial position the F0 peak aligned *after* the accented syllable, with the F0 peak aligning even later in paragraph-initial position. In sentence-final position, the F0 peak aligned earlier, namely within the accented syllable.

Figure 2.8 depicts their data on the F0 peak alignment. The accented syllable is shaded. The location of the F0 peak (H) is indicated by the red arrow. In both paragraph-initial

[2]Target words included di-, tri- and tetrasyllabic words with stress either on the first or second syllable: <carTEsian>, <COMmon>, <comPENdium>, <enLIGHtenment>.

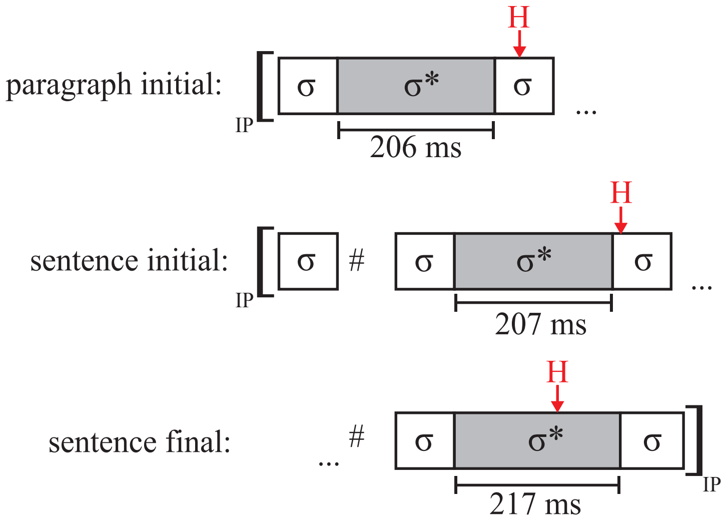


Figure 2.8: Illustration of Wichmann, House & Rietveld's (2000) data.

and sentence initial position, the F0 peak aligns shortly after the acoustic offset of the accented syllable. More specifically, it aligns 32 ms and 11 ms after the offset, respectively. In sentence final position, the F0 peak is retracted into the accented syllable and aligns 82 ms *before* the syllable offset.

Schepman, Lickley & Ladd (2006) investigated the alignment of nuclear rising pitch accents in Dutch. The purpose of this study was, on the one hand, to disentangle the effects of stress clash and word length. On the other hand, the authors wanted to unscramble the effects of syllable structure and phonological vowel length previously reported in Ladd et al. (2000). Thus, Schepman et al. (2006) designed a corpus including mono- and disyllabic target words, each having either a phonologically long or short stressed vowel. Target words were embedded into carrier phrases and were followed by either a stressed monosyllable or a disyllable with stress on the second syllable. The beginning of the nuclear rise was measured relative to the onset of the accented syllable. The end of the nuclear rise was measured relative to the end of the accented vowel. This landmark was the nearest one to the F0 peak.

As expected, the beginning of the nuclear rise was stably aligned with the onset of the accented syllable, ranging from 3 ms before the syllable onset in phonologically short vowels to 5 ms after the syllable onset in phonologically long vowels. Suprisingly, the alignment of the F0 peak was *not* affected by syllable structure nor by the "right hand" context, which involved a stress clash when the accented syllable was immediately followed by another stressed syllable. Figure 2.9 displays the data for the F0 peak alignment from Schepman et al. (2006).

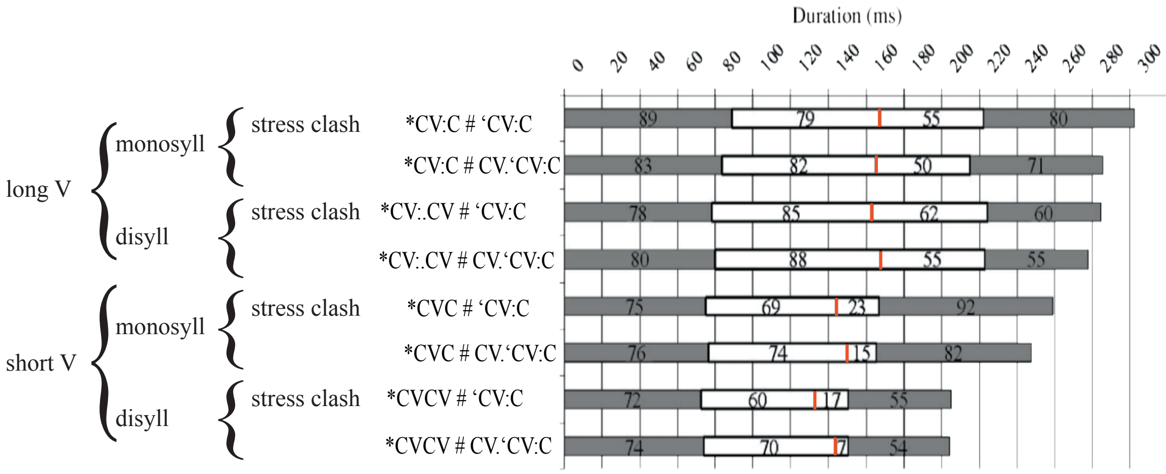


Figure 2.9: Alignment of nuclear F0 peaks as a function of prosodic structure (extended and reprinted with permission from Schepman, Lickley & Ladd 2006:11).

The red bars indicate the first and third segments, the consonants, of the target words. The white bar indicates the accented vowel, with the line representing the F0 peak. In all cases, the F0 peak aligned within the accented vowel and was only affected by the phonological vowel length. In target words with a long vowel, the F0 peak aligned on average 56 ms before the end of the vowel, whereas in target words with a short vowel, the F0 peak aligned on average 16 ms before the end of the vowel. The structure of the target word or the presence of stress clash induced by a syllable immediately following the accented one did not play a significant role in determining the alignment of the F0 peak. The authors conclude that "there is a fairly fixed pattern of alignment for long vowels and another for short vowels" (Schepman et al. 2006: 15).

These findings are in contrast with those reported by Silverman & Pierrehumbert (1990) and Prieto, van Santen & Hirschberg (1995) in that the F0 peak was not affected by the right-hand prosodic context in case of a stress clash. Schepman et al.'s (2006) explanation for this absent effect is based on the finding that nuclear accents as a whole align earlier possibly due to a following a low phrase accent. They state that

> "[...] the right context established by the seperate intonational event (phrase accent [...]) is the most important factor determining the alignment of the peak in nuclear accents, important enough that it overrides the subtle effects of syllable structure and stress clash" (Schepman et al. 2006: 17).

The finding that phonological vowel length plays a crucial role in determining F0 peak alignment was corroborated by a study from Ladd et al. (2009) in which the authors investigated F0 peak alignment in both prenuclear and nuclear rising pitch accents in two varieties of English (Scottish Standard English vs. Received Pronunciation). Again, their corpus consisted of target words with phonologically long and short stressed vowels such as /iː/ and /ɪ/. For prenuclear accents, the authors were able to clearly show that the F0 peak globally aligned earlier in syllables with a long vowel than in syllables with a short vowel. More specifically, the F0 peak aligned 27 ms before the vowel offset in long vowels and 19 ms *after* the vowel offset in short vowels. In contrast, the beginning of the prenuclear rise was not affected by any of the factors under investigation. It aligned shortly after the onset of the syllable.

The follow-up experiment on nuclear accents[3] yielded similar results. Here, monosyllabic

[3] I am referring to their third experiment, as in their second experiment, the authors could not reliably

target words were placed in phrase-final position. The F0 peak alignment was measured from both the onset and the end of the accented vowel. In phonologically long vowels, the F0 peak aligned 111 ms before the vowel offset. In phonologically short vowels, it aligned 90 ms before the vowel offset. This indicates that in long vowels, the F0 peak aligned earlier. However, phonological vowel length had an effect the actual vowel duration in that long vowels were produced with a longer duration. The F0 peak aligned 37 ms after the vowel onset in long vowels and 26 ms after the vowel onset in short vowels, indicating in earlier alignment for short vowels. The authors, however, concede that an effect of phonological vowel length can be masked when the accented syllable is produced in phrase-final position, suggesting "[...] that any underlying effect of vowel length may be overridden by the need to align the nuclear peak early enough to execute the phrase-final fall" (Ladd et al. 2009: 158).

Effects of syllable structure on F0 peak alignment were further supported by a study from Prieto & Torreira (2007). The authors investigated the alignment of prenuclear rising pitch accents in Peninsular Spanish. The beginning of the rise was related to the onset of the stressed syllable while, the end of the rise was related to two different landmarks in the segmental string: the offset of the stressed vowel and the syllable boundary. While the beginning of the rise displayed a stable alignment "[...] realized within 20 ms of the onset of the syllable" (Prieto & Torreira 2007: 491), the end of the rise was affected by syllable structure. In target words with an open syllable, such as in <PaLOma>, the vowel offset and the syllable offset coincide, while in target words with a closed syllable, such as <BelMONdo>, the syllable offset coincides with the end of the coda consonant. With respect to the vowel offset, the F0 peak aligned later in closed syllables than in open syllables, i.e. it aligned *after* the vowel offset. With respect to the syllable boundary, the

identify the F0 peak.

F0 peak aligned earlier in a closed syllable, i.e. it aligned *before* the syllable boundary, within the coda consonant. The authors conclude that their results "[. . .] demonstrate that tonal H turning points are not anchored at acoustic segmental landmarks such as the vocalic or the syllabic offsets" (Prieto & Torreira 2007: 491).

Rathcke & Harrington (2007) report on syllable structure and word boundary effects in German nuclear accents. Their corpus included target words with open and closed stressed syllables (entailing different phonological vowel lengths) produced either in phrase-final position or followed by one or two unstressed syllables (such as <Lie>, <Liener>, <Lienerer> and <Linn>, <Linner> and <Linnerer>). A professional speaker then produced a H* accent on the stressed syllable.[4] The proportional alignment of the F0 peak was investigated by means of its alignment relative to the syllable onset divided by the syllable duration.

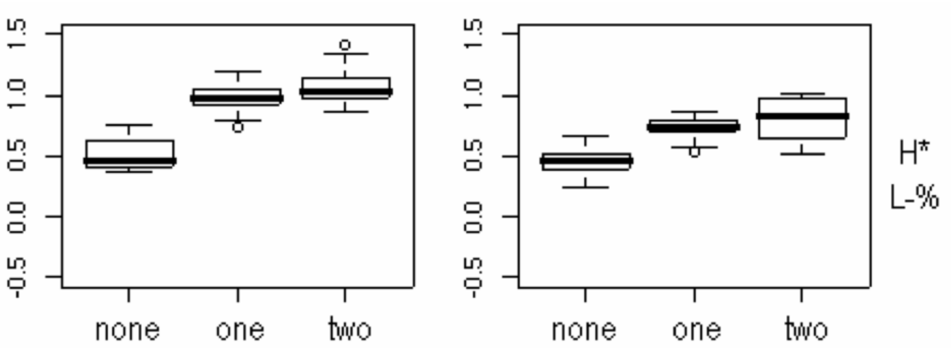


Figure 2.10: Alignment of nuclear F0 peaks as a function of syllable structure and postnuclear syllables (reprinted with permission from Rathcke & Harrington 2007: 984).

Figure 2.10 displays the data from Rathke & Harrington (2007), showing the F0 peak alignment of the H* accent in open (left-hand side) and closed syllables (right-hand side) as a function of the number of postnuclear syllables (from none to two). Zero denotes the

[4]The authors investigated both a H* and H+L* accent, but I will focus on the H* as it is a rising accent and thus more comparable to previous research.

syllable onset while one denotes the syllable offset. The authors state that the F0 peak aligned later in open syllables and that it aligned later in target words followed by one or two unstressed syllables. However, the authors report on an interaction between syllable struture and the number of postnuclear syllables. Although they do not report any posthoc comparisons, a visual inspection of the data suggests that there is no difference F0 peak alignment between the accented syllables followed by one or two postnuclear syllables belonging to the same word.

Similar results were obtained by Mücke & Hermes (2007), who investigated nuclear rising L+H* accents in German. Their corpus was based on two speakers of Austrian German (Vienna variety) included monosyllabic and trochaic disyllabic target words produced with either a long or short stressed vowel, i.e. /a/ vs. /a:/. Target words were placed in both phrase-medial and phrase-final position. The F0 peak alignment was measured relative to the end of the target word. Figure 2.11 displays their data.

Confirming previous work, the accentual F0 peak aligned earlier in phrase-final target words than in phrase-medial target words, as indicated by the dashed arrow. In addition, the authors report on an earlier F0 peak alignment in long vowels than in short vowels in both mono- and disyllabic target words.

Kleber & Rathcke (2008) investigated the alignment of prenuclear rising accents in German (Upper Saxon variety). Based on 12 speakers, their corpus, an extension of Atterer & Ladd (2004), involved target words in phrase-initial position with the stressed syllable being either the very first syllable or preceded by one or two unstressed syllables (e.g., <**Man**gelhafte ...>, <Die **man**gelhaften ...>, <In Er**man**gelung ...>. As in Rathcke & Harrington (2007) both the beginning and the end of the prenuclear rise were measured

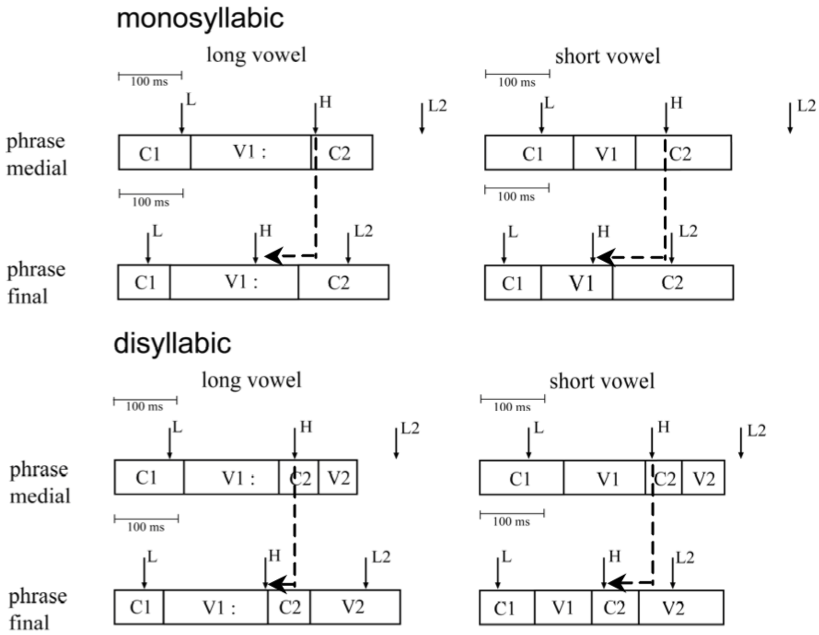


Figure 2.11: Alignment of nuclear rising accents as a function of vowel length and phrasal position (reprinted with permission from Mücke & Hermes 2007: 999).

as a proportion in relation to the syllable onset. Figure 2.12 shows their alignment data as a function of number of syllables preceding the stressed syllable. Zero denotes the onset of the syllable. While the beginning of the rise (left-hand side) was aligned earlier when the stressed syllable was preceded by one or two syllables[5], in contrast, the end of the rise (right-hand side) was not affected by the number of syllables preceding the stressed one.

Mücke et al. (2009) directly compared the alignment of nuclear and prenuclear rising pitch accents in two varieties of German. Based on two subjects from each variety (Düsseldorf and Vienna), their corpus included trochaic disyllabic target words with either an open or closed stressed syllable (e.g., /ˈma:mi/ vs. /ˈmami/). The accentual F0 peaks were

[5]The alignment difference between target words having one vs. two preceding syllables was not significant.

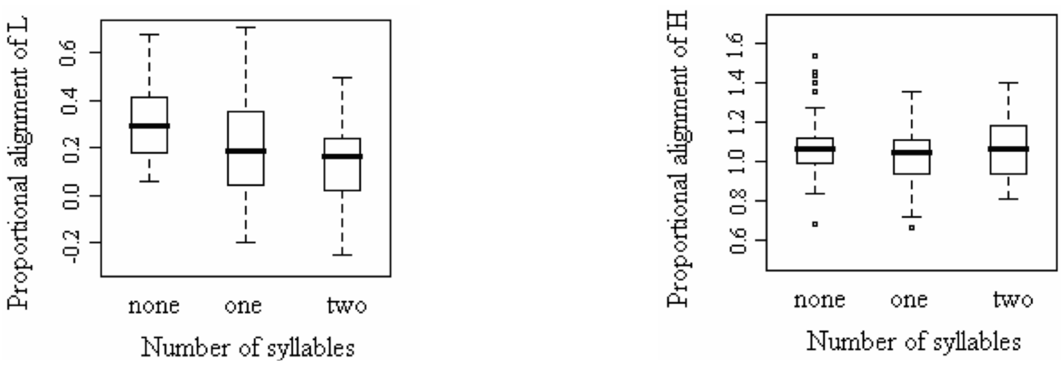


Figure 2.12: Alignment of L and H in prenuclear rising pitch accents in German, Upper Saxon variety (reprinted with permission from Kleber & Rathcke 2008: 8).

measured relative to their nearest segmental landmarks, i.e. the prenuclear F0 peak was related to both the onset and the offset of the postaccented vowel, and the nuclear peak was related to both the onset and the offset of the postaccented consonant.

With respect to the accent status, the nuclear F0 peaks regularly aligned earlier than the prenuclear F0 peaks. The nuclear F0 peaks aligned well within the postaccented consonant while the prenuclear F0 peaks aligned within the postaccented vowel. With respect to syllable structure, the nuclear F0 peaks aligned significantly later in closed than in open syllables, i.e. in both syllable structures, the F0 peak aligned in the postaccented consonant but later when the stressed syllable was closed. The same effect could not be ascertained for the prenuclear accents in that only one of the four speakers aligned the F0 peak later in closed syllables.

Figure 2.13 displays data from Mücke et al. (2009) for two speakers. In both open and closed syllables the F0 peak aligned within the postvocalic consonant C2, although slightly later in closed syllables. More strikingly, the F0 peak aligned consistently earlier in nuclear F0 peaks than in prenuclear F0 peaks, as indicated by the shift.

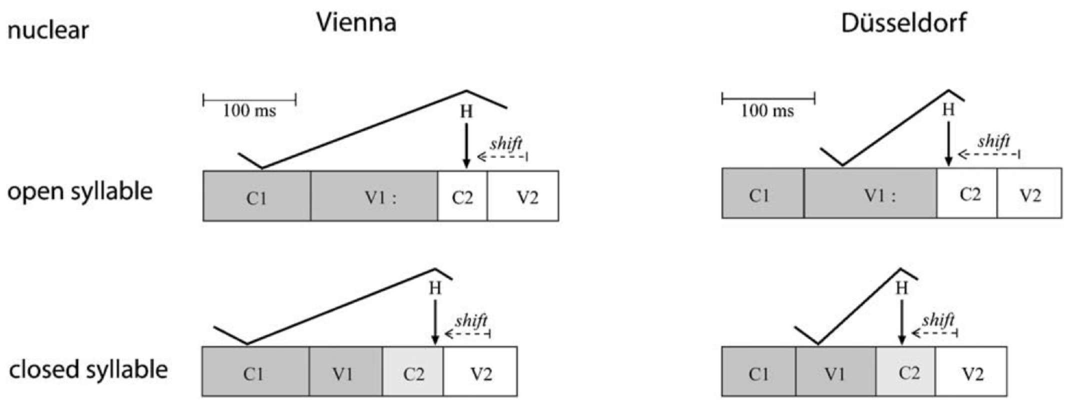


Figure 2.13: Alignment of nuclear F0 peaks as a function of syllable structure and accent status. The shift illustrates an earlier alignment of nuclear F0 peaks compared to prenuclear ones (reprinted with permission from Mücke et al. 2009: 334).

This finding is perfectly in line with research from Silverman & Pierrehumbert (1990). Although they do not compare the alignment of prenuclear and nuclear F0 peaks in their study, they relate their findings on prenuclear accents to work carried out by Steele (1986) who investigated the alignment of nuclear F0 peaks and state that

> "One difference remains between our data on prenuclear H* accents and Steele's data for the same accents in nuclear position; namely that peaks are absolutely earlier" (Silverman & Pierrehumbert 1990: 96).

So far, it has been shown that the the beginning of a rising pitch accent shows a rather stable alignment with the onset of the stressed syllable while the end of the rise, the F0 peak, is subject to a great deal of variation. In a nutshell, there is evidence that accentual F0 peaks tend to align earlier in open than in closed syllables. It seems that in open syllables the F0 peak tends to align at the end of the stressed vowel, whereas in closed syllables it tends to align in the coda consonant. F0 peaks also tend to align considerably earlier in phrase-final position as compared to in non-final positions in the phrase. A great alignment difference can also be found when comparing nuclear to prenuclear rising pitch accents in that nuclear F0 peaks align considerably earlier than prenuclear ones. More subtle effects have been found between mono- and disyllabic target words in that in the former the F0 peak aligns slightly earlier.

Recent studies have attempted to find more stable timing patterns between tonal targets and landmarks in the articulatory domain. The following section will address these studies in detail.

2.4 Tonal alignment with articulatory gestures

This section reviews studies that have examined the alignment of tonal targets with respect to landmarks in the articulatory domain, looking specifically at the timing of F0 peaks with consonantal gestures articulated with the lips and the tongue tip. These studies have attempted to find stable alignment lags with articulatory landmarks, such as the maximum lip opening, that are not affected by prosodic factors. In other words, researchers have sought find articulatory landmarks that serve as more stable anchors to F0 peaks. This section is divided into two parts: the first examines the alignment of F0 peaks with nearby articulatory gestures, the second examines F0 peak alignment with non-nearby articulatory gestures.

2.4.1 Nearby landmarks

D'Imperio et al. (2007) investigated the effects of speech rate, syllable structure and accent type on the alignment of nuclear rising pitch accents in Neapolitan Italian. Previous studies (D'Imperio 2000, 2002) have already shown that both yes/no questions and statements are produced with a rising pitch accent but that yes/no questions involve a later F0 peak alignment relative to the acoustic vowel onset than statements. Pitch accents in yes/no questions are thus analysed as L+H* while pitch accents in statements are analysed as L*+H.

Based on data from one speaker, the corpus involved trochaic target words with open and closed syllables such as /ˈma.ma/ and /ˈmam.ma/ that were embedded in carrier phrases either produced as a yes/no question or as a statement. Both renditions were recorded at a normal and at a fast speaking rate. In the acoustics, the F0 peak was related to the onset and offset of the stressed vowel. In statements, the F0 peak was found to align at the end of the vowel while in questions the F0 peak aligned later, within the following consonant. The effect of speech rate was not clear as with respect to vowel onset, the F0 peak in statements remained stable while in questions it aligned earlier at a fast speech rate. With respect to the vowel offset, the F0 peak alignment remained stable in questions while it was affected by speech rate in statements. In addition, alignment lags were rather large. The authors conclude that

> "[a] stable, albeit large, temporal alignment of a tonal and segmental event would not constitute sufficient evidence for a synchronization of laryngeal and supralaryngeal gestures" (D'Imperio et al. 2007: 580).

Thus, the F0 peak was related to two nearby articulatory landmarks: the early F0 peak in statements (L+H*) was measured relative to the peak velocity of the postvocalic consonant's closing gesture; the late F0 peak (L*+H) in yes/no questions was measured relative to the maximum constriction following the peak velocity of this gesture. The key finding was that both F0 peaks roughly co-occured with these landmarks in the articulation, i.e. the early F0 peak co-occured with the peak velocity of the consonantal closing gesture while the late peak co-occured with the following maximum constriction. The upper panel of Figure 2.14 displays the alignment of the F0 peak in statements relative to the peak velocity, while the lower panel displays the F0 peak in yes/no questions relative to the maximum constriction, i.e. the maximum lip closure in the postvocalic consonant.

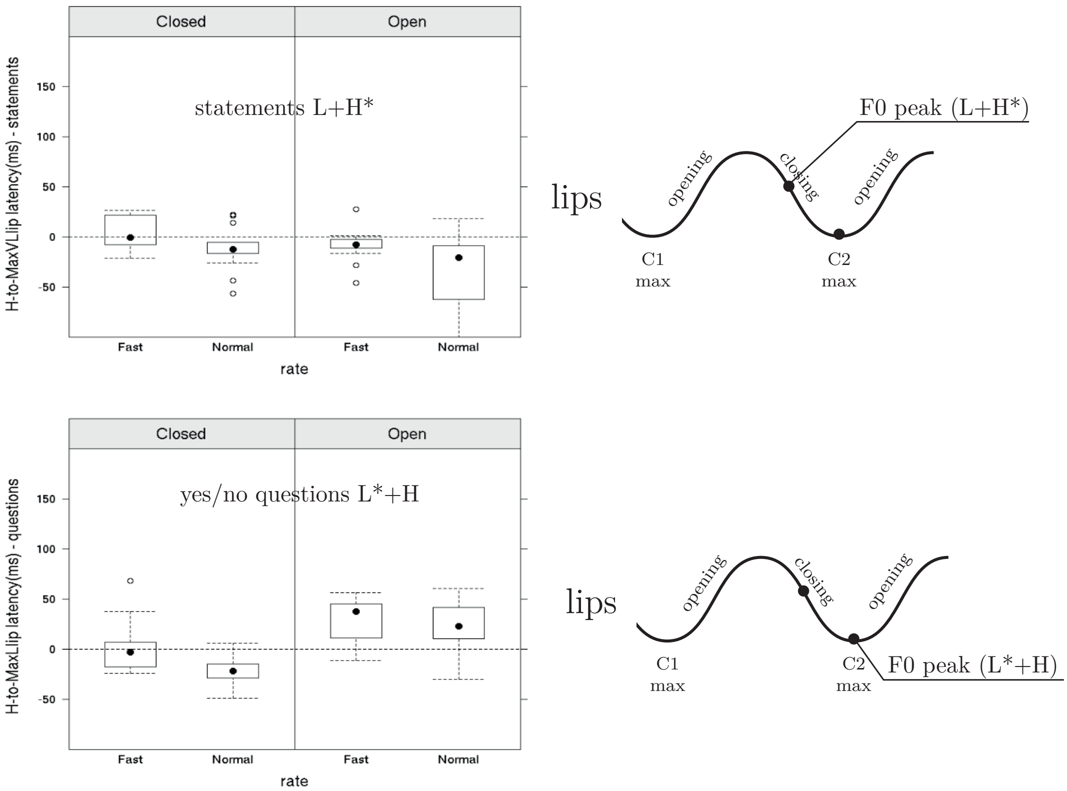


Figure 2.14: F0 peak alignment relative to the peak velocity (upper panel) and maximum constriction (lower panel) of the postvocalic consonant, labial data only (extended and adapted from D'Imperio et al. 2007: 591)

However, speech rate had an effect on these alignment patterns as in statements the F0 peak aligned slightly later at a fast rate. The authors compared their articulatory alignment data with the data for acoustic alignment and found that "all the results confirmed the observation that articulatory latencies are smaller than acoustic ones." (D'Imperio et al. 2007: 592). The authors, however, highlight that the lags between the articulatory landmarks and the F0 peaks were not any more stable than the acoustic ones:

> "H targets of nuclear rises in Neapolitan statements appear to be more closely, though not more stably, synchronized with the articulatory dimension of peak velocity within the trajectory of the primary constrictor than with two of the most commonly employed acoustic segmental landmarks" (D'Imperio et al. 2007: 592).

Similar results were obtained by Prieto et al. (2007), who looked at the alignment of rising nuclear pitch accents in Catalan. Based on data from one speaker, their corpus included target words with open and closed syllables such as /mi.ma.mi/ and /mi.mam.zi/. In the acoustics, the F0 peak was related to the syllable offset. In open syllables, the F0 peak aligned roughly with the syllable edge, i.e. it aligned with the vowel offset. In closed syllables the F0 peak aligned earlier with respect to the syllable edge; more specifically, it aligned in the coda consonant following the stressed vowel. In the articulation, the F0 peak was related to nearby landmarks in the articulation, namely the peak velocity and the maximum constriction of the consonantal closing gesture following the stressed vowel. The key finding was that the F0 peak aligned in a time window between these two articulatory landmarks. As in D'Imperio et al. (2007), alignment lags between the F0 peak and articulatory landmarks were smaller than between F0 peaks and landmarks in the segmental string. However, F0 peaks aligned later in target words with a closed syllable than with an open syllable.

Mücke & Hermes (2007) investigated the aligment of phrase-medial and phrase-final nuclear rising accents in German. For all their test conditions (long vs. short vowel, mono- and disyllabic target words), they found that the F0 peak aligned well within the consonantal closing gesture, i.e. the lip closing gesture in the consonant following the stressed vowel.

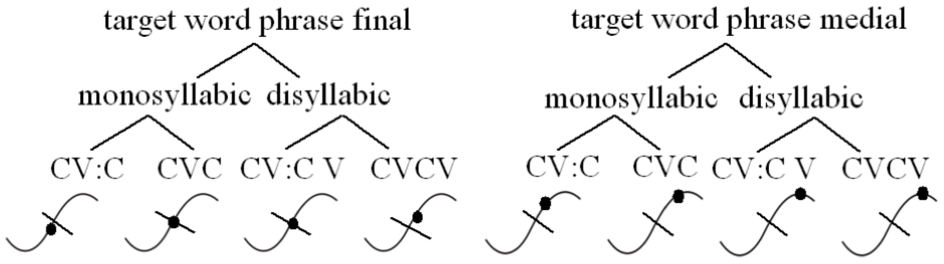


Figure 2.15: F0 peak alignment relative to lip closing gesture (reprinted with permission from Mücke & Hermes 2007: 1000).

Figure 2.15 presents the data from Mücke & Hermes (2007). The F0 peak (the black dot along the S-shaped curve) was related to the consonantal closing gesture following the stressed vowel. As a rule, the F0 peak aligned earlier in phrase-final position than in phrase-medial position, and earlier in monosyllables than in disyllables.

A more detailed analysis of German rising accents is provided by Mücke et al. (2009) who investigated the timing of both nuclear and prenuclear rising accents in trochaic disyllabic target words with an open or closed stressed syllable. The accentual F0 peaks were related to their nearest articulatory landmarks. The alignment of nuclear F0 peaks was measured relative the consonantal target in the unstressed syllable, corresponding to the maximum lip/tongue tip closure. As prenuclear F0 peaks aligned later, they were measured relative to the following maximum opening of the consonantal gesture ("transvocalic target"). For nuclear accents, the effect of syllable structure found in the acoustics (see Section 2.3 was still present in that the F0 peak aligned later in closed syllables. For prenuclear accents, the attested effect of syllable structure diminished.

Figure 2.16 summarises the findings from D'Imperio et al. (2007), Prieto et al. (2007),

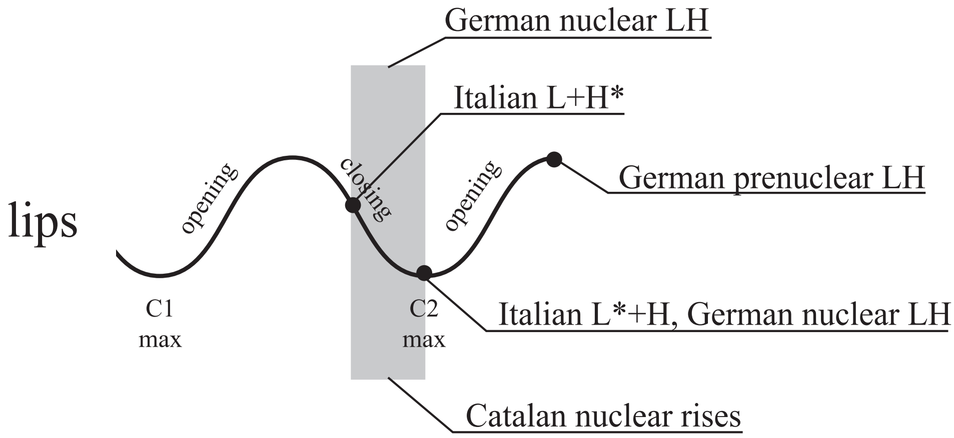


Figure 2.16: F0 peak alignment relative to consonantal constriction gestures. Italian data from D'Imperio et al. (2007), Catalan data from Prieto et al. (2007), German data from Mücke et al. (2009) and Mücke & Hermes (2007)

Mücke & Hermes (2008) and Mücke et al. (2009). All nuclear F0 peaks align well within a time window between the peak velocity and the maximum constriction gesture for the postvocalic consonant. These studies have shown that alignment lags between the F0 peak and articulatory landmarks tend to be smaller compared to those relative to acoustic landmarks, however the alignment is not necessarily more stable. For example, syllable structure and phrasal position have an effect on the alignment lag between the F0 peak and the articulatory landmarks.

2.4.2 Non-nearby landmarks

A different approach was taken by Mücke, Grice & Hermes (2008). Based on data from two speakers from Mücke et al. (2009), they investigated the alignment of accentual F0 peaks relative to both nearby and what they call "non-nearby landmarks". More specifically, the F0 peak was aligned to the consonantal gesture it was produced with (nearby landmark) as well as to the transvocalic target (non-nearby landmark); this transvocalic target corresponds to the maximum opening of the consonantal gesture during the accented vowel.[6]

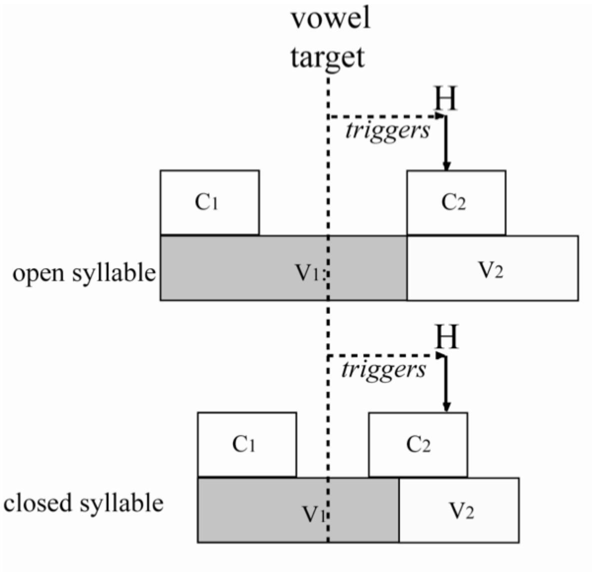


Figure 2.17: F0 peak alignment relative to articulatory gestures. The dashed line indicates the transvocalic target, i.e. the maximum opening of the consonantal gesture (reprinted with permission from Mücke, Grice & Hermes 2008).

As can be seen in Figure 2.17, the F0 peak aligns well within the postvocalic consonantal gesture for C2 in both open and closed syllables. The F0 peak, however, aligns later in closed syllables than in open syllables with respect to this landmark. Surprisingly, this alignment difference between open and closed syllables disappeared when the F0 peak was related to the transvocalic target, which instead indicated a stable timing pattern. The authors state that "the effect [of syllabe structure] disappears for H relative to the underlying gestural representation" (Mücke, Grice & Hermes 2008).

Even though the alignment lag between the F0 peak and the transvocalic target was larger than that between the F0 peak and the onset of the postaccented consonant, it

[6]This articulatory landmark roughly corresponds to the vocalic target derived from the tongue body movement for the production of the vowel.

was not affected by syllable structure and was thus found to be more stable. This finding runs against D'Imperio et al.'s (2007) claim that only nearby landmarks can serve as an anchor for tonal events. However, the interplay between the tonal tier and the text tier can be captured in coordinatory terms, that is, tonal targets must not necessarily co-occur with articulatory landmarks to indicate a coupling between the two systems. Rather, a variable alignment of tonal targets could be the result of specific underlying coordination between the tone and articulatory gestures. Evidence for this comes from a number of studies claiming tonal targets to be the onset and offset of a so-called tone gestures whose coordination with articulatory gestures can be explained in terms of coupling structures with oral constriction gestures.

The next chapter will introduce the basic concepts of Articulatory Phonology and will review studies on tonal alignment treating tones, be they pitch accents or lexical tones, as tone gestures.

3 Articulatory Phonology

This Chapter provides an overview of Articulatory Phonology (see e.g., Browman & Goldstein 1986, 1989, 1992), a framework that treats articulatory gestures as fundamental phonological units. This chapter is divided into four sections. The first introduces the basic concepts of Articulatory Phonology. The main focus of the second section is the dynamical nature of gestures, which has been modelled by employing nonlinear dynamics. The third part focuses on the relationship between intergestural timing patterns and low-level prosodic units such as the syllable by presenting the coupled oscillator model of syllable structure. The fourth section illustrates how high-level prosodic units involving prosodic phrasing, tone and intonation have been successfully implemented into the framework of Articulatory Phonology.

3.1 Basic concepts

Classical approaches to phonology such as generative phonology (Chomsky & Halle 1968) assume a gap between the phonology and the phonetics of a given language. That is, phonology and phonetics operate on two different levels, or domains, as phonology is treated as being discrete and phonetics as being continuous, with a sharp distinction made between the two domains. Thus, these approaches require an interface between the two domains, i.e. a kind of translator that takes the phonological primitive, the segment, as its input and transforms it via rule-based routines into actual speech output. In contrast, Articulatory Phonology integrates phonetics and phonology in one single system by stating that “these apparently different domains are [...] in fact the low and high dimensional description of a single (complex) system” (Browman & Goldstein 1995: 177).

According to Articulatory Phonology, gestures rather than segments are the primitives of a given language. These gestures are isomorphous in that they are cognitive and physical at the same time: on the one hand they are abstract phonological entities specifying phonological contrast and, on the other hand, they are concrete units of action performed by the articulators:

> “Gestures are characterizations of discrete, physical real events that unfold during the speech production process. Articulatory Phonology attempts to describe lexical units in terms of these events and their interrelations, which means that gestures are basic units of contrast among lexical items as well as units of articulatory action” (Browman & Goldstein 1992: 156).

Articulatory gestures are linked to the human articulators in a functional way, that is, they do not describe the actual movement of an articulator but rather the linguistic task that is assigned to it. For example, a speech task could be a full oral closure between the tongue tip and the alveolar ridge and at the same time a lowering of the velum to produce the alveolar nasal consonant /n/, or raising and fronting of the tongue body and a simultaneous protrusion of the lips to produce the vowel /y/. In order to fulfill the speech task, a set of variables is involved, called *tract variables*.

Figure 3.1 displays the five tract variables implemented in Articulatory Phonology (left-hand side) and their linked organ groups (right-hand side). Tract variables are specified either one- or two-dimensionally.

tract variable		articulators involved
LP	lip protrusion	upper & lower lips, jaw
LA	lip aperture	upper & lower lips, jaw
TTCL	tongue tip constrict location	tongue tip, body, jaw
TTCD	tongue tip constrict degree	tongue tip, body, jaw
TBCL	tongue body constrict location	tongue body, jaw
TBCD	tongue body constrict degree	tongue body, jaw
VEL	velic aperture	velum
GLO	glottal aperture	glottis

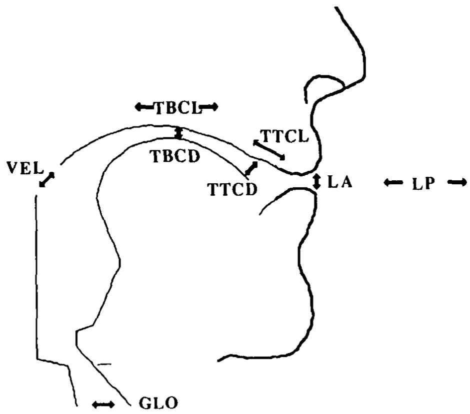


Figure 3.1: Tract variables in Articulatory Phonology (reprinted with permission from Browman & Goldstein 1990: 344).

The velum (VEL) and the glottis (GLO) are aech specified one-dimensionally, where the velic aperture controls for nasality and the glottal aperture controls for voicing. The lips, the tongue tip and the tongue body are specified two-dimensionally. The lips are controlled for degree of protrusion (LP) and aperture (LA); for example, the lips are protruded in the case of /y/ but are retracted in the case of /i/, and the lip aperture is greater in the case of /a/ than in the case of /i/. The tongue tip and the tongue body are each controlled for their constriction location (TTCL and TBCL, respectively) and their constriction degree (TTCD and TBCD, respectively) each. The production of the velar fricative /x/, for example, involves a constriction between the tongue body and the velum in order to generate aerial friction. A change in degree of constriction could result in full closure thus switching from /x/ to /k/.

In Articulatory Phonology the structure of a linguistic unit, a syllable or a word, is represented in gestural scores involving the activation of gestures corresponding to their tract variables. In other words, a gestural score reflects the workflow of specific speech tasks assigned to the articulators.

> "[...] we represent linguistic structures in terms of coordinated articulatory movements called gestures, that are themselves organized into a gestural score that resembles an autosegmental representation" (Browman & Goldstein 1990: 341).

Two gestural scores are presented in Figure 3.2 for the utterances <mad> and <ban>, respectively. The tract variables velum, tongue tip, tongue body, lips and glottis are displayed on the right-hand side. The gestural activation is indicated by the grey boxes. For the production of /m/ in /mad/ the lips need to be closed ("bilabial closure") while at the same time the velum needs to be lowered ("wide"). For the production of the vowel /a/, the tongue body needs to be lowered ("pharyngeal wide"). The final consonant /d/ is produced by a full closure between the tongue tip and the alveolar ridge ("alveloar closure"). Both utterances /mad/ and /ban/ involve the same number of articulatory gestures. The crucial difference, however, is down to the timing between the gestures. In /mad/, the velic gesture controlling for nasal airflow is timed with the bilabial gesture; together they produce the word-initial /m/. In /ban/, the velic gesture is timed with the alveolar closure, thus producing the word-final /n/. It must be noted, however, that the temporal coordination of this velic gesture with the alveolar closure is still a matter of research. There is evidence for both a simultaneous and sequential timing of the two gestures (see Goldstein, Byrd & Saltzman 2006 for a discussion).

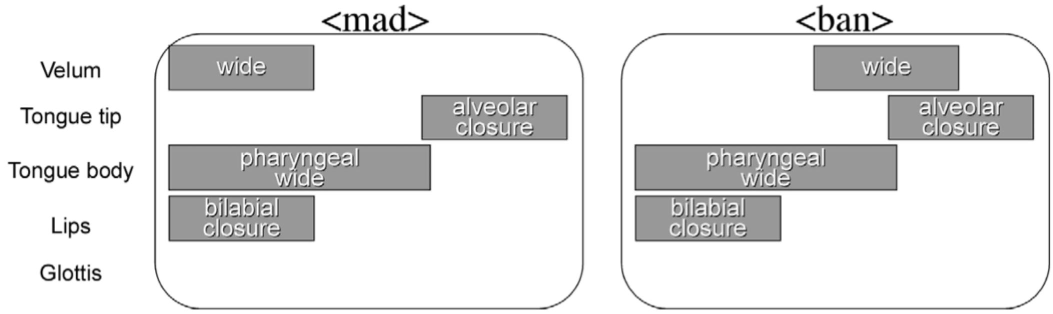


Figure 3.2: Gestural scores for the utterances <mad> and <ban> (reprinted with permission from Goldstein, Byrd & Saltzman 2006: 226).

These gestural scores illustrate another central part of Articulatory Phonology: Lexical contrast is *not* specified in terms of phonologial processes such as elision (deleting segments), epenthesis (adding segments) or methathesis (swapping segments). Rather, it is the timing between articulatory gestures (intergestural timing) that specifies lexical contrast. Before intergestural timing and its relation to syllable structure are explained later on in Section 3.3, the next section will introduce the concepts of nonlinear dynamics. More specifically, it will be shown that the employment of potential functions gives rise to the phonological grammar and eventually to the linguistic structure of a given language.

3.2 Gestural dynamics

Similar to more traditional phonological approaches, Articulatory Phonology differentiates between vocalic and consonantal units by "[. . .] posit[ing] two functional tiers, a vocalic one and a consonantal one" (Browman & Goldstein 1990: 351). In gestural terms, these tiers refer to vocalic and consonantal gestures. The gestural distinction between vowels and consonants dates back to analyses by Öhman (1966) who investigated formants and formant transitions in VCV sequences such as /agi/ or /ida/ in American English speakers. More specifically, Öhman (1966) showed that the frequency of the second formant (F2) of a given vowel depends on the vowel preceding or following it. In the example /agi/, /i/ shows a lower F2 than it would show if produced in isolation. From an articulatory point of view, this indicates that the tongue gesture for /i/ in /agi/ is slightly retracted, thus lowering F2. Vice versa, /a/ shows a *higher* F2 than it would show in isolation. Metaphorically speaking, Öhman (1966) concluded that vowels behave like a rubber band with consonants being put on it.

> "a VCV utterance [. . .] can *not* be regarded as a linear sequence of three successive gestures. We have clear evidence that the stop-consonant gestures are actually superimposed on a context-dependent vowel substrate that is present during all of the consonantal gesture" (Öhman 1966: 165).

The description of articulatory trajectories derived from consonantal and vocalic gestures involves the application of task dynamics, "[. . .] a general model of skilled movement control that was developed originally to explain nonspeech tasks such as reaching and standing upright" (Hawkins 1992: 9). Task dynamics (Saltzman 1985, Saltzman & Kelso 1987, Saltzman & Munhall 1989, Saltzman 1995) account for the inherent dynamics of gestures and give rise to their spatial dimensions by applying nonlinear mathematics:

> "[...] gestures are dynamical in that their model employs the mathematics of linear and nonlinear dynamics. The model takes the form of an invariant mathematical law. This law is hypothesized to give rise to the continuous movement of gestures" (Gafos & Beňuš 2006: 915).

Hence, the dynamics of a gesture can be captured by a differential equation of a critically damped mass-spring system given in Equation 3.1.

$$m\ddot{x} + b\dot{x} + k(x - x_0) = 0 \tag{3.1}$$

where m represents the mass associated with a given tract variable, b represents the system's damping and k represents the stiffness of the spring. The current position of the tract variable is given by x whereas x_0 represents its target, for example, a constriction at the alveolar ridge for the production of a plosive /t/. The spring's velocity ($\dot{x}$) and its acceleration $\ddot{x}$ can be derived from x.

In Articulatory Phonology, the timing between articulatory gestures is crucial to their linguistic analysis. Studies from non-speech research fields (see Turvey, Schmidt & Rosenblum 1989 and Turvey 1990 for overviews) have already shown that the most stable timing pattern between two coordinated movements (e.g., oscillatory interlimb coordination) is the in-phase timing, that is, where movements start simultaneously. Movements can also easily coordinated in anti-phase mode, that is, where they start sequentially. However, as speed increases, movements that are coordinated anti-phase, eventually switch into in-phase coordination, indicating that the in-phase mode is a stronger attractor for coordination.

Figure 3.3 displays the time course of experiments reported in Kelso (1981), Kelso (1984) and Kelso & Scholz (1985). Subjects had to oscillate their left and right index fingers in opposite directions, i.e. while one index finger was raised the other one had to be lowered simultaneously. The speed of the fingers oscillating was set by a metronome becoming faster over the time course of the experiment. Figure 3.3 (A) shows the displacement,

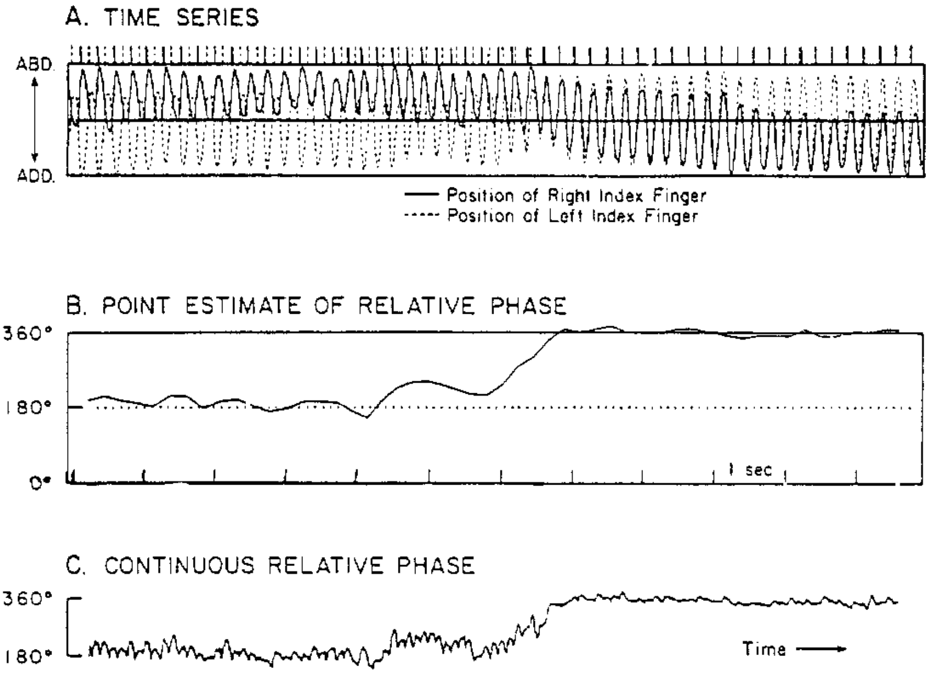


Figure 3.3: Time course of two index fingers transitioning from anti-phase to in-phase (reprinted with permission from Kelso et al. 1987: 80).

or position, of the two index fingers. While the right index finger was raised (ABD: abduction), the left index finger was lowered (ADD: adduction). Figure 3.3 (B) and Figure 3.3 (C) provide the point estimate of the relative phase and the continuous relative phase, respectively. The relative phase is the difference between the two phases of the index fingers. In the beginning, the relative phase amounts to 180°, i.e. both index fingers are oscillating in anti-phase. As speed increases, the index fingers abruptly switch to an in-phase mode. That is, their relative phase amounts to 360° (or 0°). The result is an overlap between the position of the right and left index fingers in Figure 3.3 (A).

This coordinated behaviour of the two index fingers (the two oscillators) has been modelled in terms of a potential function (Haken-Kelso-Bunz potential function) capturing the transition from anti-phase to in-phase oscillations (Haken, Kelso & Bunz 1985). Figure

3.4 displays three possible outcomes of the HKB potential function given in Equation 3.2.

$$V(\psi) = -a\cos(\psi) - b\cos(2\psi) \tag{3.2}$$

The variable ψ represents the relative phase between the two oscillators, i.e. the difference between the two phases ($\phi_2 - \phi_1$). The relationship between the variables a and b determines the shape of the potential function, i.e. they serve as the control parameter to the function. In all three examples given in Figure 3.4 the control parameter b amounts to 1, while a changes. Specifically, a amounts to the value 1 in the example on the left-hand side, to the value 2 in the middle and to the value 4 on the right-hand side.

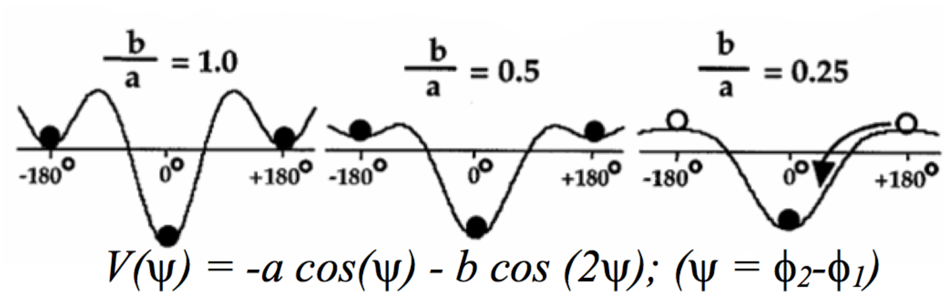


Figure 3.4: Landscapes of the HKB potential function (adapted from Nam, Goldstein & Saltzman 2009: 304).

The graph on the left-hand side displays three local minima at −180°, 0° and 180°. However, the minimum, or basin, at 0° is lower than at −180° or 180°. In other words, the 0° phasing is a stronger attractor than the −180° or 180° phasing. Applied to the index fingers oscillating, this indicates that at a slow rate the fingers can be oscillated both in-phase (both fingers up and down the same time, 0°) or in anti-phase (fingers go up and down sequentially, −180° or 180°). As speed increases, represented by increasing the value of a in the given example, the graph changes its form. The basins at $-180°$ and 180° flatten out, indicating that the index fingers are more likely to oscillate in-phase. On the right-hand side of Figure 3.4 there are no longer any basins for $-180°$ and 180°, i.e. the index fingers only oscillate in-phase. In other words, there is a critical shift from anti-phase to in-phase. Rosenbaum (2009) employs the metaphor of a marble to illustrate the transition from in-phase to anti-phase mode. In an initial state the marble (the black dot in Figure 3.4) can roll into the basins at $-180°$, 0° or 180° (left-hand side). As the shape shifts, there is only one basin left at 0° into which the marble can roll (right-hand side).

Potential functions such as the HBK potential function have not only been used to model intergestural timing but have also been applied to more general linguistic phenomena such as vowel harmony in Hungarian. By means of electromagnetic articulatory recordings, Gafos & Beňuš (2006) investigated the displacement of vocalic gestures, i.e. the horizontal tongue body movement accompanying the production of vowels. Vowel harmony refers to the phenomenon that in some languages vowels in suffixes have similar or even the same phonetic features as the vowels in the word stem. An example given by Gafos & Benus is the use of the dative suffixes /nak/ and /nek/. In traditional phonological terms, Hungarian uses the feature [back] to differentiate between front and back vowels. Front vowels have the feature [-back] whereas back vowels have the feature [+back]. These features *spread* to adjacent vowels in suffixes; a word stem with a vowel specified as [+back] will have a suffix containing a vowel that has a [+back] specified vowel, too. Thus, the suffix /nak/ occurs with word stems having a back vowel such as /ház/, resulting in /ház-nak/, whereas the suffix /nek/ occurs with word stems having a front vowel such as /öröm-nek/. However, there are so-called transparent vowels that “may intervene between the trigger and the target vowel even when they bear the opposite value for the harmonizing feature" (Gafos & Beňuš 2006: 925). In other words, a vowel specified as [-back] can occur between the stem vowel and the suffix vowel, although both are specified as [+back]. An example would be /buli-nak/, where the intervening vowel /i/ is specified as [-back], whereas the stem vowel /u/ (the trigger) and the suffix vowel /a/ (the target) are specified as [+back]. However, the authors point out that even though these transparent vowels are perceptually equivalent (i.e., the /i/ in /buli-nak/ and the /i/ in /bili-nek/ sound the same), they may have different underlying articulatory characteristics. The authors assume that front vowels occuring between front vowels are more fronted than front vowels occuring between back vowels and that both vowels can captured by using nonlinear dynamics. They hypothesize that:

> "[...] transparency emerges from nonlinearities in the relation between articulation and sound. In a nutshell, we hypothezise that the /i/ in /zafir-ban/ [...] is retracted articulatory as compared to /i/ in zefir-ben [...], but that this retraction falls within that limited region of articulatory variation that does not result in any significant acoustic consequences. If this hypothesis is correct it would provide the basis for a principled understanding of the cooccuring of two properties of the phenomenon, the nature of the harmonizing parameter (tongue body retraction) and the set of transparent vowels in Hungarian (/í, i, é, e/)" (Gafos & Beňuš 2006: 915).

Their articulatory data from three speakers showed that the tongue body is indeed more fronted for a transparent vowel that was followed by a suffix with a front vowel as compared to a transparent vowel that was followed by a suffix with a back vowel. In other words, more fronted transparent vowels are likely to trigger suffixes with a vowel specified as [-back] whereas less fronted vowels are likely to trigger suffixes with a [+back] vowel. To model the continuous tongue body movement that leads to the choice of suffixes with either [+back] or [-back] vowels, the authors employed the potential function given in

Equation 3.3.[1]

$$V(x) = Rx - \frac{x^2}{2} + \frac{x^4}{4} \tag{3.3}$$

where R represents the control parameter, i.e. the degree of retraction of the tongue body while producing the transparent vowel. Figure 3.5 displays this potential function with the control parameter R ranging between 0.2 and 1.2. The x-axis represents the horizontal displacement (backness and frontness) of the tongue body movement for the transparent vowel.

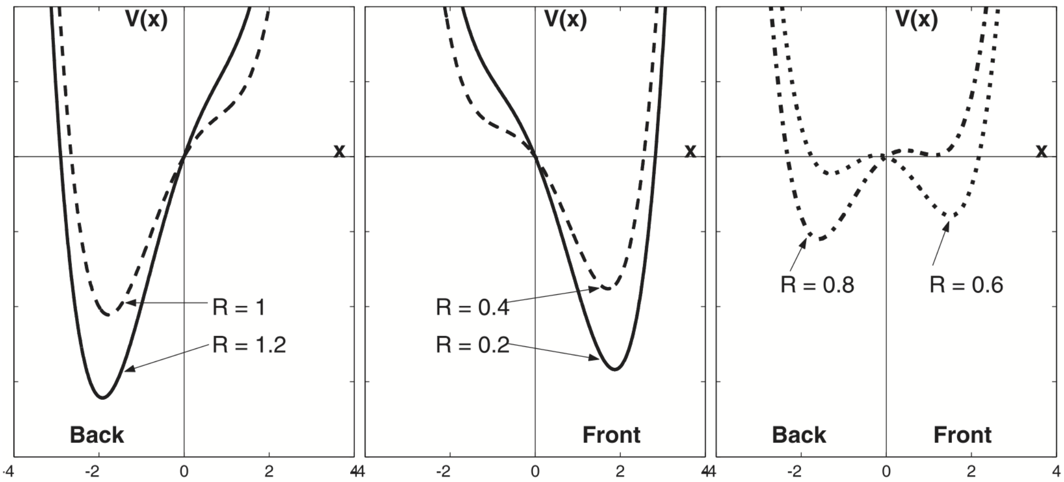


Figure 3.5: Landscapes of the potential function predicting the selection of either [+back] or [-back] vowels in Hungarian suffixes (reprinted with permission from Gafos & Beňuš 2006: 934).

The plots on the left-handside and in the middle show monostable systems, i.e. one basin in either the back or the front region. In the system of the left-hand side, a suffix with a back vowel is chosen, or triggered, as its attractor is situated in the back region of the x-axis. The opposite is true for the system in the middle of Figure 3.5. Here, the

[1]It must be noted that Gafos & Beňuš 2006 make use of the additional parameter λ referring to Tuller et al. 1994 who introduced this parameter to account for covariates in the experimental conditions.

attractor is situated in the front region, thus a suffix with a front vowel is most likely to occur. Particularly, the potential function can also form a bistable system (see right-hand side of Figure 3.5) to account for the occurence of both back or front vowels in suffixes.

3.3 Coupled Oscillator model of syllable structure

Within the coupled oscillator model of syllable structure, gestures, whether consonantal or vocalic gestures, are associated with nonlinear oscillators, or clocks, that are coupled in a pairwise fashion with each other to determine intergestural timing (Goldstein et al. 2008, Nam & Saltzman 2003). The model states that intergestural timing, determined by the coupling of consonantal and vocalic gestures, gives rise to the syllable structure (Goldstein, Chitoran & Selkirk 2007, Goldstein et al. 2008, Nam, Goldstein & Saltzman 2009).

Crucially, this intergestural coupling remains syllable-internal, i.e. the coupling of consonantal and vocalic gestures only arises within the syllable. However, it should be noted that it is not the vocalic and consonantal gesture themselves that are being coupled. Rather, their associated oscillators are coupled with each other. Along the line of the experiments conducted on interlimbic coordination (see Section 3.2.), Articulatory Phonology proposes two coupling modes: in-phase and anti-phase, i.e. gestures start either simultaneously or sequentially.

An example of a coupling structure and its relation to syllable constituents is given in Figure 3.6 displaying the gestural timing in the word /bɪt/. For reasons of clarity, only oral gestures are shown in the gestural score and the coupling structure. The

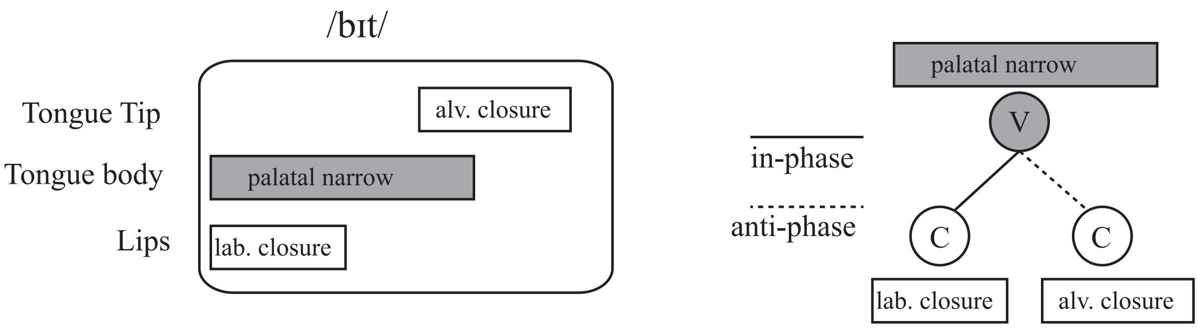


Figure 3.6: Gestural scores (left-hand side) and corresponding coupling structure (right-hand side) for the syllable /bIt/ (oral gestures only).

gestural score on the left-hand side provides the timing of the articulatory gestures. The consonantal gesture for /b/, *labial closure*, starts at the same time as the vocalic gesture for /ɪ/, *palatal narrow*. However, the consonantal gesture has a higher degree of stiffness, thus achieving its target earlier than the vocalic gesture, ensuring that both sounds /b/ and /ɪ/ are are perceptually recovered. The following consonantal gesture for /t/, *alveolar closure*, is initiated when the vocalic gesture has nearly completed its execution. The right-hand side of Figure 3.6 displays the coupling structure for the articulatory timing in /bɪt/. Crucially, the coupling structure correlates with the syllable structure in that the nucleus is coupled in-phase with the onset while it is coupled anti-phase with the coda consonant. In this example, the vocalic gesture for the nucleus /i/ (V) is coupled in-phase with the onset consonant /b/ (C) while it is coupled anti-phase with the coda consonant /t/ (C). This coupling structure leads to the timing displayed in the gestural score.

Figure 3.7 illustrates the outcome of an in-phase coupling by means of articulatory gestures involved in the production of the word /lina/ taken from the utterance <Er geht mit der Lina viel lieber> (*He goes with Lina preferentially*). The upper trajectory displays the vertical position of the tongue tip while the lower trajectory displays the vertical position of the tongue body. Both the consonantal gesture for /l/ (the syllable onset) and the vocalic gesture for /i/ (the syllable nucleus) start at same time. The starting point is indicated by the black dot. However, the consonantal gesture reaches its target earlier than the vocalic target, leading to the perceptual sequence of /l/ and /i/.

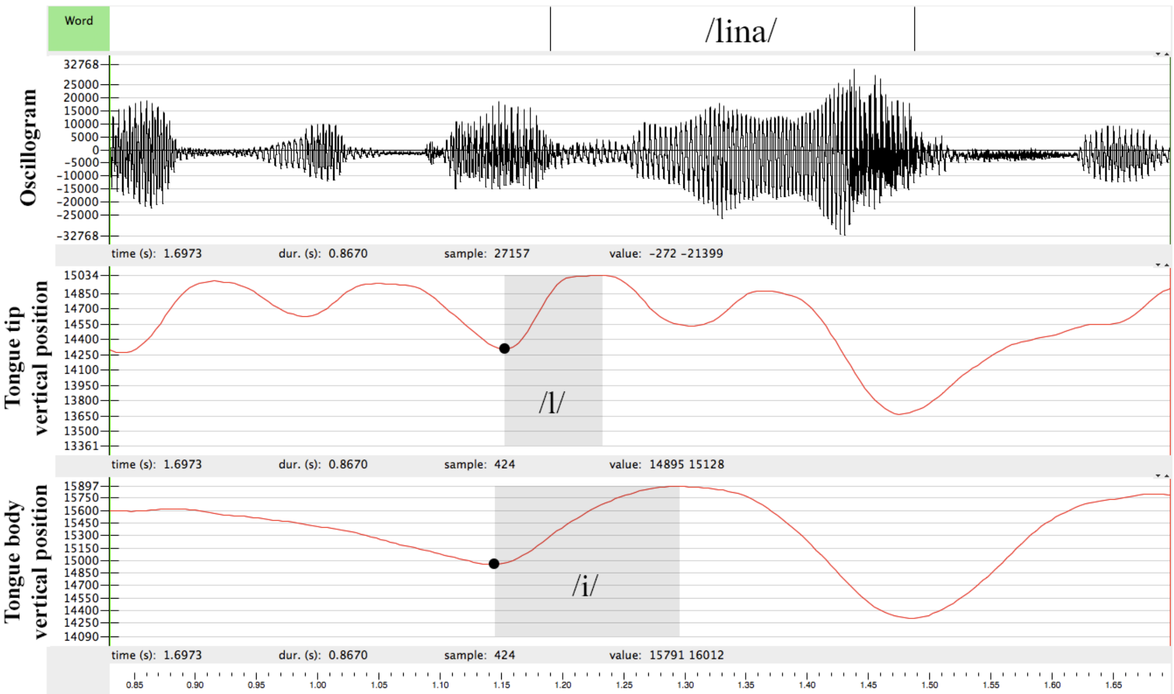


Figure 3.7: Trajectories derived from sensors on the tongue tip (upper trajectory) and the tongue body (lower trajectory).

When another consonant is added to the syllable, the intergestural timing changes as a function of whether this consonant is added to the left edge (filling another onset position in the syllable hierarchy such as in CCV) or the right edge (filling another coda position in the syllable hierarchy such as in VCC). Browman & Goldstein (1988) investigated the relative timing of articulatory gestures in American English. Their target words involved structures such as /pi # C(CC) ats/ with the target either produced with one, two or three consonants. In addition, the word boundary was sometimes shifted resulting in the structure /piC # C(C) ats/. For both conditions two line-up points, or anchors, were defined. The midpoint of the /p/ closure in /pi/ (or /piC/) served as the left anchor, while the articulatory target of /t/ in /ats/ served as the right anchor. The authors calculated three temporal intervals: the distance or duration between the left edge, the right edge and the c-center, i.e. the temporal midpoint of all consonantal targets. With respect to the right anchor, the authors found that the c-center of the consonant involved in the sequence /pi # C(CC) ats/ showed the lowest standard variation. With respect to the left anchor, the left-edge showed the lowest standard variation. The authors conclude that pre- and postvocalic consonants are timed differently and state that

> "[...] while initial consonants are related to their words in terms of a single global metric for the entire cluster, the C-center, final consonants appear to be related to their words in terms of the local metric of achievement of targets [...] rather than in terms of C-centers" (Browman & Goldstein 1988: 149).

While prevocalic consonants, i.e. consonants that form the syllable onset, are timed globally, postvocalic consonants, i.e. consonants that form the syllable coda, are timed locally in that their left edge aligns with the preceding anchor. What Browman & Goldstein found in syllable onsets was dubbed the *c-center effect* describing "the fact that, as consonants are added to onsets, the resulting timing of all consonant gestures changes with respect to the following vowel in a way that preserves the overall timing of the center of the consonant sequence with respect to the vowel" (Saltzman et al. 2006: 69).

Two different coupling structures have been proposed for the timing patterns of consonant clusters in complex onsets and codas, respectively. Figure 3.8 shows the syllabification (upper panel), the gestural scores (middle panel) and the underlying coupling structures (lower panel) for the syllables /spɪt/ and /tɪps/.

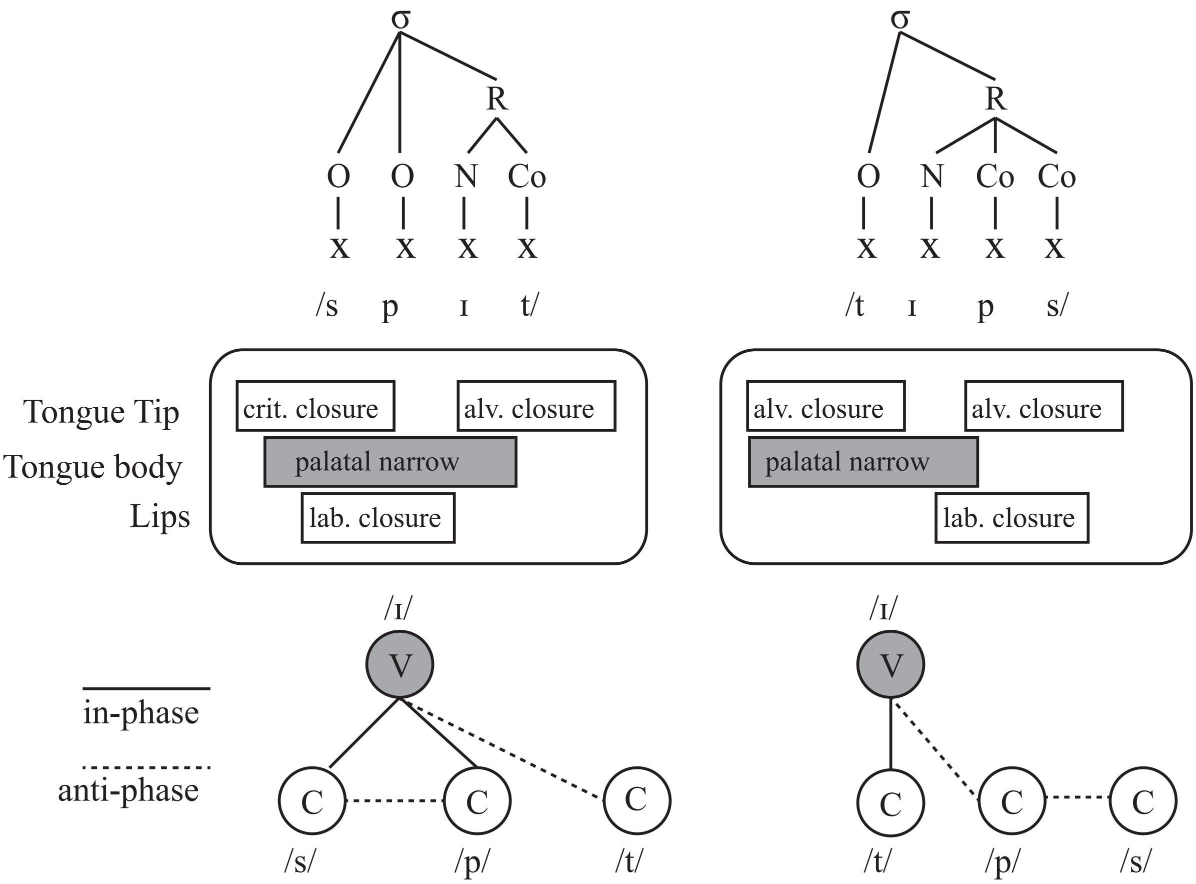


Figure 3.8: Syllabification (upper panel), gestural scores (middle panel) and corresponding coupling structures (lower panel) for the syllables /spɪt/ and /tɪps/ (oral gestures only).

In /spɪt/, both oral consonantal gestures belonging to the syllable onset, the /s/ (*critical closure*) and the /p/ (*labial closure*), are coupled in-phase with the vocalic gesture /ɪ/ (*palatal narrow*) leading to a competitive interaction between the gestures as both consonantal gestures cannot start with the vocalic gesture at the same time. Thus, an anti-phase coupling between the two consonantal gestures is provided.[2] This coupling structure leads to the c-center effect, i.e. the vowel adjacent consonantal gesture shifts to the right to make room for the added consonant. The consonantal gesture belonging to the syllable coda, /t/ (*alveolar closure*) is coupled anti-phase with the vocalic gesture.

[2]It must be noted that the exact phasing relation between consonants in a cluster is still a matter of debate, see Goldstein (2011) and Saltzman et al. (2006).

Prototypically, the consonantal gesture for /s/ starts first, then the vocalic gesture for /ɪ/ is initiated followed by the consonantal gesture for /p/. Eventually, the consonantal gesture for /t/ is initiated.

In contrast, the coupling structure for /tɪps/ provides no competitive coupling between the consonantal gestures as there is only one consonantal gesture in the syllable onset. This gesture for /t/ is coupled in-phase with the vocalic gesture for /ɪ/. The postvocalic consonantal gesture for /p/ is coupled anti-phase with the vocalic gesture and the last consonantal gesture for /s/ is only coupled anti-phase to the preceding consonantal gesture. As a result, the consonantal and vocalic gestures start simultaneously with the following gestures initiated sequentially.

Figure 3.9 schematizes the c-center effect resulting from a competitive coupling structure. The dashed lined represents the c-center as the temporal midpoint of all consonantal targets incorporated in the syllable onset. When consonants are added to word-initial CV structures and the language under investigation allows the consonants to enter the syllable internal gestural network (such as in <lay> to <play> and <splay>), then the vowel-adjacent consonant is expected to shift rightwards to make room for the added consonant(s). Vice versa, a consonant is expected to shift leftwards when other consonants are added betweem it and the vowel. In sum, the temporal distance between the c-center and the following vowel is said to remain stable across all syllable structures.

In the last twenty years, researchers have intensively investigated intergestural timing patterns in word-initial and word-final consonant clusters to shed light on their status in the syllable as complex onsets or codas (see also Tilsen et al. 2012 for an extensive overview). The rationale behind these studies is that

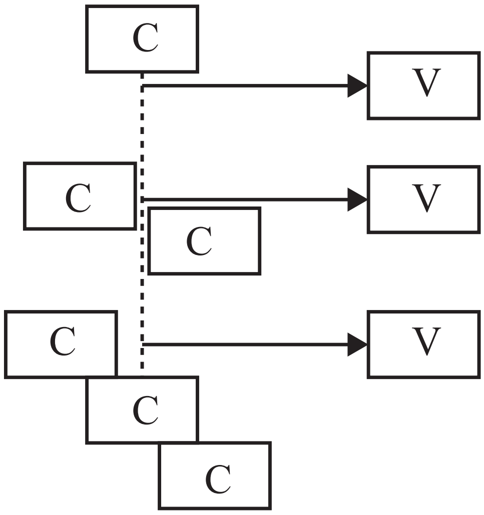


Figure 3.9: Schematic illustration of the c-center effect (after Saltzman et al. 2006: 69).

> "If the coupling model is correct in hypothesizing that complex onsets have a competitive coupling graph, then it should be possible to use the consequences of the competition (e.g., the rightward shift of the final C) as a diagnostic that a C sequence is syllabified as an onset" (Goldstein, Chitoran & Selkirk 2007: 243.)

For American English, Browman & Goldstein (1988), and more recently Marin & Pouplier (2010), have shown that word-initial clusters form complex onsets in that a C-center effect can be observed. In contrast, word-final consonants in American English do not show a c-center effect.[3] Further evidence for word-initial clusters being timed globally and thus forming complex syllable onsets comes from studies on Romanian (Marin 2013), German (Pouplier 2012), French (Kühnert, Hoole & Mooshammer 2006), Georgian (Goldstein, Chitoran & Selkirk 2007) and Polish (Mücke et al. 2010, but see Pastätter & Pouplier 2014). Languages that do not allow word-initial consonants to form a complex onset include Moroccan Arabic (Shaw et al. 2009) and Tashlhyit Berber (Hermes et al. 2011, Goldstein, Chitoran & Selkirk 2007). In these languages, no C-center effect was observed,

[3] However, it should be noted that some wordfinal clusters, especially those produced with the lateral approximant /l/ do show a C-Center like organisation, i.e. the consonant following the vowel is retracted towards the vowel.

i.e. the consonant preceding the vowel does not shift towards the vowel. Rather, there was a stable timing pattern between the vowel and the preceding consonant indicating that the added consonants are not integrated into the syllable internal gestural network. However, there is evidence that some languages allow for both simple and complex onsets: Hermes et al. (2008) investigated word-initial consonant clusters in Italian and found that clusters containing the sibilant /s/ such as in /spina/ do not show the c-center effect while clusters without this sibilant such as in /prima/ do. In Italian, there is a morphological alternation between the masculine articles <il> and <lo>. The article <il> occurs before nouns beginning with one consonant or a consonant cluster except for /s/+C(C) clusters. In the case of /s/+C(C) clusters the article <lo> is used. Hermes et al. (2008) shed light on the syllabification of such /s/+C(C) clusters showing that these clusters do not form complex onsets. Adding an /s/ to the cluster /pr/ does not affect the intergestural timing between the vowel and the preceding consonant. More specifically, the vowel-adjacent consonant does not shift towards the vowel to make room for the added /s/. Thus, the /s/ is analysed as not being part of the syllable onset. Rather, the /s/ is directly linked to the syllable or word node in the syllable hierarchy (see Hermes, Mücke & Grice 2013 for a detailed phonological analysis).

3.4 Prosody in Articulatory Phonology

This section deals with coupling structures that have been proposed to model articulatory timing involving higher levels of the prosodic hierarchy, i.e. the articulatory timing at prosodic boundaries and the coordination of articulatory gestures with lexical tones, pitch accents and boundary tones.

A huge number of studies have shown that articulatory timing is sensitive to prosodic structure. More specifically, gestural execution slows down at the edges of prosodic domain and thus overlap less in time (cf. Byrd 2000, Cho 2006, Cho & Keating 2001, Fougeron 2001, Tabain 2003, Krivokapić & Byrd 2012). Rather than modelling these boundary effects on the articulatory timing for each articulatory gesture separately, Byrd (2000), Byrd et al. (2000) and Byrd & Saltzman (2003) propose a prosodic gesture, the π-gesture, that models interarticulatory timing in the vicinity of prosodic boundaries (see also Saltzman et al. 2008 for a task dynamics application). The π gesture comes without any constriction, rather it modifies articulatory gestures in terms of their stiffness, such that boundary-adjacent gestures become less stiff and thus slow down:

> "We refer to such prosodic boundary gestures as π-gestures. Like articulatory gestures, π-gestures have an inherent temporal extent. The π-gestures are hypothesized to cause slowing at phrase edges by affecting the ongoing stiffness parameter values of all constriction gestures that are active within the π-gesture's activation interval. Stronger π-gestures slow the local speaking rate more (i.e., lower stiffness more) than weaker π-gestures. The prosodic boundary strength defines the activation level of a π-gesture" (Byrd 2000: 13).

Figure 3.10 schematizes the π-gesture's mode of operation. As articulatory gestures (*constrictions*) are produced at a prosodic boundary, their associated oscillators are affected by the π-gesture in that it slows down their execution. The degree of influence depends on the proximity to the boundary. Boundary-adjacent gestures are more affected by the π-gesture than non-adjacent gestures. In other words, as the activation of the π-gesture increases, it increasingly slows down the articulatory gestures. However, the exact domain or scope of the π gesture and whether pre- and postboundary gestures are affected equally is still a matter of debate (Byrd & Saltzman 2003, Krivokapić 2007, Krivokapić & Byrd 2012).

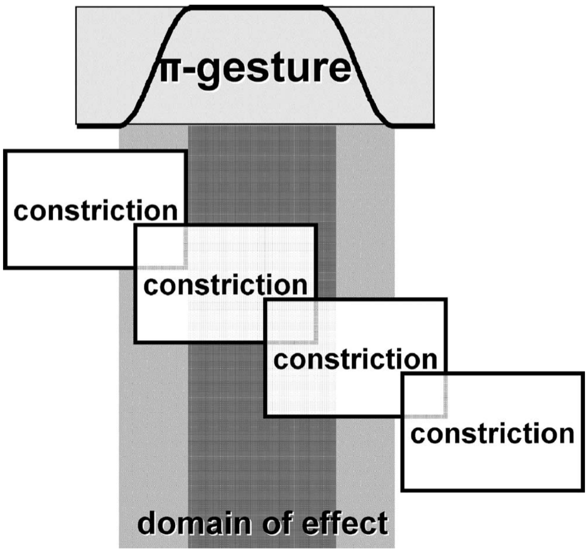


Figure 3.10: Illustration of the π-gesture affecting articulatory timing (reprinted with permission from Byrd, Krivokapić & Lee 2006: 1591).

Another prosodic of gestures includes the μ-gestures (*modulation gestures*), proposed to model the effect of stress on intergestural timing (cf. Saltzman et al. 2008). Rather than generally slowing down the articulators at prosodic boundaries, μ-gestures can modulate specific constrictions both temporally (μ_T-gesture) and spatially (μ_S-gesture).

Besides modelling the articulatory timing at prosodic boundaries, recent work has focussed on the coordination of articulatory gestures with so-called tone or tonal gestures. Gao (2009) first applied the concept of gestures as supralaryngeal movements to laryngeal movements. A tonal gesture is described as a "coordinated articulatory action to achieve a desired tonal goal and thus defined as dynamical system in F0 space" (Mücke et al. 2012: 209). This definition is reasonable as F0 is explicitly linked to glottal gestures in that it it is determined by rate of vocal fold vibration (see Laver 1994 for a discussion).

In contrast to an autosegmental-metrical analysis of tonal events, this approach treats tones, whether lexical tones or pitch accents, as action units. Figure 3.11 provides an analysis of a bitonal pitch accent analysed in the two different frameworks. In the AM model, this bitonal pitch accent is analysed as having two tonal targets, L and H, corresponding to the beginning and the end of the accentual rise in the F0 trace. The gestural analysis treats the accentual rise as *one tone gesture*, a high tone gesture (H gesture) with L and H denoting its beginning and end, respectively. Along this line, a falling pitch accent would be analysed as a low tone gesture (L gesture).

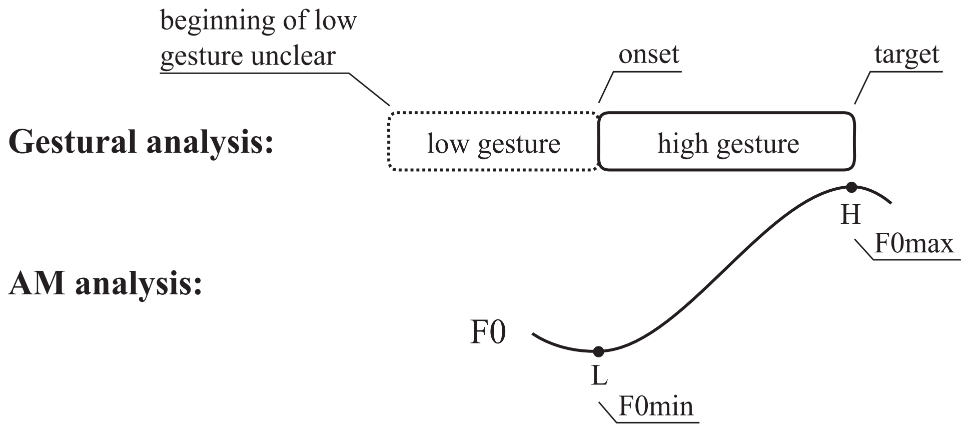


Figure 3.11: Autosegmental-metrical and gestural analysis of a bitonal pitch accent (adapted from Niemann et al. 2011 and Mücke et al. 2012)

Gao (2009) modelled the coordination of the four lexical tones in Mandarin Chinese with articulatory gestures using the concept of tone gestures. For the lexical tones she proposed four different tone gestures (see Figure 3.12). Both tone 1, a high-level tone, and tone 3, a low-falling tone, are modelled as single tone gestures, a high gesture for tone 1 and a low tone gesture for tone 3. Tone 2, a rising tone, and tone 4, a high falling tone, are modelled as a combination of low and high tone gestures. For tone 2, Gao assumes the two tone gestures to be coupled in-phase as indicated by a simultaneous start. For tone 4, she proposes an anti-phase coupling as indicated by the sequential execution of the tone gestures.

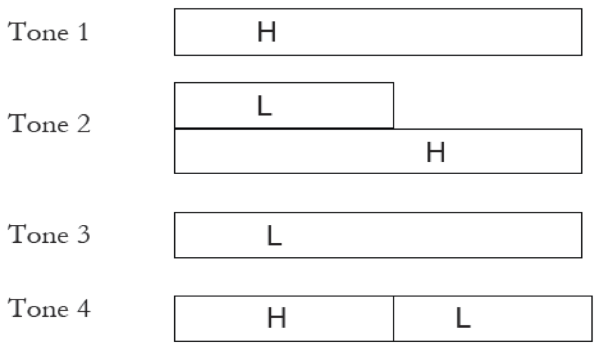


Figure 3.12: Proposed tone gestures for the four lexical tones in Mandarin Chinese (from Gao 2009: 43).

In her articulatory study, Gao examined the temporal lags between the onsets of the vocalic, consonantal and tone gestures. More specifically, she calculated the lags between the onset of the tone gesture relative to the onset of the consonantal and vocalic gestures, respectively. In addition, the lag between the onset of the vocalic gesture and that of the consonantal gesture was calculated. Her results show that tone gestures behave like consonants that are added to the syllable onset, that is, there is a c-center like coordination between the tones, the consonant and the vowel. First, the consonantal gesture starts, followed by the vocalic gesture and the tone gesture(s). Particularly, for tone 1, tone 2 and tone 3 the onset of the vocalic gesture roughly co-occurs with the midpoint between the onsets of the consonantal and tone gestures. For tone 4, the vocalic gesture starts significantly later and roughly starts with the first tone gesture. Gao (2008) proposes two different coupling structures (see Figure 3.13). For tone 1, tone 2 and tone 3 the tone gesture is coupled in-phase with the vowel gesture but is coupled in an anti-phase mode with the consonant gesture. This coupling structure results in a sequential activation of the articulatory gestures as can be seen in the score on the right-hand side in Figure 3.13. For tone 4, Gao proposes the additional tone gesture to be coupled in-phase with the vowel and to be coupled anti-phase with the other tone gesture. This competitive coupling results in the vocalic gesture starting with the tone gesture.

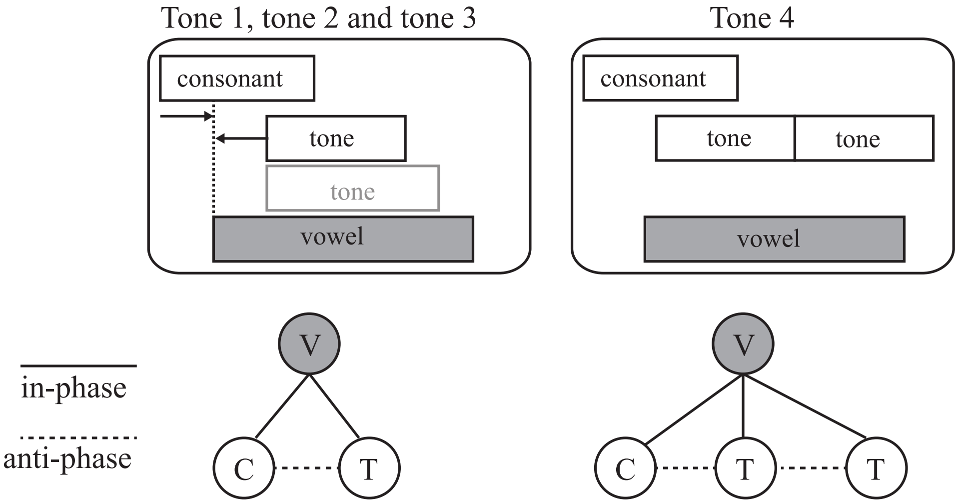


Figure 3.13: Gestural scores and corresponding coupling structures for lexical tones in Mandarin Chinese (after Gao 2009).

Hsieh (2011) extends the findings on lexical tones in Mandarin Chinese by modelling their coordination with task dynamics (TADA, cf. Nam et al. 2004). This study highlights the fact that the realisation of tone 3 depends on the contextual and contonal environment. Hsieh (2011) identifies three realisations of tone 3: a falling-rising contour ("full tone 3"), a low falling contour ("low tone 3") and tone 3 surfacing as tone 2, a rising contour ("sandhi tone 3"). Unlike Gao's (2009) approach, Hsieh (2011) provides two tone gestures, high and low, for the three realisations of tone 3. Figure 3.14 provides an overview of the underlying coupling structure for three realisations of tone three. The falling-rising contour ("full tone 3") can be modelled with the low tone gesture (L) being coupled in-phase with the vowel, and the high tone gesture (H) being coupled anti-phase with the vowel. This structure resembles the coupling structure given for coda consonants that are also coupled anti-phase with the vowel. When tone 3 surfaces as a falling tone ("low tone 3"), e.g. when it is followed by another high tone, its high gesture is obscured by the existing high gesture being coupled in-phase with the following vowel. Thus, only the falling part of the contour is realized. The coupling structure for the rising contour in “sandhi tone 3” is similar to the structure for the falling-rising contour in “full tone 3”. Hsieh (2011) hypothesized that this variation can be explained in terms of an additional in-phase coupling between the low and high tone gesture.

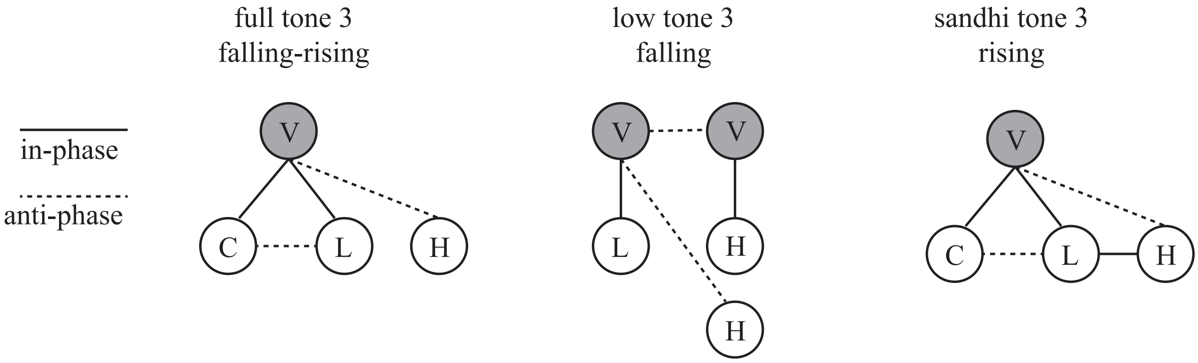


Figure 3.14: Coupling structure for tone 3 variations in Mandarin Chinese (after Hsieh 2011).

Apart from these studies that have looked investigated the coordination of lexical tones with articulatory gestures, there are studies that looked at the coordination of post-lexical tones such as pitch accents and boundary tones. Niemann et al. (2011) examined the coordination of rising nuclear pitch accents in German and Italian. Based on articulatory data from two speakers of each language, our corpus included trochaic CV.CV target words such as /ˈlina/ or /ˈnina/. In the acoustic domain, the beginning of the accentual rise was related to the acoustic onset of the stressed syllable. In the articulatory domain, the beginning of the accentual rise was related to the onset of the vocalic gesture (tone-vowel-lag), i.e. the onset of the tongue body movement for the accented vowel. In addition, the lag between the onset of the vocalic gesture and the onset of the consonantal gesture was calculated for the stressed syllable (consonant-vowel-lag). On the acoustic surface, the beginning of the accentual rise aligns shortly before the acoustic syllable onset in Italian. In German, however, the rise starts significantly later. It aligns in the onset consonant or even later in the stressed vowel. In the articulatory domain, we showed that in Italian the onset of the high tone gesture co-occured with both the onset of the consonantal and the onset of the vocalic gesture while in German, the high tone gesture was initiated significantly later. We account for these differences by proposing two different coupling structures for the two languages. Figure 3.15 displays gestural scores based on means from one speaker of each language and the proposed coupling structures.

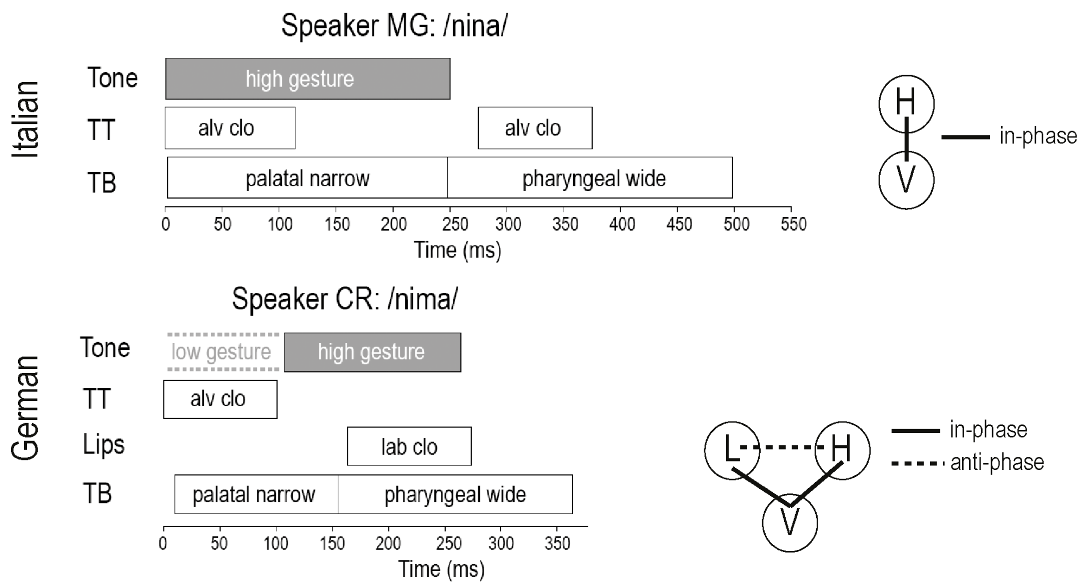


Figure 3.15: Gestural scores and coupling structures for Italian (upper panel) and German (lower panel) (from Niemann et al. 2011).

In the Italian data, upper panel, the tone gesture (“high gesture”) for the accentual rise, the consonantal gesture (“alv clo”) for the alveolar closure in /n/ and the vocalic gesture (“palatal narrow”) for the vowel /i:/ start simultaneously. In terms of coupling, the high gesture (H) is coupled in-phase with the vocalic gesture (H) leading to a synchronous start for the two. For German, we propose an additional tone gesture, a low tone gesture (L), that is coupled in-phase with the vocalic gesture. As in a consonant clusters, the tone gestures are coupled anti-phase with each other. This competitive coupling structure leads to the delayed start of the high tone gesture resulting in a later alignment of the accentual rise on the acoustic surface.

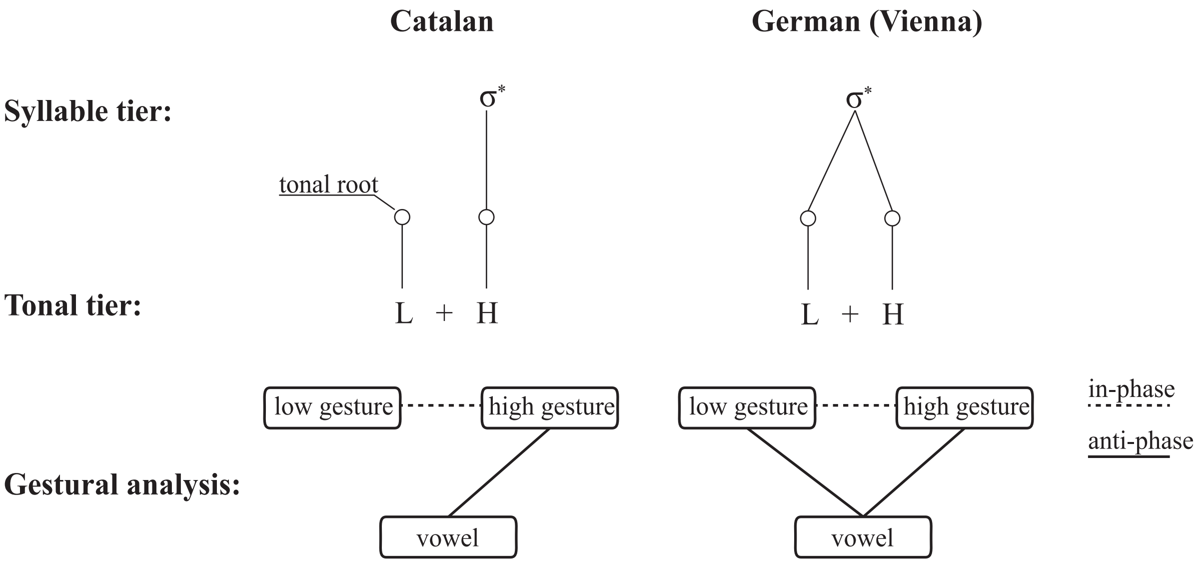


Figure 3.16: Gestural and autosegmental-metrical analysis for accentual rises in Catalan (left-hand side) and German (right-hand side) (adapted from Mücke et al. 2012: 225).

Similar coupling structures have been proposed by Mücke et al. (2012) by comparing the coordination of high tone gestures in Catalan and German (Vienna variety). Similar to the Italian data in Niemann et al. (2011), Mücke et al. (2012) found that in Catalan the accentual rise starts earlier than in German. In gestural terms, the high tone gesture is initiated with the vocalic gesture at the same time resulting in an earlier accentual rise compared to German where the high tone gesture is initiated later. The authors also outline the connection between an autosegmental-metrical analysis and the gestural analysis of early and late accentual rises on the acoustic surface. Figure 3.16 illustrates their analysis in terms of tonal association in the autosegmental-metrical model and in terms of gestural coupling in the framework of Articulatory Phonology.

Mücke et al. (2012) propose two tone gestures, low (L) and high (H), for the two languages. In Catalan, the low tone gesture is not coupled with the vocalic gesture resulting in a simultaneous activation of the high tone gesture and the vocalic gesture. In terms of tonal association, only H is associated with the accented syllable. In German, the low tone gesture is coupled in-phase with the vowel leading to a competitive structure that results in a later aligned accentual rise. In terms of tonal association, both tones L and H are associated with the accented syllable.

Katsika et al. (2014) modelled the coordination of boundary tones with consonantal and vocalic gestures in Greek by employing both the π-gesture and the μ-gesture. More specifically, they examined the onset of low and high boundary tone gestures in phrase-final trisyllabic target words with a CV.CV.CV structure. In addition, they varied the stress position resulting in target words such as /ˈma.mi.ma/, /ma.ˈmi.ma/ and /ma.mi.ˈma/. Katsika et al. (2014) found that the onset of the boundary tone gesture accompanying the final syllable co-occured roughly with the target of the vocalic gesture in that syllable. For example, the onset of a high boundary tone gesture co-occured with the maximum tongue body lowering for the production of the vowel /a/ in the final syllable /ma/. The authors thus propose an anti-phase coupling between the boundary tone and the vocalic gesture similar to the coupling structure for the Mandarin lexical tone 3 provided by Hsieh (2011). The position of stress in the trisyllabic target words also affected the onset of the boundary tone gesture. More specifically, stress attracted the activation of the boundary tone gesture, i.e. it was activated earlier in /ˈma.mi.ma/ than in /ma.mi.ˈma/. Figure 3.17 provides a coupling structure involving the vocalic gesture, the boundary tone gesture as well as the use of the π-gesture and the μ-gesture. The upper panel shows the coupling for target words with stress on the antepenultimate syllable such as in /ˈma.mi.ma/, the middle panel coupling for target words with penultimate stress (/ma.ˈmi.ma/) and the lower panel coupling for target words with final stress (/ma.mi.ˈma/). As a rule, the μ-gesture is always coupled in-phase with the vocalic gesture of the stressed syllable to mark stress, and the π-gesture is always coupled anti-phase with the vocalic gesture of the final syllable to mark the boundary.

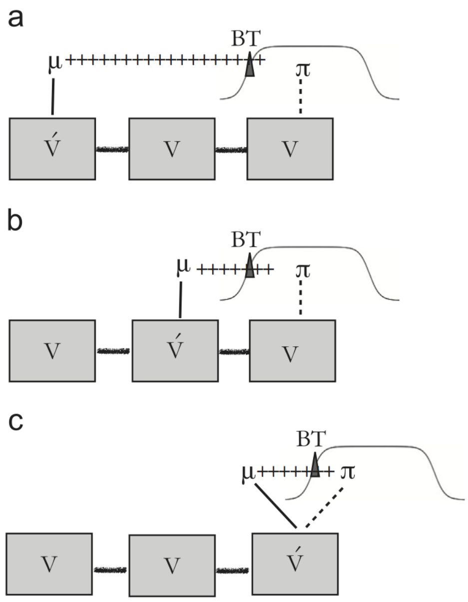


Figure 3.17: Coordination and coupling between vocalic and boundary tone gestures involving μ-gestures and π-gestures (reprinted with permission from Katsika et al. 2014: 80).

In addition, Katsika et al. (2014) assume a specific coupling of “[a] currently uncertain type” (ibid: 79) between the π-gesture and the μ-gesture as denoted by the “+” symbols. Importantly, Katsika et al. (2014) hypothesized that boundary tone gestures might not be coupled with the vocalic gesture, rather they assume that “[...] their timing is controlled indirectly via the coordination of the π-gesture” (ibid: 79). Thus, the onset of the boundary tone gesture is initiated earlier in target words with antepenultimate and penultimate stress. In target words with final stress, it is initiated later as both the π-gesture and the μ-gesture are coupled with the vocalic gesture in the final syllable.

The present thesis investigates the coordination of both rising nuclear and prenuclear pitch accents with oral constriction gestures by means of articulatory data collected from four speakers of German. The next chapter will explain the methods and recording procedure in detail.

4 Methods

4.1 Speakers and recordings

This thesis reports on articulatory and acoustic data that were recorded at the Institut für Linguistik - Abteilung Phonetik at the University of Cologne. Four German female subjects (S1-S4) aged between 26 and 32 years participated in the study. All speakers shared a similar educational background and grow up north of the Benrather isogloss.[1] Speakers S1 and S4 come from Lower Franconia (Wesel), S2 comes from Eastphalia (Brilon) and S3 from Lower Saxony (Löningen). None of the speakers reported on any hearing disorders or speaking disfluencies. Speakers were renumerated for their efforts.

The kinematic data were obtained with a 3D Electromagnetic Articulograph[2]. The articulograph creates a low-energy magnetic field capable of capturing the movements of sensor coils attached to the speakers' articulators. Using fibrin glue, sensor coils were placed on the upper and lower lips, the tongue tip (1 cm behind the tip), the tongue blade (2 cm behind the tip) and the tongue body (3 cm and 4 cm behind the tip). Two additional sensors on the upper gums and on the bridge of the nose served as reference

[1] The Benrather isogloss is a major linguistic border separating Middle German from Low German.
[2] AG 500 from (Carstens Medizinelektronik GmbH).

coils to correct for head movements. Figure 4.1 shows the positioning of coils for one subject before the recording session.

Figure 4.1: Position of the sensor coils attached to the subject's articulators.

Kinematic data were recorded with a sampling rate of 200 Hz. Acoustic data were recorded with the built-in, time-synchronized microphone and digitized at 16 kHz and 16 bit. At the end of each recording session, the sensors from the upper and lower lips served as reference points for the calculation of the occlusal plane[3]. Using the standard software provided with the articulograph, all kinematic data were processed and corrected for head movement. Data were then converted to the Simple Signal File Format by means of the software EMA2SSFF.[4] Both acoustic and articulatory data were labelled using the EMU Speech Database System (Harrington 2010).

[3]These sensors were attached to a bite plate and inserted into the subject's mouth between the tongue and palate. Subjects were instructed to keep their tongues still while the position of the sensor coils was recorded.

[4]Version 3.08.2, developed and programmed by Frank Christian Stoffel.

4.2 Speech material and data elicitation

The speech material was designed to test for the effects of accent (nuclear vs. prenuclear), syllables structure (open vs. closed syllable) and word length (monosyllables vs. disyllables). In addition, the corpus was designed to test for the effects of phrasal position (phrase-initial, phrase-noninitial and phrase-final position). The following section illustrates the speech material including the carrier sentences in more detail.

The corpus consisted of four target words differing in word length and syllable structure. Target words were either troachic disyllabes and had either an open or closed stressed syllable. (/ma:/, /mam/, /ˈma:.mi/ and /ˈmam.zi/). Table 4.1 gives an overview of all target words with, together with the orthographic representation that was presented to the subjects.

Table 4.1: Speech material.

Orthographic	Phonemic	CV
Ma	/ma:/	CV:
Mahmi	/ˈma:mi/	CV:CV
Mamm	/mam/	CVC
Mamsi	/ˈmam.zi/	CVC.CV

All target words were embedded in carrier sentences designed to elicit either a nuclear or a prenuclear rising pitch accent on the stressed syllables /ma/ and /mam/, respectively. In both accent conditions, the target word was produced both in phrase-initial and in phrase-noninitial position. In addition, target words bearing a nuclear accent were also produced in phrase-final position. All carrier sentences were uttered as an answer in small question and answer pairs. The basic structure of the carrier sentences for target words in phrase-noninitial position is given in Table 4.2.[5]

In Table 4.2 target words are shaded and the accented syllable of each target word is in bold. Ambisyllabic consonants are shown in parentheses. Note that all target words were embedded in carrier sentences having the same sequence of strong and weak syllables, thus ensuring no unintended effects of stress clash on the pitch accent realisation on the accented syllable.

Table 4.2: Basic corpus for target words in phrase-noninitial position. Target words are shaded. σ* denotes the accented syllable. Stressed and unstressed syllables are marked by σ and σ', respectively.

target word	σ	σ*	σ	σ	σ	σ	σ'	σ	σ	σ'	σ
/ma:/	Die	**Mah**	mi(n)	(n)e	ra	li	sier	te	das	Wa(s)	(s)er
/ma:.mi/	Die	**Mah**	mi	ma	ni	pu	lier	te	die	Brem	se
/mam/	Die	**Mamm**	si	mul	ta	ni	sier	te	die	Hand	lung
/mam.zi/	Die	**Mamm**	si	ma	ni	pu	lier	te	das	Fahr	zeug

Prior to the recordings, subjects were familiarised with the target words, i.e. the target words were presented as female first names by the experimenter. The elicitation of either a nuclear or a prenuclear rising pitch accent on the syllables /ma:/ or /mam/, respectively, was ensured by pairing the carrier sentence with a question in a small question-answer pair. Questions were recorded by a professional speaker and were designed to elicit either a nuclear or a prenuclear rising accent on the test syllable. Each question-answer-pair was presented to the subjects on a computer screen.[6]. Simultaneously, subjects heard the question via the computer's built-in loudspeaker and were asked to read the answer in a comfortable and natural way. Figure 4.2 shows an example slide from the question-answer pairs presented to the subjects during the recording sessions.

[5]Note that the verb <simultanisierte> in the carrier sentence with the target word /mam/ is a pseudoword. However, it follows German phonotactics and its content can easily be deduced as

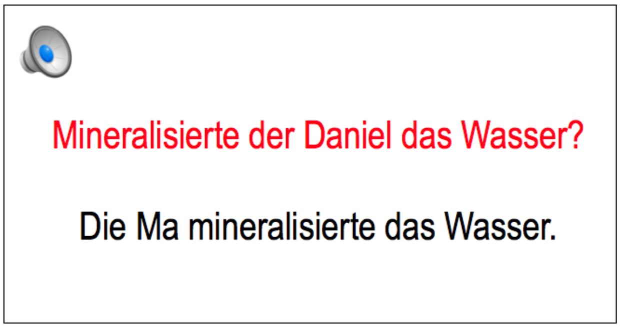


Figure 4.2: Example of a question and answer pair presented to the subjects.

The order of the question-answer-pairs was pseudo-randomized. Carrier sentences with the same target word or with target words in the same phrasal position did not appear adjacently. Each question-answer pair was repeated 8 times. In sum, 640 tokens were recorded (nuclear: 4 target words x 3 phrasal positions x 8 repetitions x 4 speakers = 384 tokens; prenuclear: 4 target words x 2 phrasal positions x 8 repetitions x 4 speakers = 256 tokens). Tokens that were not produced with a rising pitch accent or were misproduced due to mispronounciation were excluded from the analysis. In total, 544 tokens went into the analysis. In the following, I will discuss the question-answer pairs for target words in phrase-initial and phrase-noninitial position (both nuclear and prenuclear) and then for target words in phrase-final position (only nuclear).

Carrier sentences in phrase-initial and phrase-noninitial position were identical between the nuclear and prenuclear data. In phrase-initial position target words were placed at the left edge of an intonation phrase boundary.[7] In phrase-noninitial position, an unstressed syllable, the definite article <Die>, occurred before the target word. For all phrasal positions, questions were designed to elicit a rising pitch accent on the syllable

making something simultaneous.

[6]Microsoft PowerPoint for Mac 14.2.1

[7]Target words could not be placed in utterance-initial position as the articulatory onset of consonantal and vocalic gestures could not have been properly identified.

/ma:/ or /mam/. For nuclear accents, questions were designed so that the target words were part of the contrastive focus; in prenuclear accents, target words were part of the broad focus of the answer (cf. Büring 2007).

In the following, I will present the carrier sentences for target words in-phrase initial position first (both nuclear and prenuclear data), then I will present the carrier sentences for target words in phrase-noninitial positoin (both nuclear and prenuclear). Subsequently, I will present them for target words in phrase-final position (only nuclear).

Table 4.3: Question and answer pairs for target words in phrase-initial position (nuclear data).

target	Q/A	question and answer pair
/ma:/	question	Dachte sie dann: Daniel mineralisierte das Wasser? *Did she think: Daniel mineralised the water?*
	answer	[Dann dachte sie:]$_{IP}$ [Ma mineralisierte das Wasser.]$_{IP}$ *Then she thought:* ***Ma*** *mineralised the water.*
/ma:.mi/	question	Dachte sie dann: Lukas manipulierte die Bremse? *Did she think: Lukas manipulated the brakes?*
	answer	[Dann dachte sie:]$_{IP}$ [**Mah**mi manipulierte die Bremse.]$_{IP}$ *Then she thought: Mahmi manipulierte die Bremse.*
/mam/	question	Dachte sie dann: Sarah simultanisierte die Handlung? *Did she think: Sarah simultanised the plot?*
	answer	[Dann dachte sie:]$_{IP}$ [**Mamm** simultanisierte die Handlung.]$_{IP}$ *Then she thought: Mamm simultanised the plot.*
/mam.zi/	question	Dachte sie dann: Lena manipulierte das Fahrzeug? *Did she think: Lena manipulated the car?*
	answer	[Dann dachte sie:]$_{IP}$ [**Mam**si manipulierte das Fahrzeug.]$_{IP}$ *Then she thought: Mamsi manipulated the car.*

Table 4.3 shows the question-answer pairs with target words in phrase-initial position (nuclear data). The stressed syllable is in bold. Square brackets indicate an intonational phrase boundary. The nuclear rising pitch accent on the target word was elicited by asking a question such as <Dachte sie dann: X mineralisierte das Wasser> (<*Did she*

think: X mineralised the water?>) where X denotes another first name; thus the target word in the answer, the carrier sentence, occurred as part of the contrastive focus.

Table 4.4: Question and answer pairs for target words in phrase-initial position (prenuclear data).

target	Q/A	question and answer pair
/ma:/	question	Was dachte sie dann? *What did she think then?*
	answer	[Dann dachte sie:]$_{IP}$ [Ma mineralisierte das Wasser.]$_{IP}$ *Then she thought:* ***Ma*** *mineralised the water.*
/ma:.mi/	question	Was dachte sie dann? *What did she think then?*
	answer	[Dann dachte sie:]$_{IP}$ [**Mah**mi manipulierte die Bremse.]$_{IP}$ *Then she thought: Mahmi manipulierte die Bremse.*
/mam/	question	Was dachte sie dann? *What did she think then?*
	answer	[Dann dachte sie:]$_{IP}$ [**Mamm** simultanisierte die Handlung.]$_{IP}$ *Then she thought: Mamm simultanised the plot.*
/mam.zi/	question	Was dachte sie dann? *What did she think then?*
	answer	[Dann dachte sie:]$_{IP}$ [**Mam**si manipulierte das Fahrzeug.]$_{IP}$ *Then she thought: Mamsi manipulated the car.*

Table 4.4 displays the question-answer set for target words bearing a prenuclear accent in phrase-initial position. Note that the carrier sentences are identical to those in the nuclear data. However, a prenuclear rising pitch accent was elicited here by asking the question <Was dachte sie dann?> (<*What did she think then?*>) in all cases; thus the target word in the answer occurred as part of the broad focus. While speakers produced a prenuclear rising pitch accent on the stressed syllable /ma:/ or /mam/, in almost all instances, they produced a nuclear falling accent on the stressed syllable in the phrase-final word, e.g. in <Wasser> in the carrier sentence <Dann dachte sie: Ma mineralsierte das Wasser>.

Table 4.5 shows the question-answer pairs with target words in phrase-noninitial position

Table 4.5: Question and answer pairs for target words in phrase-noninitial position (nuclear data).

target	Q/A	question and answer pair
/ma:/	question	Mineralisierte der Daniel das Wasser? *Did Daniel mineralise the water?*
	answer	[Die **Ma** mineralisierte das Wasser]$_{IP}$ *The Ma mineralised the water*
/ma:.mi/	question	Manipulierte der Lukas die Bremse? *Did Lukas manipulate the brakes?*
	answer	[Die **Mah**mi manipulierte die Bremse]$_{IP}$ *The Mahmi manipulated the brakes.*
/mam/	question	Simultanisierte die Sarah die Handlung? *Did Sarah simultanise the plot?*
	answer	[Die **Mamm** simultanisierte die Handlung]$_{IP}$ *The Mamm simultanised the plot.*
/mam.zi/	question	Manipulierte die Lena das Fahrzeug? *Did Lena manipulate the car?*
	answer	[Die **Mam**si manipulierte das Fahrzeug.]$_{IP}$ *The Mamsi manipulated the car.*

(nuclear data). The carrier sentences are similiar to those in Table 4.3, however, they were produced as one intonational phrase, instead of two, and the unstressed syllable <Die> preceded the target word. Again, a nuclear rising pitch accent was elicited by asking a question in this case such as <Mineralisierte X das Wasser?> (<*Did X mineralise the water?*>).

Table 4.6 displays the question-answer set for target words bearing the prenuclear accent in phrase-noninitial position. The carrier sentences are identical to the nuclear ones in the same position. The prenuclear rising pitch accent was elicited by asking the question <Was hat die X gestern gemacht> (<*What did X do yesterday?*>) where X denotes the target word. Thus, the target word occurred as given information in broad focus. Similar to the carrier sentences with the target word in phrase-initial position, speakers produced a prenuclear rising pitch accent on the stressed syllable /ma:/ or /mam/, and a nuclear falling accent on the stressed syllable in the phrase-final word, e.g. in <Wasser> in the carrier sentence <Die Ma mineralisierte das Wasser>.

Table 4.6: Question and answer pairs for target words in phrase-noninitial position (prenuclear data).

target	Q/A	question and answer pair
/ma:/	question	Was hat die Ma gestern gemacht? *What did Ma do yesterday?*
	answer	[Die **Ma** mineralisierte das Wasser]$_{IP}$ *The Ma mineralised the water*
/ma:.mi/	question	Was hat die Mahmi gestern gemacht? *What did Mahmi do yesterday?*
	answer	[Die **Mah**mi manipulierte die Bremse]$_{IP}$ *The Mahmi manipulated the brakes.*
/mam/	question	Was hat die Mamm gestern gemacht? *What did Mamm do yesterday?*
	answer	[Die **Mamm** simultanisierte die Handlung]$_{IP}$ *The Mamm simultanised the plot.*
/mam.zi/	question	Was hat die Mamsi gestern gemacht? *What did Mamsi do yesterday?*
	answer	[Die **Mam**si manipulierte das Fahrzeug.]$_{IP}$ *The Mamsi manipulated the car.*

In phrase-final position (only nuclear data), target words were placed in utterance-final position. Table 4.7 shows the question-and-answer pairs with target words in phrase-final position (nuclear data). While the carrier sentences are different from those with target words in phrase-initial and phrase-noninitial position, they share the same sequence of strong and weak syllables. In the case of monosyllabic target words, the accented syllables coincides with the final syllable in the intonation phrase. A nuclear rising pitch accent was elicited by asking a question such as <Hat sie dann den X gesehen?> (<*Did she see the Daniel?*>) where X denotes another first name. Again, target words occurred as part of the contrastive focus in the answer.

Table 4.7: Question and answer pairs for target words in phrase-final position (nuclear data).

target	Q/A	question and answer pair
/ma:/	question	Hat sie dann den Daniel gesehen? *Did she see the Daniel?*
	answer	[Sie sah dann die Ma.]$_{IP}$ *Then she saw the Ma.*
/ma:.mi/	question	Hat sie dann den Lukas gesehen? *Did she see the Lukas?*
	answer	[Sie sah dann die Mahmi.]$_{IP}$ *Then she saw the Mahmi.*
/mam/	question	Hat sie dann die Sarah gesehen? *Did she see the Sarah?*
	answer	[Sie sah dann die Mamm.]$_{IP}$ *Then she saw the Mamm.*
/mam.zi/	question	Hat sie dann die Lena gesehen? *Did she see the Lena?*
	answer	[Sie sah dann die Mamsi.]$_{IP}$ *Then she saw the Mamsi.*

The next section presents detailed information on both the data annotation and the calculation of alignment lags for the beginning and end of the accentual rise relative to landmarks in both the acoustic and articulatory domain.

4.3 Data labelling and processing

The *EMU Speech Database System* (Cassidy & Harrington 2001, Bombien et al. 2006, Harrington 2010) was used to create a database for each subject and condition. To look at the effects of phrasal position, word length and syllable structure on the alignment of nuclear and prenuclear accents, target words were labelled in the acoustic (segmental boundaries, F0 targets) and articulatory (onsets and targets of consontal and vocalic gestures) domain. With the exception of tonal targets (local minima and maxima) in the F0 trace, both acoustic and articulatory data were manually labelled in EMU. The beginning and end of each accentual rise was manually labelled in PRAAT (Boersma & Weenink 2010).[8] All data were extracted and transferred to *R* (R Core Team 2013)[9] via the EMU/R interface[10]. The following section will present the data labelling in more detail.

4.3.1 Acoustic labelling

In the acoustic domain, segmental boundaries were annotated by means of an oscillogram and wide-band spectrogram with a bandwidth of 300 Hz. Figure 4.3 depicts the annotation of segmental boundaries of the target word /ma:mi/ in phrase-initial position (nuclear pitch accent). The target sentence was <Dann dachte sie, Mahmi manipulierte die Bremse> ("Then she thought, Mahmi manipulated the brakes").

[8] Version 6.0.17
[9] Version 2.15.1 for extraction and version 3.1.2. for analysis
[10] Jonathan Harrington, Tina John, others and IPS LMU Muenchen & IPDS CAU Kiel (2012). emu: Interface to the Emu Speech Database System. R package version 4.3. http://CRAN.R-project.org/package=emu

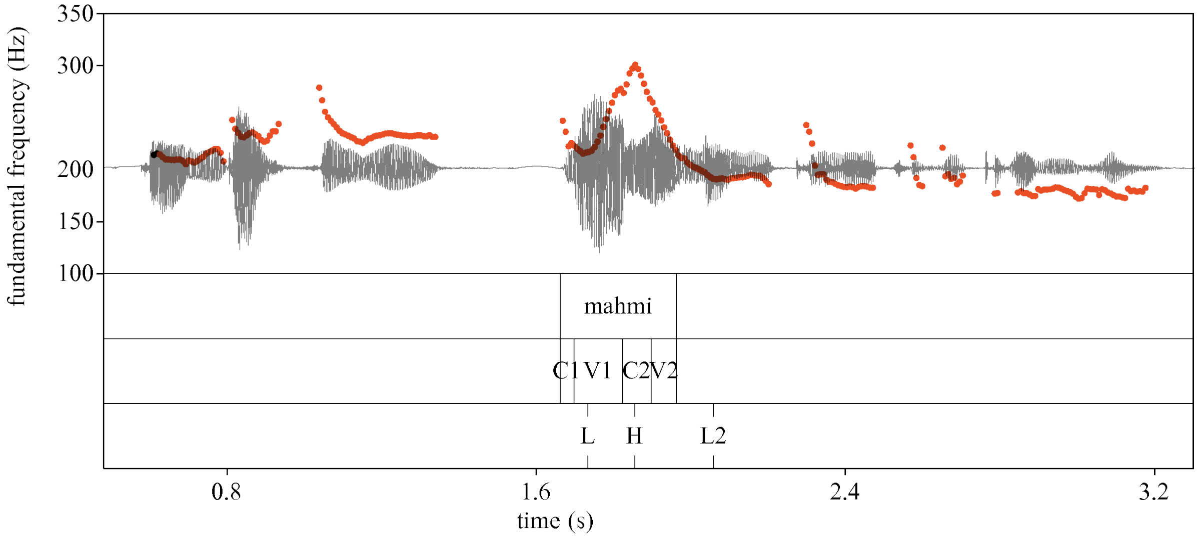


Figure 4.3: Annotation of acoustic boundaries and the beginning and end of the nuclear rise (L and H, respectively) for /ma:mi/ in phrase-initial position (data from S1).

In the F0 trace, local minima and maxima of the rising nuclear (or prenuclear) accent were identified and labeled as the beginning and the end of the accent. As can be seen in Figure 4.3 above, the accentual rise starts shortly after the start of the accented vowel V1 and, in this example, reaches its peak in the postaccented consonant C2.

Figure 4.4 provides an example of the target word /ma:mi/ produced in phrase-final position (nuclear pitch accent). The target sentence was <Sie sah dann die Mahmi> ("Then she saw the Mahmi"). Compared to the F0 peak in phrase-initial position, the F0 peak aligns earlier, namely at the end of the accented vowel V1.

An example of a prenuclear rise is given in Figure 4.5. The target sentence was <Die Mahmi manipulierte die Bremse> ("The Mahmi manipulated the brakes.") with a prenuclear rise accompanying the target word /ma:mi/. Compared to the nuclear F0 peak position, the F0 peak aligns considerably later in the prenuclear rise.

The next section provides detailed information on the annotation carried out in the

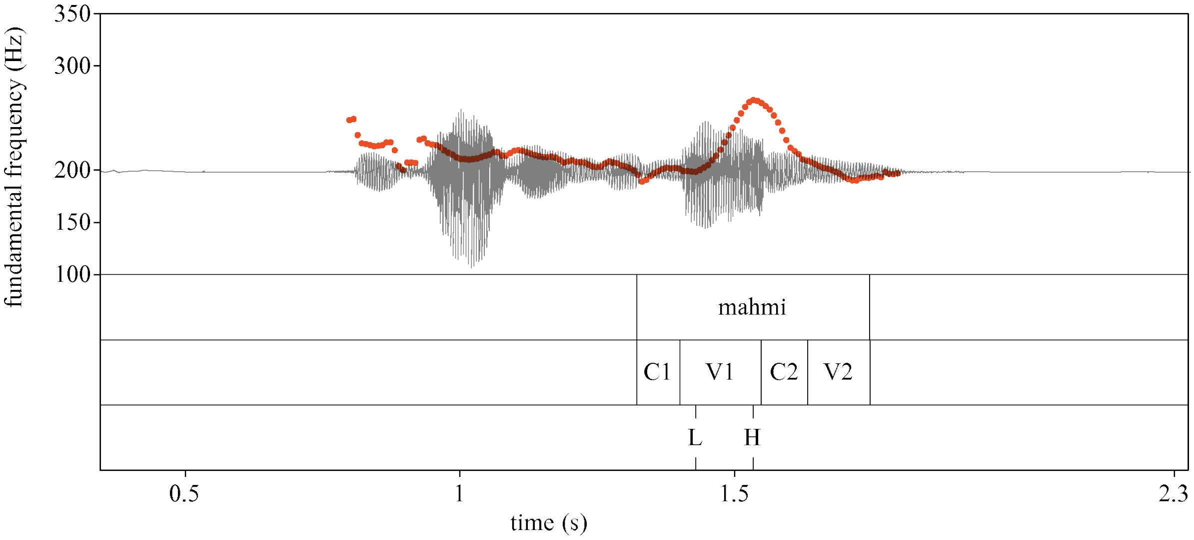


Figure 4.4: Annotation of acoustic boundaries and the beginning and end of the nuclear rise (L and H, respectively) for /ma:mi/ in phrase-final position (data from S1).

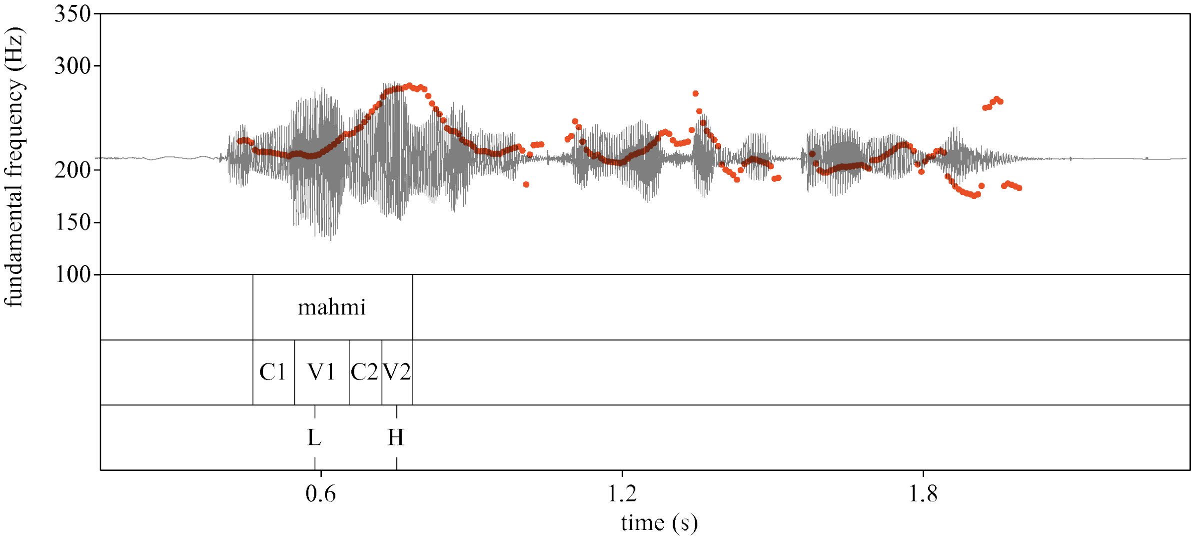


Figure 4.5: Annotation of acoustic boundaries and the beginning and end of the prenuclear rise (L and H, respectively) for /ma:mi/ in phrase-noninitial position (data from S1).

articulatory domain.

4.3.2 Articulatory labelling

In the articulatory domain, both consonantal gestures and vocalic gestures were labelled for each accented syllable. Consonantal gestures were identified via the Lip Aperture Index (Byrd 2000), which displays the interlip distance. In this trajectory, the target (the maximum lip closure) of the wordinitial consonant C1 [m], the occurence of the peak velocity of the opening gesture, the maximum lip opening, the occurence of the peak velocity of the following closing gesture and the target of the following consonant C2 [m] were labelled.

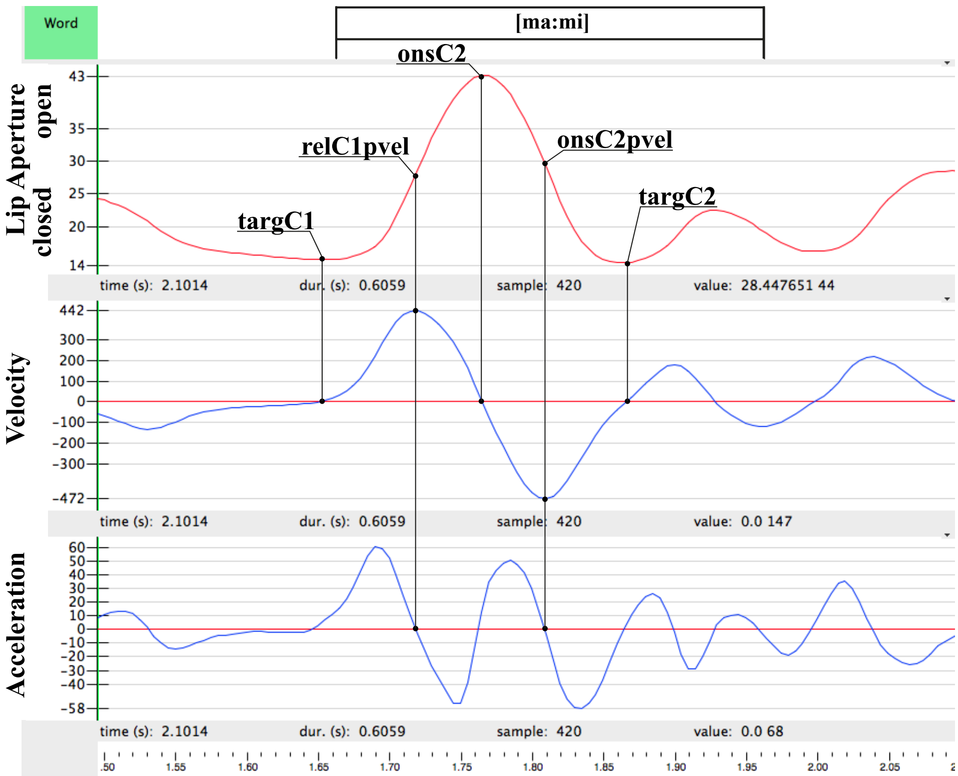


Figure 4.6: Annotated landmarks for bilabial gestures in the target word /ma:mi/. Upper panel: lip aperture, middle panel: velocity, lower panel: acceleration.

Figure 4.6 displays the target word /ma:mi/ with the interlip distance, the derived velocity and acceleration. The maximum lip closure for the production of C1 ("targC1"), the maximum lip opening ("onsC2") and maximum closure for the production of C2 ("targC2") were labelled by means of zero-crossings in the velocity trace. The time point where the opening gesture reaches its peak velocity ("relC1pvel") and the time point where the closing gesture reaches its peak velocity ("onsC2pvel") were labelled by means of zero-crossing in the acceleration trace.

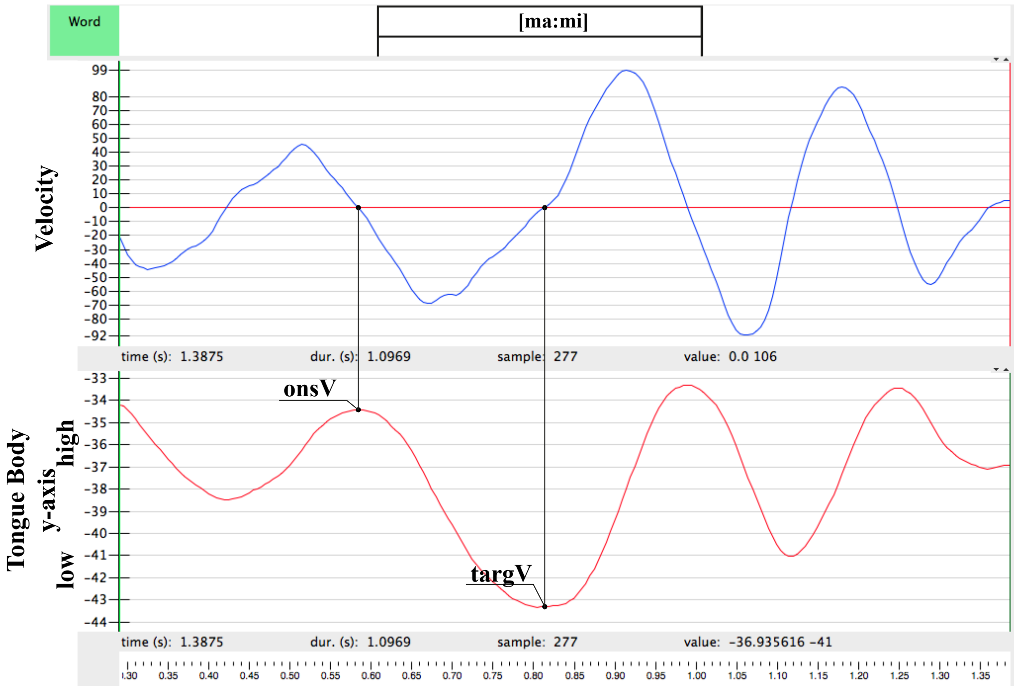


Figure 4.7: Annotated landmarks for vocalic gestures in /ma:mi/. Upper panel: Velocity, lower panel: vertical movement of the most back sensor coil on tongue.

Vocalic gestures were labelled in the vertical trajectory of the tongue body (most back sensor coil on the tongue). Vowel-to-vowel articulation differs from consonantal articulation in that the articulatory target of a vocalic gesture corresponds to the onset of the following vocalic gesture. That is, the target of a vocalic gesture is at the same time the onset of a following vocalic gesture. Vocalic gestures were identified in terms of local maxima and minima, the former corresponding to the onset (“onsV”) of the articulatory gesture for /a/ (= the articulatory target for the preceding /i/) and the latter corresponding to the articulatory target for /a/ (“targV”). Figure 4.7 displays the annotation of the vocalic gesture in the target word /ma:mi/. The onset of the articulatory gesture for /a/ was labeled by means of a negative zero-crossing in the velocity trace, while its target was labeled by means of a positive zero-crossing.

4.3.3 Derived variables and calculations

This section presents the derived variables under investigation. Both nuclear and prenuclear rises were related to landmarks in the acoustic signal and articulatory traces summarised in Table 4.8 and Table 4.9. The landmark closest to the beginning of the accentual rise (L) was the acoustic vowel onset. Thus, L was measured relative the acoustic vowel onset both in absolute terms (**LtoV1ons**) and as a proportion of the vowel duration (**LtoV1.prop**). In addition, L was measured as a proportion of the total syllable duration (**LtoSyll.prop**). This measurement accounts for different expected syllable durations for open and closed syllables. The landmark nearest to the end of the accentual rise (H) was the end of the syllable. Thus, H was measured relative to the end of the accented vowel [a:] in [ma:] or [ˈma:.mi] or the end of the coda consonant [m] in [mam] or [ˈmam.zi] (**HtoEndSyll**). In addition, H was calculated as a proportion of the total syllable duration (**HtoSyll.prop**).

In the articulation, L was measured relative to nearby landmarks in the articulatory domain, namely the articulatory target of the consonantal closing gesture for C1 (**LtotargC1**) and the peak velocity of its release (**LtorelC1pvel**). H was measured relative to a non-nearby or distant landmark. It was related to the articulatory target of the vocalic gesture of V1, i.e. the maximum tongue body lowering for the production of /a:/ or /a/.(**HtotargV**).

For each analysis, data were submitted to a repeated measures ANOVA with SPEAKER as a random factor and POSITION (phrase-initial, phrase-noninitial, phrase-final), SYLL (open, closed) and WORD LENGTH (monosyllabic, disyllabic) as independent variables. An effect was deemed to be significant at alpha = 0.05. Effect sizes will be reported by means

Table 4.8: Measurements applied in the acoustic domain.

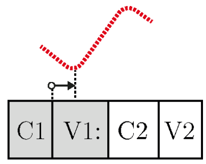
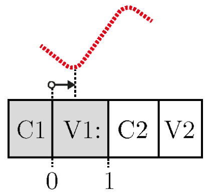
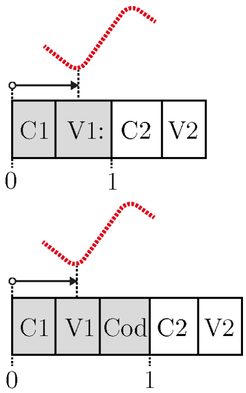
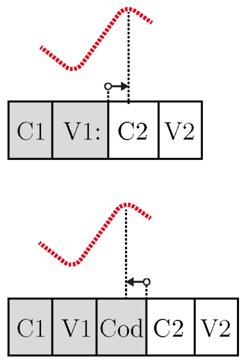
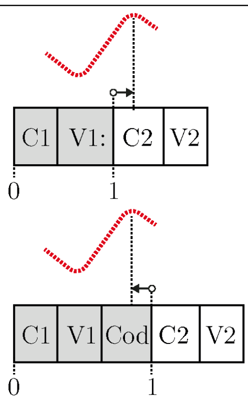
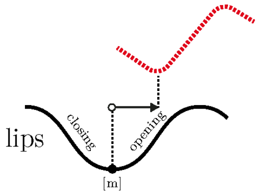
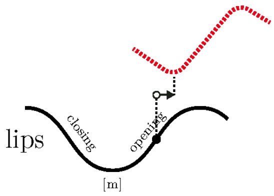
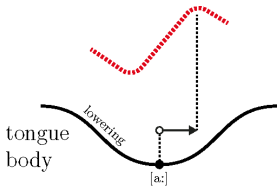
Measure	Description	Schema
LtoV1ons	Onset of the accentual rise relative to onset of accented vowel	
LtoV1.prop	Onset of the accentual rise relative to onset of accented vowel in relation to the vowel duration	
LtoSyll.prop	Onset of the accentual rise relative to the syllable onset in relation to the syllable duration	
HtoEndSyll	End of the accentual rise relative to the end of the accented syllable. In open syllables, this landmark corresponds to the end of the vowel. In closed syllables, it corresponds to the end of the coda consonant.	
HtoSyll.prop	End of the accentual rise relative to the syllable onset in relation to the syllable duration	

Table 4.9: Measurements applied in the articulatory domain.

Measure	Description	Schema
LtotargC1	Onset of the accentual rise relative to the maximum closure in C1	
LtorelC1pvel	Onset of the accentual rise relative to peak velocity of the C1 release gesture	
HtotargV	End of the accentual rise relative to the articulatory target of the accented vowel /a/ or /a:/	

of the generalized eta square coefficient η^2. This coefficient describes the proportion of a variance for a model that is explained by a given factor and ranges between 0 and 1. Values between 0.01 and 0.05 indicate a small effect size, values between 0.06 and 0.13 indicate a medium effect size and values above 0.14 a large effect size (Ellis 2010: 40). Where significant interactions between the factors were found, posthoc tests[11] were conducted in order to find out which data sets differed significantly from one another. Linear mixed effects modeling (see Chapter 7) was carried out using the R package *lme4* (Bates et al. 2015).

[11]Tukey's HSD (Honestly Significant Difference) tests

5 Results: Nuclear Accents

This chapter presents results for the effects of phrasal position, word length and syllable structure on the alignment of nuclear pitch accents. Section 5.1 investigates the alignment in the acoustic domain while section 5.2 investigates the alignment in the articulatory domain. Section 5.3 summarises the findings.

5.1 Alignment relative to acoustic landmarks

This section presents results for the nuclear pitch accent alignment relative to acoustic landmarks. First, we look at the alignment of the beginning of the nuclear rise which was measured relative to the acoustic onset of the accented vowel. In addition, it was measured as a proportion of both the vowel duration and the total syllable duration. Then we look at the alignment of the end of the nuclear rise which was measured relative to the end of the accented syllable. In addition, the end of the rise was measured as a proportion of the total syllable duration.

5.1.1 Acoustic alignment of L

In this section, we look at the alignment of the start of the nuclear rise, first in absolute terms (relative to the acoustic vowel onset), then in relative terms (relative to the vowel duration and the total syllable duration, respectively).

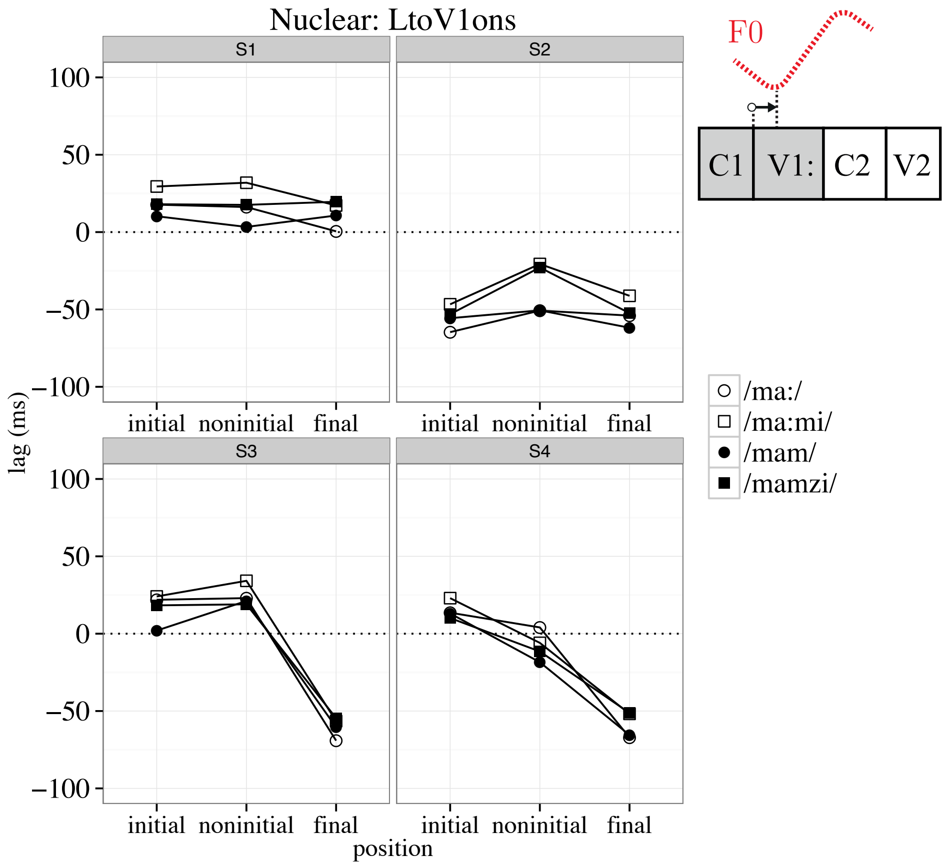


Figure 5.1: Mean alignment lags (in ms) for L relative to the onset of the accented vowel (LtoV1ons).

Figure 5.1 shows means by speaker for the beginning of the accentual rise relative to the acoustic onset of the accented vowel V1 in phrase-initial, phrase-noninitial and phrase-final position. Zero denotes the acoustic onset of the accented vowel V1. Positive values indicate that L occurs after the vowel onset, while negative values indicate that it occurs before it, that is in the preceding consonant. Full data including averages across speakers, phrasal position and syllable structure are given in Table 5.1.

Table 5.1: Mean alignment lags (in ms) for L relative to the onset of the accented vowel (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	18 (6)	-65 (9)	22 (15)	14 (3)	-4 (38)	1 (37)	-2 (35)
		/ma:mi/	29 (9)	-47 (8)	24 (9)	23 (16)	6 (34)		
	closed	/mam/	10 (7)	-56 (22)	2 (38)	13 (10)	-10 (36)	-5 (34)	
		/mamzi/	18 (5)	-53 (17)	18 (7)	10 (4)	-1 (32)		
noninitial	open	/ma:/	16 (6)	-51 (7)	23 (6)	4 (25)	-1 (32)	5 (32)	-1 (30)
		/ma:mi/	32 (15)	-21 (31)	34 (15)	-6 (23)	10 (32)		
	closed	/mam/	3 (18)	-51 (12)	21 (5)	-18 (18)	-12 (30)	-6 (28)	
		/mamzi/	18 (6)	-23 (21)	19 (6)	-12 (19)	0 (32)		
final	open	/ma:/	0 (20)	-54 (8)	-69 (6)	-67 (9)	-46 (33)	-36 (36)	-36 (35)
		/ma:mi/	17 (8)	-41 (24)	-56 (8)	-52 (5)	-24 (36)		
	closed	/mam/	11 (6)	-62 (12)	-60 (5)	-66 (13)	-42 (35)	-36 (35)	
		/mamzi/	20 (8)	-52 (10)	-55 (9)	-51 (11)	-30 (35)		

In phrase-initial and phrase-noninitial position, L is aligned at the boundary between C1 and V1 (on average 2 ms before V1 in phrase-initial position and 1 ms before V1 in phrase-noninitial position). In phrase-final position, speakers align L 36 ms before V1. However, an rmANOVA revealed no effect of POSITION [$F(2,6) = 3.56$, $p \geq 0.05$, $\eta^2=0.52$] but small effects of SYLL [$F(1,3) = 49.79$, $p<0.01$, $\eta^2=0.02$] and WORD LENGTH [$F(1,3)$

= 17.70, $p<0.05$, η^2=0.08]. Specifically, L is aligned slightly later in open syllables as compared to closed syllables, and it aligns later in disyllables than in monosyllables. In phrase-initial position, L aligns 6 ms later in open syllables as compared to closed syllables. In phrase-noninitial position, it aligns 11 ms later in open syllables as compared to closed syllables. In all phrasal positions, L aligns later in disyllabic target words than in monosyllabic ones. Averaged across all positions, it aligns 51 ms before the onset of V1 in /ma:/ while it aligns only 3 ms before the onset of V1 in /ma:mi/. The same pattern applies to closed syllables. L aligns 21 ms before V1 in /mam/ while it aligns 10 ms before it in /mamzi/.

The alignment of L is, however, highly speaker-specific. For example, in all phrasal positions, Speaker S1 aligns L after the acoustic onset of V1 while speaker S2 aligns L before it. In non-final positions, speakers S3 and S4 align L at the boundary between C1 and V1 and shift it leftwards into C1 in phrase-final position.

The subtle effects of SYLL and WORD LENGTH possibly show that L is not aligned at a fixed distance from the onset of the accented vowel. Instead, it could be the case that L aligns at a fixed proportion of the vowel duration. Thus, the alignment of L as a function of the vowel duration was investigated (Figure 5.2). This measurement sets the vowel duration to 100 % (or 1). Values below zero indicate that L is aligned before the acoustic onset of the vowel, and values between zero and one indicate that it is aligned within the accented vowel. For example, a value of 0.5 would indicate that L is aligned in the middle (50 %) of the accented vowel. Data for all speakers are given in Table 5.2. These were submitted to an rmANOVA with the same factors used in the analysis above. Unlike for the absolute measurement, none of the factors reached significance (POSITION: [$F(2,6)$ = 2.80, $p\geq0.05$, η^2=0.39, SYLL [$F(1,3)$ = 3.32, $p\geq0.05$, η^2=0.18] and WORD LENGTH [$F(1,3)$

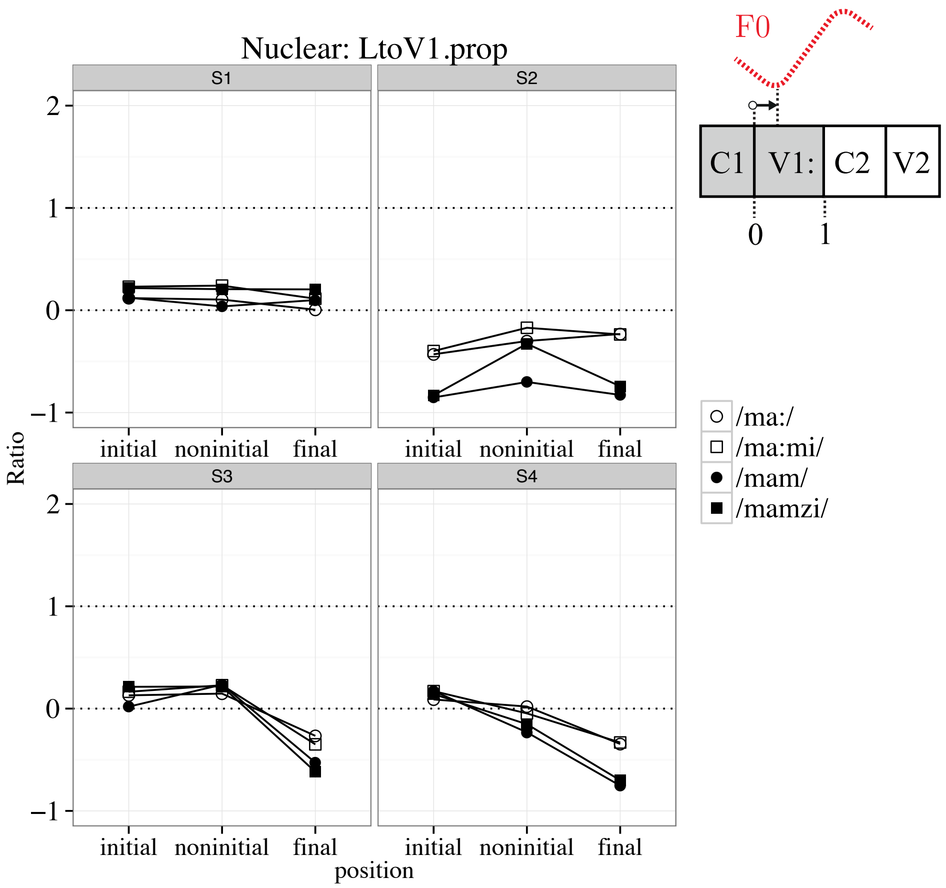


Figure 5.2: Mean alignment lags for L relative to the onset of the accented vowel as a function of the vowel duration (LtoV1.prop).

$= 6.27$, $p \geq 0.05$, $\eta^2=0.03$]), showing evidence of a stable alignment for L relative to the vowel. However, there were speaker-specific differences. As with the absolute alignment of L relative the vowel onset, speakers S3 and S4 align L earlier in phrase-final position while speakers S1 and S2 do not adjust alignment of L as a function of position in the phrase.

The stable alignment of L is further supported by the analysis of its alignment as a proportion of the entire syllable duration. This analysis not only takes different vowel durations into account but also considers durational differences due to different syllable

Table 5.2: Mean alignment lags for L relative to the onset of the accented vowel as a function of the vowel duration (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	0.12 (0.04)	-0.43 (0.06)	0.13 (0.08)	0.09 (0.02)	-0.03 (0.25)	0.00 (0.26)	-0.06 (0.40)
		/ma:mi/	0.23 (0.08)	-0.40 (0.08)	0.16 (0.06)	0.17 (0.11)	0.03 (0.28)		
	closed	/mam/	0.12 (0.08)	-0.85 (0.32)	0.02 (0.42)	0.17 (0.13)	-0.17 (0.51)	-0.11 (0.49)	
		/mamzi/	0.22 (0.06)	-0.83 (0.29)	0.21 (0.09)	0.14 (0.05)	-0.05 (0.47)		
noninitial	open	/ma:/	0.10 (0.04)	-0.30 (0.04)	0.15 (0.03)	0.02 (0.16)	0.00 (0.19)	0.03 (0.22)	-0.04 (0.31)
		/ma:mi/	0.24 (0.13)	-0.17 (0.26)	0.23 (0.10)	-0.05 (0.18)	0.07 (0.24)		
	closed	/mam/	0.04 (0.20)	-0.70 (0.18)	0.23 (0.05)	-0.23 (0.24)	-0.18 (0.39)	-0.10 (0.36)	
		/mamzi/	0.21 (0.07)	-0.33 (0.31)	0.22 (0.06)	-0.15 (0.26)	-0.02 (0.31)		
final	open	/ma:/	0.01 (0.07)	-0.23 (0.06)	-0.27 (0.03)	-0.35 (0.05)	-0.20 (0.15)	-0.18 (0.19)	-0.32 (0.36)
		/ma:mi/	0.11 (0.05)	-0.24 (0.14)	-0.35 (0.07)	-0.33 (0.03)	-0.14 (0.23)		
	closed	/mam/	0.10 (0.05)	-0.83 (0.19)	-0.53 (0.06)	-0.75 (0.14)	-0.47 (0.40)	-0.44 (0.42)	
		/mamzi/	0.20 (0.08)	-0.74 (0.15)	-0.62 (0.12)	-0.70 (0.17)	-0.40 (0.45)		

structures. Figure 5.3 displays the alignment of L as a proportion of syllable duration. Zero denotes the onset of the accented syllable, and one denotes the end of the accented syllable. In target words with an open syllable, the end of the accented syllable refers to the acoustic offset of the vowel /a:/. In closed syllables, it refers to the end of the coda consonant /m/. Table 5.3 provides full data for each condition.

In all phrasal positions, speakers show a stable alignment in that they align L within the

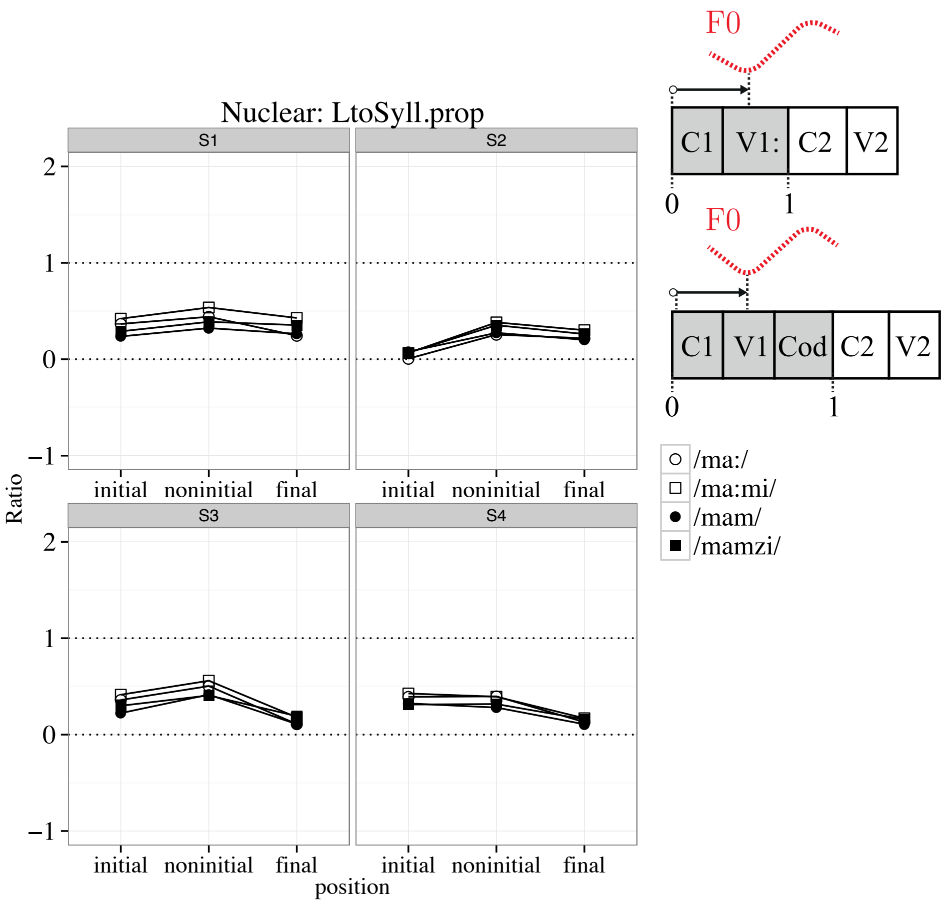


Figure 5.3: Mean alignment lags for L as a function of the syllable duration (LtoSyll.prop).

first half of the syllable. Phrase-initially, L aligns at 26 % of the syllable duration while phrase-noninitially, it aligns at 39 % of the syllable duration and phrase-finally, it aligns at 22 % of the duration. The rmANOVA revealed neither an effect of POSITION [$F(2,6) = 3.12$, $p \geq 0.05$, $\eta^2=0.47$] nor one of SYLL [$F(1,3) = 7.93$, $p \geq 0.05$, $\eta^2=0.13$]. There was a subtle but significant effect of WORD LENGTH [$F(1,3) = 18.65$, $p<0.05$, $\eta^2=0.13$], suggesting that L is aligned slightly earlier in monosyllables than in disyllables. Averaged across all positions and both syllable structures, L aligns at 25 % of the syllable duration in monosyllables as compared to 34 % of the syllable duration in disyllables.

Taken together, these findings show evidence of a stable alignment for the beginning of the accentual rise in the acoustic domain. Both absolute and proportional measurements show a relatively stable alignment of L with the beginning of the accented vowel. Of the three applied measurements, the alignment of L measured as the proportion of the accented vowel sticks out as none of the investigated factors reached significance here.

Table 5.3: Mean alignment lags for L as a function of the syllable duration (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	0.37 (0.04)	0.01 (0.05)	0.36 (0.05)	0.39 (0.04)	0.27 (0.17)	0.30 (0.17)	0.26 (0.14)
		/ma:mi/	0.42 (0.06)	0.07 (0.03)	0.41 (0.03)	0.43 (0.07)	0.33 (0.17)		
	closed	/mam/	0.24 (0.04)	0.08 (0.04)	0.22 (0.08)	0.32 (0.01)	0.21 (0.10)	0.23 (0.11)	
		/mamzi/	0.29 (0.04)	0.07 (0.02)	0.30 (0.05)	0.31 (0.05)	0.24 (0.11)		
noninitial	open	/ma:/	0.44 (0.03)	0.26 (0.02)	0.50 (0.03)	0.40 (0.09)	0.40 (0.10)	0.44 (0.12)	0.39 (0.11)
		/ma:mi/	0.54 (0.09)	0.38 (0.15)	0.56 (0.06)	0.39 (0.09)	0.47 (0.13)		
	closed	/mam/	0.32 (0.06)	0.27 (0.02)	0.41 (0.02)	0.28 (0.07)	0.32 (0.07)	0.34 (0.07)	
		/mamzi/	0.39 (0.02)	0.35 (0.08)	0.41 (0.04)	0.32 (0.09)	0.36 (0.07)		
final	open	/ma:/	0.24 (0.05)	0.22 (0.02)	0.11 (0.02)	0.13 (0.03)	0.17 (0.07)	0.23 (0.12)	0.22 (0.10)
		/ma:mi/	0.43 (0.03)	0.30 (0.08)	0.18 (0.03)	0.17 (0.03)	0.30 (0.13)		
	closed	/mam/	0.26 (0.02)	0.20 (0.02)	0.11 (0.02)	0.11 (0.05)	0.17 (0.08)	0.21 (0.09)	
		/mamzi/	0.35 (0.03)	0.26 (0.03)	0.19 (0.03)	0.16 (0.04)	0.36 (0.09)		

5.1.2 Acoustic alignment of H

This section investigates the alignment of the accentual peak relative to acoustic landmarks. It was measured both from the acoustic offset of the syllable–the most nearby landmark–and as function of the total syllable duration.

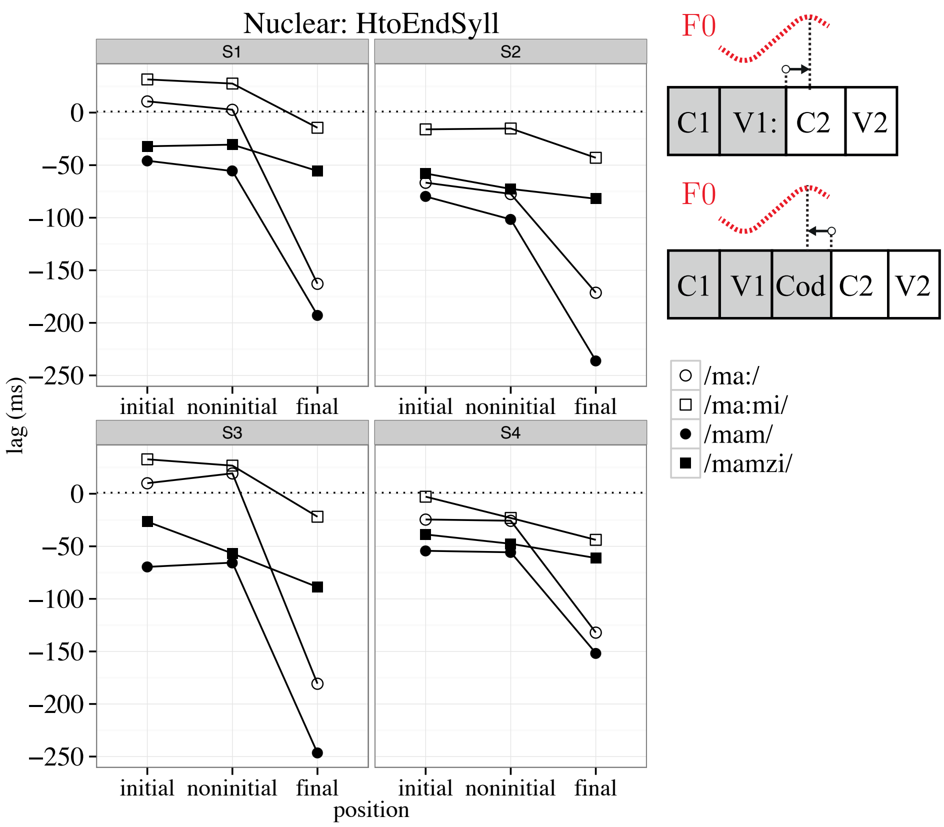


Figure 5.4: Mean alignment lags (in ms) for H relative to the end of the accented syllable (HtoEndSyll).

Figure 5.4 and Table 5.4 provide data for the alignment of the accentual peak relative to the acoustic offset of the accented syllable. In target words with an open syllable, this landmark corresponds to the end of the vowel /a:/ while in target words with a closed syllable, it corresponds to the end of the coda consonant /m/. Positive values indicate that the F0 peak occurs after the syllable boundary, while negative values indicate that it occurs before it.

In contrast to the alignment of L, the alignment of H is highly variable. The rmANOVA confirmed large effects of POSITION [$F(2,6) = 49.45$, $p<0.001$ $\eta^2=0.86$], SYLL [$F(1,3) = 22.91$, $p<0.05$, $\eta^2=0.66$] and WORD LENGTH [$F(1,3) = 58.17$, $p<0.01$, $\eta^2=0.76$], as well as an interaction between WORD LENGTH and POSITION [($F(2,6) = 53.26$, $p<0.001$, $\eta^2=0.69$].

Table 5.4: Mean alignment lags (in ms) for H relative to the acoustic offset of the accented syllable (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	11 (21)	-67 (17)	10 (6)	-24 (8)	-18 (36)	-4 (33)	-28 (36)
		/ma:mi/	32 (4)	-16 (4)	33 (9)	-3 (14)	11 (23)		
	closed	/mam/	-46 (9)	-80 (21)	-70 (20)	-54 (9)	-62 (21)	-50 (21)	
		/mamzi/	-32 (8)	-58 (9)	-27 (7)	-39 (11)	-38 (15)		
noninitial	open	/ma:/	3 (18)	-77 (24)	19 (5)	-26 (12)	-20 (40)	-7 (35)	-35 (40)
		/ma:mi/	28 (7)	-15 (7)	27 (11)	-23 (20)	4 (27)		
	closed	/mam/	-56 (14)	-102 (19)	-66 (8)	-56 (14)	-70 (24)	-61 (24)	
		/mamzi/	-30 (10)	-73 (10)	-57 (18)	-48 (17)	-51 (21)		
final	open	/ma:/	-163 (28)	-171 (19)	-181 (9)	-132 (8)	-162 (25)	-101 (71)	-122 (75)
		/ma:mi/	-14 (5)	-43 (30)	-22 (6)	-44 (12)	-28 (18)		
	closed	/mam/	-193 (20)	-236 (11)	-247 (12)	-152 (11)	-202 (41)	-140 (74)	
		/mamzi/	-56 (15)	-82 (8)	-89 (14)	-61 (12)	-70 (18)		

On average, H aligns 28 ms before the syllable boundary in phrase-initial position and 35 ms before it in phrase-noninitial position. Phrase-finally, though, H is shifted leftwards and aligns 122 ms before the syllable boundary. Posthoc tests showed that the difference between the peak alignment in phrase-initial and phrase-noninitial position was not significant while the two groups significantly differ from the peak alignment in phrase-final position.

However, the magnitude of this leftward shift depends on the WORD LENGTH. While there is a rather small position-induced peak shift in disyllables, there is huge shift in monosyllables. In the phrase-initial /mamzi/, for example, the accentual peak aligns, on average, 38 ms before the syllable boundary. In phrase-final position, it aligns 32 ms earlier, resulting in an alignment of 70 ms before the syllable boundary. In the phrase-initial /ma/, on the other hand, it aligns, on average, 18 ms before the syllable boundary. In phrase-final position, it aligns 144 ms earlier, i.e. it occurs 162 ms before the syllable boundary. Thus, monosyllabic target words show a greater position-induced leftward shift of the accentual peak as compared to disyllabic target words.

Next to phrasal position, syllable structure plays a significant role in determining the peak alignment. Relative to the syllable boundary, H aligns later in open than in closed syllables in all three positions. More specifically, it aligns after the accented vowel in open syllables and within the coda consonant in closed syllables. In phrase-initial position, speakers align the F0 peak, on average, 4 ms before the syllable boundary in open syllables, while they align it 50 ms before it in closed syllables. In phrase-noninitial position, similarly, the F0 peak aligns, on average, 7 ms before the syllable boundary in open and 61 ms before it in closed syllables. In phrase-final position, the F0 peak generally aligns earlier overall, but the effect of syllable structure persists in the same direction. Specifically, the F0 peak aligns, on average, 101 ms before the syllable boundary in open syllables and 140 ms before it in closed syllables.

Even though all speakers show later F0 peaks in open syllables with respect to the syllable boundary, they produce somewhat different alignment patterns: Speaker S1 and speaker S2 consistently align H after the syllable boundary in open syllables, while they align it before it in closed syllables. On the other hand, speaker S2 and speaker S4 align *all* accentual F0 peaks before the syllable boundary but produce an alignment difference between open and closed syllables such that the F0 peak aligns later in open syllables.

As indicated by the interaction, the effect of WORD LENGTH was only apparent in phrase-final positions. Here, the peak aligns earlier in monosyllables than in disyllables. In the other phrasal position, the F0 peak alignment does not differ significantly between mono- and disyllablic target words. Phrase-initially, H aligns shortly before or after the syllable boundary in open syllables (18 ms before the boundary in /ma:/ and 11 ms after it in /ma:mi/). Phrase-finally, H aligns significantly earlier in /ma:/ than in /ma:mi/. On average, the F0 peak aligns 162 ms before the syllable boundary in /ma:/ and 28 ms *after* it in /ma:mi/.

In sum, these results indicate that F0 peak alignment is highly sensitive to the prosodic factors investigated. First, it aligns earlier in phrase-final position than in non-final position within the phrase. Alignment differences between initial and non-initial position were not significant. Second, it aligns later with respect to the syllable boundary in target words with an open syllable. More specifically, speakers align the F0 peak roughly at the end of the accented vowel in open syllables and within the coda consonant in closed syllables. Third, the F0 peak aligns earlier in monosyllables as compared to disyllables in phrase-final position, while WORD LENGTH played almost no role in nonfinal position.

To account for possible differences due to influences of speaking rate and syllable structure, the F0 peak position was measured relative to the total syllable duration. Specifically, the F0 peak alignment was measured as the lag between the F0 peak and the syllable onset, and divided by the syllable duration. Figure 5.5 presents the F0 peak alignment as a proportion of the total syllable duration. Full data are provided in Table 5.5.

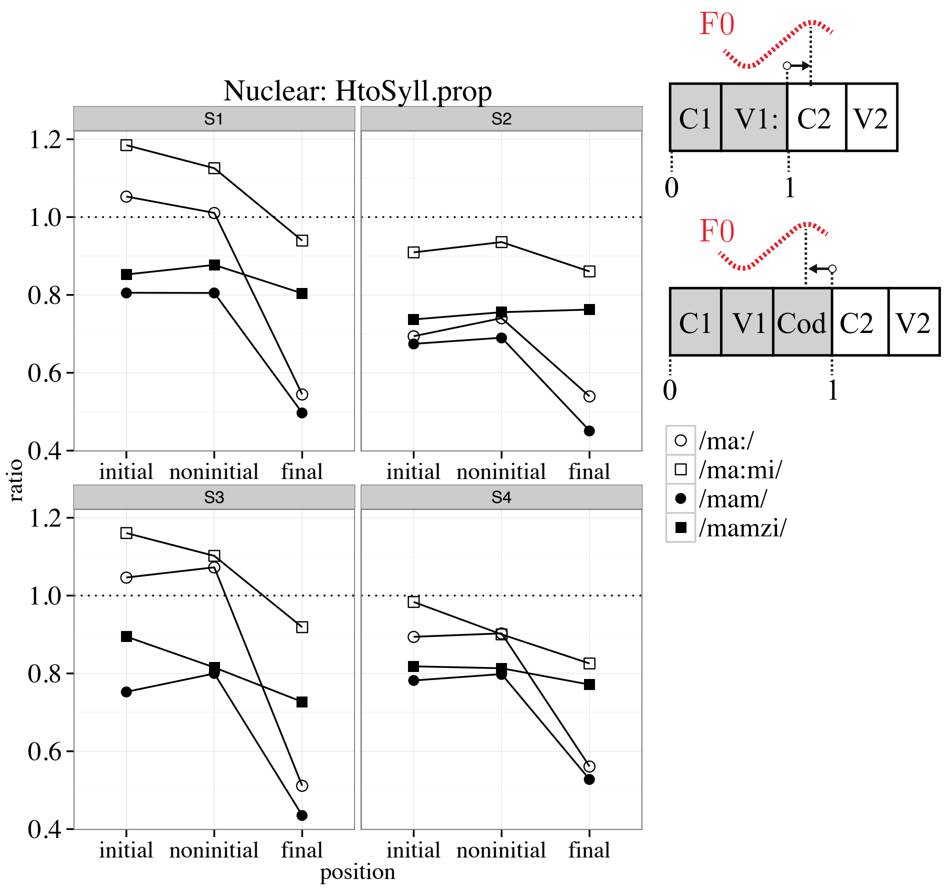


Figure 5.5: Mean alignment lags for H as a function of the syllable duration (HtoSyll.prop).

Zero denotes the acoustic onset of the stressed syllable, i.e. the onset of /m/. One denotes the end of the stressed syllable, i.e. either the offset of /a:/ in /ma:/ and /ma:mi/ or the offset of the second /m/ in /mam/ and /mamzi/. Values above one indicate that the peak aligns outside the stressed syllable.

Table 5.5: Mean alignment lags for H as function of the syllable duration (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	$\bar{x}$	mean	mean
initial	open	/ma:/	1.05 (0.10)	0.69 (0.07)	1.05 (0.03)	0.89 (0.03)	0.92 (0.17)	0.99 (0.16)	0.89 (0.16)
		/ma:mi/	1.18 (0.02)	0.91 (0.02)	1.16 (0.05)	0.98 (0.07)	1.06 (0.13)		
	closed	/mam/	0.81 (0.04)	0.67 (0.07)	0.75 (0.07)	0.78 (0.03)	0.75 (0.07)	0.79 (0.08)	
		/mamzi/	0.85 (0.03)	0.74 (0.04)	0.89 (0.02)	0.82 (0.04)	0.83 (0.07)		
noninitial	open	/ma:/	1.01 (0.07)	0.74 (0.06)	1.07 (0.02)	0.90 (0.05)	0.93 (0.14)	0.98 (0.13)	0.88 (0.14)
		/ma:mi/	1.13 (0.03)	0.94 (0.03)	1.10 (0.04)	0.90 (0.09)	1.02 (0.11)		
	closed	/mam/	0.81 (0.04)	0.69 (0.05)	0.80 (0.03)	0.80 (0.05)	0.77 (0.06)	0.79 (0.07)	
		/mamzi/	0.88 (0.04)	0.76 (0.03)	0.82 (0.05)	0.81 (0.06)	0.82 (0.06)		
final	open	/ma:/	0.54 (0.04)	0.54 (0.04)	0.51 (0.03)	0.56 (0.02)	0.54 (0.03)	0.70 (0.19)	0.66 (0.17)
		/ma:mi/	0.94 (0.02)	0.86 (0.09)	0.92 (0.02)	0.83 (0.04)	0.90 (0.06)		
	closed	/mam/	0.50 (0.04)	0.45 (0.01)	0.44 (0.02)	0.53 (0.03)	0.48 (0.05)	0.62 (0.15)	
		/mamzi/	0.80 (0.05)	0.76 (0.03)	0.73 (0.04)	0.77 (0.04)	0.77 (0.05)		

As expected, the rmANOVA revealed large effects of POSITION [F(2,6) = 24.97, $p<0.01$, η^2=0.77], SYLL [F(1,3) = 19.97, $p<0.05$, η^2=0.67] and WORD LENGTH [F(1,3) = 72.73, $p<0.01$, η^2=0.68] as well as interactions between POSITION and WORD LENGTH [(F(2,6) = 61.62, $p<0.001$, η^2=0.50] and POSITION and SYLL [(F(2,6) = 6.81, $p<0.05$, η^2=0.16]. The results are similar to those in Figure 5.4 where the F0 peak was measured relative to the syllable offset. Speakers align accentual peaks earlier in phrase-final position. The alignment difference between F0 peaks in initial and noninitial position was not significant. On average, speakers align the F0 peak at 89 % of the syllable duration in phrase-initial, at 88 % of the syllable in phrase-noninitial and at 66 % of the syllable in phrase-final position.

In non-final positions, the peak aligns earlier in closed than in open syllables. For example, phrase-initially, H aligns at 79 % of the syllable duration in closed syllables while it aligns at 99 % of the duration in open syllables. The factor WORD LENGTH did not play a significant role in F0 peak alignment in the non-final position.

The picture changes in phrase-final position: Here, syllable structure did not have a significant effect on F0 peak alignment. Instead, WORD LENGTH had an effect on the alignment such that H aligns earlier in mono- than in disyllabic target words. On average, speakers align H at 54 % of the syllable duration in /ma:/ and align it significantly later in /ma:mi/ (at 90 % of syllable duration). The same applies to target words with a closed syllable. In /mam/, H aligns at 48 % of the syllable duration and is pushed rightwards in /mamzi/ (to 77 % of syllable duration).

In sum, these results confirm that there is no stable F0 peak alignment with respect to acoustic landmarks. Peak alignment is highly variable in that it is sensitive not only to phrasal position but also to word length in terms of number of syllables and syllable structure. In general, H aligns earlier in phrase-final position than in nonfinal position (phrase-initial and phrase-noninitial). In the nonfinal positions, speakers align H at the end or shortly after the end of the accented syllable in open syllables (= within or after the stressed vowel). In closed syllables, speakers align H considerably later within the coda (= before the syllable boundary). Word length was only important in phrase-final position: Here, speakers align H earlier in monosyllables than in disyllables.

More specifically, the F0 peak aligns considerably early before the offset of the stressed vowel in /ma:/, shortly before the vowel offset in /ma:mi/ and /mam/, but still in the coda consonant in /mamzi/.

This chapter has shown that the beginning of the accentual rise tends to be stably aligned with the stressed vowel. The alignment of the end of the accentual rise, the F0 peak, however, is highly variable. The next section will examine the alignment of both beginning and end relative to articulatory landmarks.

5.2 Alignment relative to articulatory landmarks

This section examines the alignment of the accentual rise relative to articulatory landmarks. The beginning of the accentual rise was measured relative to its nearby landmark in the articulatory domain, namely the articulatory target of the consonantal closing gesture for /m/ and the peak velocity of its release. The end of the accentual rise, on the other hand, was measured relative to the articulatory target of the vocalic gesture, determined to be the maximum of the tongue body lowering gesture for the production of the accented vowel (/a:/ or /a/).

5.2.1 Articulatory alignment of L

Figure 5.6 displays the mean alignment lags for the beginning of the accentual rise relative to the maximum labial closure, that is the articulatory target for the consonant /m/. Full data are given in Table 5.6. Zero denotes the point in time where the lip closure reaches its maximum. Positive values indicate that L aligns after this landmark. An rmANOVA only detected a small effect of WORD LENGTH ([$F(1,3) = 10.93$, $p<0.05$], $\eta^2=0.05$]). L aligns slightly earlier in monosyllables than in disyllables. For example, phrase-noninitially, speakers align L 50 ms after the maximum closure in /ma:/ while they align it 10 ms later in /ma:mi/.

The factors SYLL and POSITION did not reach significance (SYLL: [$F(1,3) = 2.42$, $p\geq0.05$, $\eta^2=0.01$], POSITION: [$F(2,6) = 4.19$, $p\geq0.05$, $\eta^2=0.56$]. On average, speakers align L 65 ms after the maximum closure in phrase-initial position. In phrase-noninitial position, L is aligned 50 ms after the closure, and in phrase-final position, it is aligned 15 ms after it.

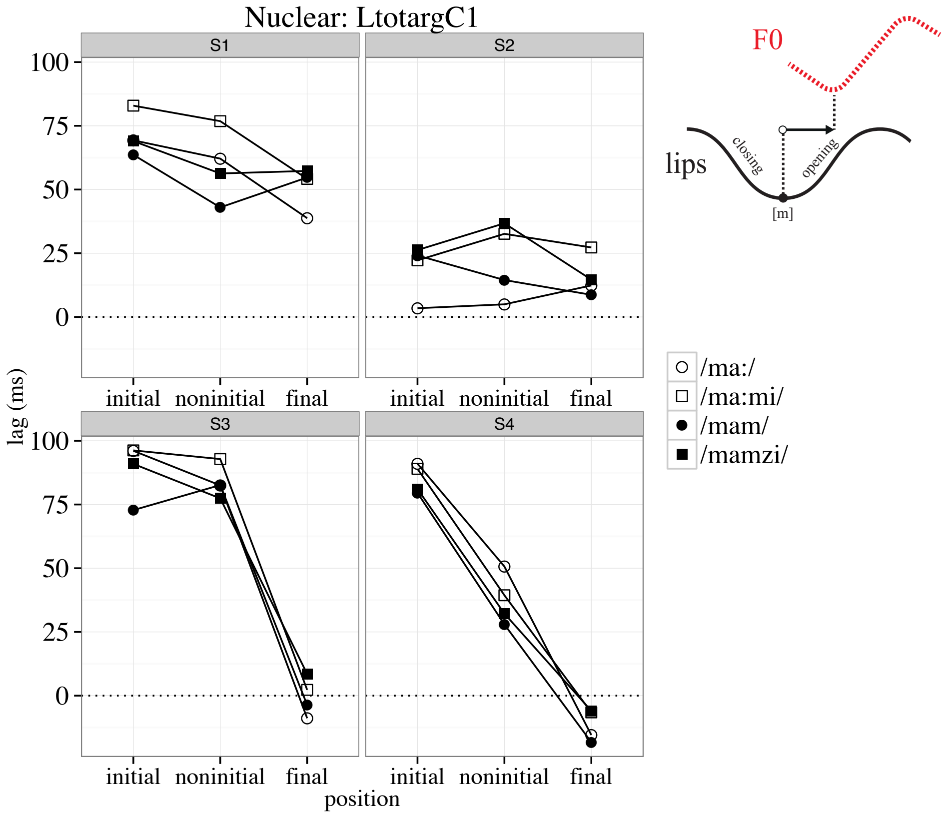


Figure 5.6: Mean alignment lags (in ms) for L relative to the maximum closure in /m/(LtotargC1).

However, as in the acoustics, speakers show different alignment strategies with respect to phrasal position. Speaker S1 and speaker S2 do not (or only slightly) adjust the timing of L, i.e. L is not shifted leftwards in phrase-final position. In contrast, speakers S3 and S4 shift L such that it aligns earlier in phrase-final position. For example, speakers S1 aligns L 69 ms after the maximum closure in the phrase-initial /mamzi/ and she aligns it 57 ms after it in phrase-final position. Speaker S3, in contrast, aligns L 96 ms after the maximum closure in phrase-initial /mamzi/. She then retracts L in phrase-final position, as it aligns only 7 ms after the closure.

In summary, there is evidence for a rather stable alignment for L relative to the maximum labial closure for /m/. Support comes from the rmANOVA indicating only a small effect

Table 5.6: Means alignment lags (in ms) for L relative to the maximum closure in /m/ (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	69 (5)	3 (12)	96 (17)	91 (25)	63 (41)	67 (37)	65 (33)
		/ma:mi/	83 (14)	22 (6)	96 (4)	89 (19)	71 (33)		
	closed	/mam/	64 (9)	24 (17)	73 (40)	80 (8)	58 (30)	63 (28)	
		/mamzi/	69 (9)	26 (11)	91 (9)	81 (8)	67 (26)		
noninitial	open	/ma:/	62 (4)	5 (15)	83 (11)	51 (25)	50 (31)	55 (33)	50 (30)
		/ma:mi/	77 (17)	33 (35)	93 (16)	39 (22)	60 (33)		
	closed	/mam/	43 (18)	14 (10)	83 (6)	28 (20)	41 (29)	45 (26)	
		/mamzi/	56 (7)	37 (24)	77 (6)	32 (19)	49 (23)		
final	open	/ma:/	39 (24)	12 (9)	-9 (4)	-15 (10)	6 (27)	14 (29)	15 (29)
		/ma:mi/	54 (10)	27 (24)	2 (10)	-7 (5)	25 (29)		
	closed	/mam/	55 (5)	9 (10)	-4 (7)	-18 (14)	11 (31)	16 (30)	
		/mamzi/	57 (12)	15 (7)	9 (9)	-6 (12)	22 (27)		

of WORD LENGTH. The factor POSITION failed to reach significance. This result, however, is only side of the coin. The speaker-specific analysis revealed that indeed two of the speakers (S3 and S4) align L considerably earlier in phrase-final position, that is, shortly before or after the maximum labial closure. For example, speaker S4 aligns L 81 ms after the maximum labial closure in /mamzi/ in phrase-initial position but in phrase-final position, this speaker aligns L 6 ms *before* the maximum labial closure. Averaged across all speakers and syllable structures, however, L aligns after the maximum closure. Thus, it was related to the point in time where the release of the closure reaches its peak velocity.

Figure 5.7 displays the alignment of L relative to the peak velocity of the consonantal release gesture, i.e. the point in time where the lip opening movement reaches its maximum velocity. Full data are given in Table 5.7.

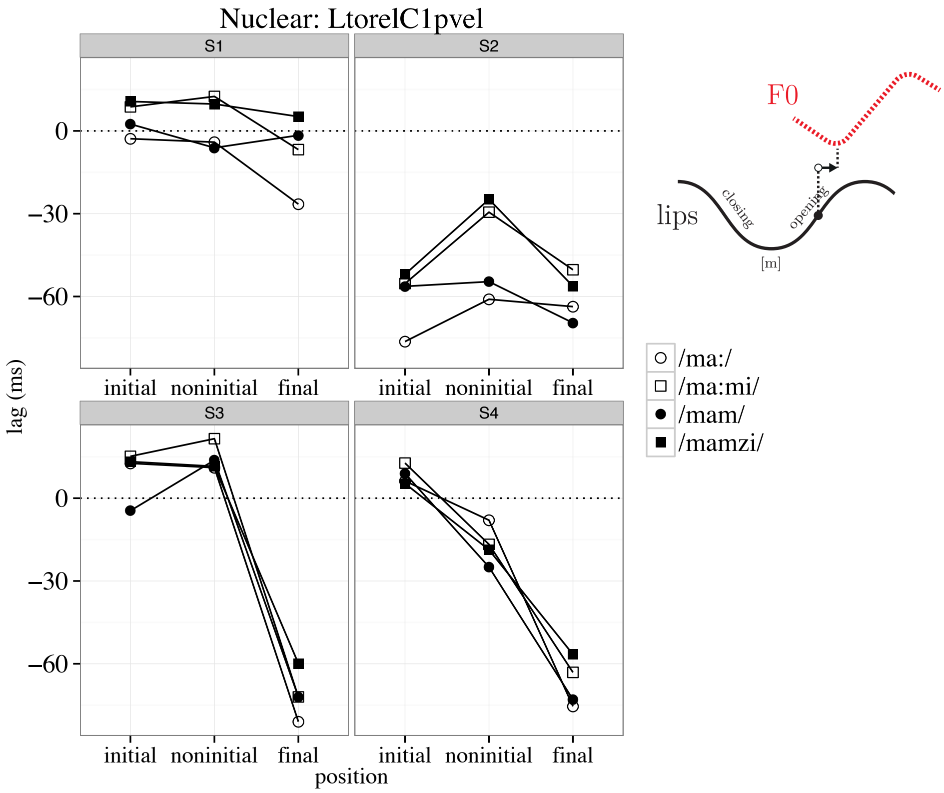


Figure 5.7: Mean alignment lags (in ms) for L relative to the peak velocity of the opening gesture from /m/ to /a/ (LtorelC1pvel).

On average, speakers align L 11 ms before reaching the gesture's peak velocity in phrase-initial and phrase-noninitial position, and 50 ms before it in phrase-final position. Again, the rmANOVA revealed only a medium effect of WORD LENGTH ([$F(2,6) = 13.61$, $p<0.05$, $\eta^2=0.09$]). The factors SYLL and POSITION did not reach significance (SYLL: [$F(1,3) =$

0.71, p≥0.05, η^2=0.00], POSITION: [F(2,6) = 4.21, p≥0.05, η^2=0.55]). More specifically, L aligns slightly earlier in monosyllables than in disyllables with respect to this landmark. In the phrase-initial /ma:/, L aligns on average 17 ms before the lip opening gesture reaches its peak velocity. In /ma:mi/, it aligns only 6 ms before it. The same pattern holds for target words with a closed syllable, too. In the phrase-final /mam/, for example, L aligns 52 ms before the peak velocity, while in /mamzi/ it aligns 38 ms before it.

Table 5.7: Mean alignment lags (in ms) for L relative to the peak velocity of the opening gesture from /m/ to /a/ (standard deviation in parentheses)

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	-3 (6)	-76 (11)	13 (15)	6 (3)	-17 (38)	-12 (35)	-11 (33)
		/ma:mi/	9 (10)	-55 (7)	15 (10)	13 (16)	-6 (32)		
	closed	/mam/	2 (8)	-56 (21)	-5 (38)	9 (9)	-14 (34)	-10 (32)	
		/mamzi/	11 (4)	-52 (17)	13 (7)	5 (3)	-5 (29)		
noninitial	open	/ma:/	-4 (7)	-61 (9)	11 (8)	-8 (25)	-16 (30)	-9 (30)	-11 (28)
		/ma:mi/	12 (16)	-29 (32)	22 (15)	-17 (23)	-3 (30)		
	closed	/mam/	-6 (19)	-55 (13)	14 (5)	-25 (19)	-19 (29)	-13 (26)	
		/mamzi/	10 (7)	-25 (20)	12 (6)	-19 (19)	-6 (22)		
final	open	/ma:/	-27 (21)	-64 (4)	-81 (4)	-75 (9)	-62 (26)	-52 (30)	-48 (32)
		/ma:mi/	-7 (7)	-50 (22)	-72 (8)	-63 (5)	-40 (30)		
	closed	/mam/	-2 (5)	-70 (12)	-72 (6)	-73 (14)	-52 (34)	-45 (33)	
		/mamzi/	5 (10)	-56 (8)	-60 (7)	-57 (10)	-37 (31)		

Individual speakers, however, use different alignment strategies. Again, speakers S1 and

S2 do not (or just slightly) adjust the beginning of the accentual rise according to phrasal position. For example, speaker S1 aligns L 5 ms after the peak velocity in phrase-initial and 3 ms after the peak velocity in phrase-noninitial position. Phrase-finally, she aligns it 7 ms before it. In contrast, speaker S3 aligns L 9 ms and 15 ms after the peak velocity in phrase-nonfinal positions and retracts it up to 72 ms before it in phrase-final position.

Summarising the results for the alignment of L relative to landmarks in the articulatory domain, there is evidence for a stable alignment for L between the maximum lip closure–the articulatory target for /m/–and the peak velocity of the opening gesture. The alignment was affected neither by the phrasal position nor by the syllable structure of the target words. Word length only had a marginal effect such that L aligns slightly earlier in monosyllables than in disyllables. A positional effect was observable for two of the speakers; they align L earlier in phrase-final position. However, this factor did not reach significance in the rmANOVA. Also, the exact timing of L seems to be speaker-specific but constrained to the time interval between the consonantal gesture's maximum constriction and the release's peak velocity.

5.2.2 Articulatory alignment of H

This section examines the alignment of H relative to the articulatory target of the vowel gesture, that is, the time point of the maximum tongue body lowering during the vocalic opening for the accented vowel (/a:/ or /a/). Figure 5.8 displays the mean alignment lags, full data are provided in Table 5.8. Positive values indicate that the F0 peak aligns after the vowel target while negative values indicate that it aligns before it. All lags were submitted to an rmANOVA.

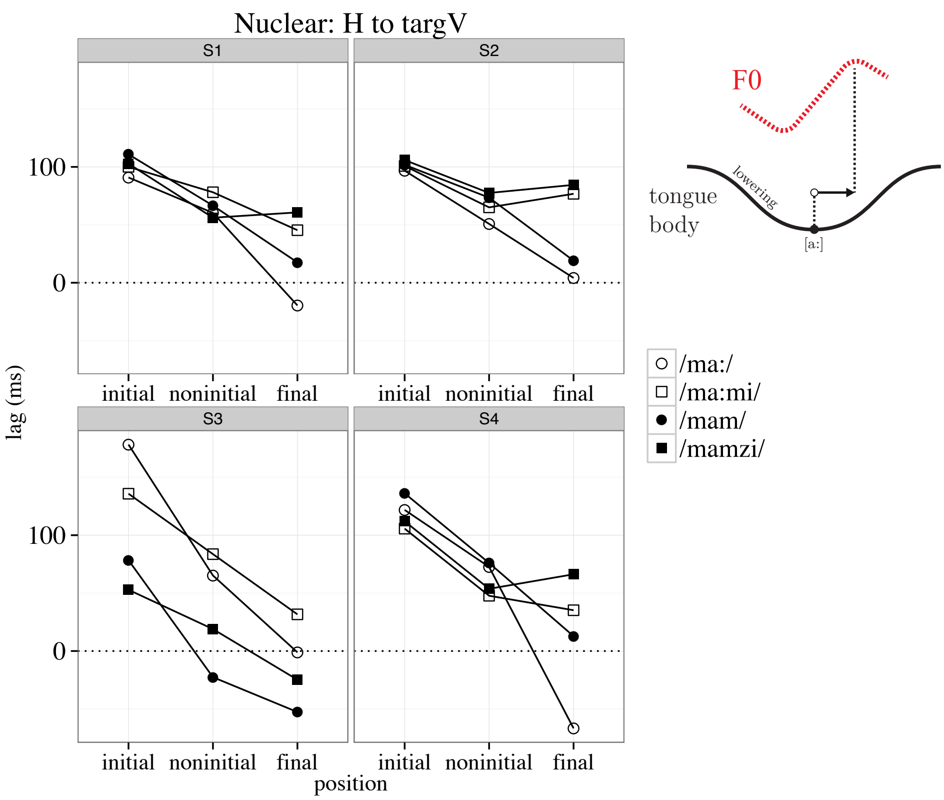


Figure 5.8: Mean alignment lags (in ms) for H relative to the articulatory vowel target for /a/ (HtotargV)

Quite notably, the rmANOVA **did not** reveal an effect of SYLL ([F(1,3) = 0.11, $p \geq 0.05$], η^2=0.02) but did reveal a large effect of POSITION ([F(2,6) = 32.83, $p < 0.001$], η^2=0.69) and a small effect of WORD LENGTH ([F(1,3) = 15.69, p<0.05], η^2=0.10), as well as an interaction between POSITION and WORD LENGTH ([F(2,6) = 10.44, p<0.05], η^2=0.26).

Table 5.8: Mean alignment lags (in ms) for H relative to the articulatory vowel target for /a/ (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	91 (16)	97 (21)	178 (66)	122 (15)	120 (48)	115 (40)	107 (41)
		/ma:mi/	100 (8)	101 (16)	136 (53)	106 (12)	110 (30)		
	closed	/mam/	111 (22)	102 (18)	78 (90)	136 (20)	107 (47)	99 (40)	
		/mamzi/	103 (4)	106 (19)	53 (36)	112 (5)	92 (32)		
noninitial	open	/ma:/	60 (15)	51 (20)	65 (7)	73 (15)	62 (16)	65 (19)	58 (32)
		/ma:mi/	78 (13)	65 (17)	84 (12)	48 (19)	68 (21)		
	closed	/mam/	67 (15)	73 (21)	-23 (15)	76 (21)	51 (44)	52 (39)	
		/mamzi/	56 (31)	78 (11)	19 (27)	54 (36)	53 (34)		
final	open	/ma:/	-20 (15)	4 (19)	-1 (19)	-67 (33)	-22 (35)	8 (45)	16 (50)
		/ma:mi/	45 (4)	77 (37)	32 (11)	35 (15)	45 (21)		
	closed	/mam/	18 (14)	19 (7)	-53 (82)	13 (14)	-1 (50)	23 (53)	
		/mamzi/	61 (17)	84 (10)	-25 (22)	66 (16)	51 (42)		

Posthoc tests showed a three-way distinction between phrasal positions. H aligns significantly earlier in phrase-final position than in phrase-noninitial position and significantly earlier in phrase-initial position. On average, H aligns only 16 ms after the vowel target in phrase-final position while it aligns 58 ms after in phrase-noninitial position and 107 ms after it in phrase-initial position.

The effect of WORD LENGTH is limited to the target words in phrase-final position. Here, H aligns earlier in monosyllables than in disyllables. On average, H aligns 22 ms before the articulatory vowel target in /ma:/ while it aligns 45 ms after it in /ma:mi/. The same applies to target words with a coda consonant. In /mam/, speakers align H 1 ms before the vowel target whereas in /mamzi/, they align it 51 ms after it.

It is worth noting that syllable structure did not play a significant role in determining F0 peak alignment. Phrase-initially, H aligns on average 115 ms after the vowel target in open syllables while it aligns 99 ms after it in closed syllables. Phrase-noninitially, it aligns, on average, 65 ms after the vowel target in open syllables and 52 ms after it in closed syllables. Phrase-finally, H aligns, on average, 8 ms after the vowel target in open syllables and 23 ms after it in closed syllables.

5.3 Summary (acoustics and articulation)

This section summarises the findings on the alignment of the nuclear rise relative to landmarks both in the acoustics and articulation. It starts with a summary of the results for the alignment relative to segmental boundaries, then presents the results for the alignment relative to landmarks in the articulatory trace.

The alignment of the nuclear pitch accent relative to acoustic landmarks in the segmental string reveals two distinct characteristics for the beginning and the end of the accentual rise. While the beginning of the rise tends to be robustly aligned with the onset of the stressed vowel, the accentual peak is sensitive to all prosodic factors under investigation: phrasal position, syllable structure and word length.

Table 5.9 sums up the statistical analyses for the acoustic alignment of L. A significant effect is indicated by the effect size η^2 (n.s. = not significant). Only word length and syllable structure had an effect on the alignment of L and these were both found to have small effects. No significant interactions were found in the rmANOVAs. In open syllables, L aligns slightly earlier as compared to closed syllables. In monosyllables, too, it aligns slightly earlier as compared to disyllables.

Table 5.9: Overview of the statistical analyses for the acoustic alignment of L.

	LtoV1ons	LtoV1ons.prop	LtoSyll
Position	n.s.	n.s.	n.s.
Syllable structure	$\eta^2 = 0.02$	n.s.	n.s.
Word length	$\eta^2 = 0.08$	n.s.	$\eta^2 = 0.13$

Figure 5.9 illustrates the findings for the stable alignment of L using data from speaker S1.

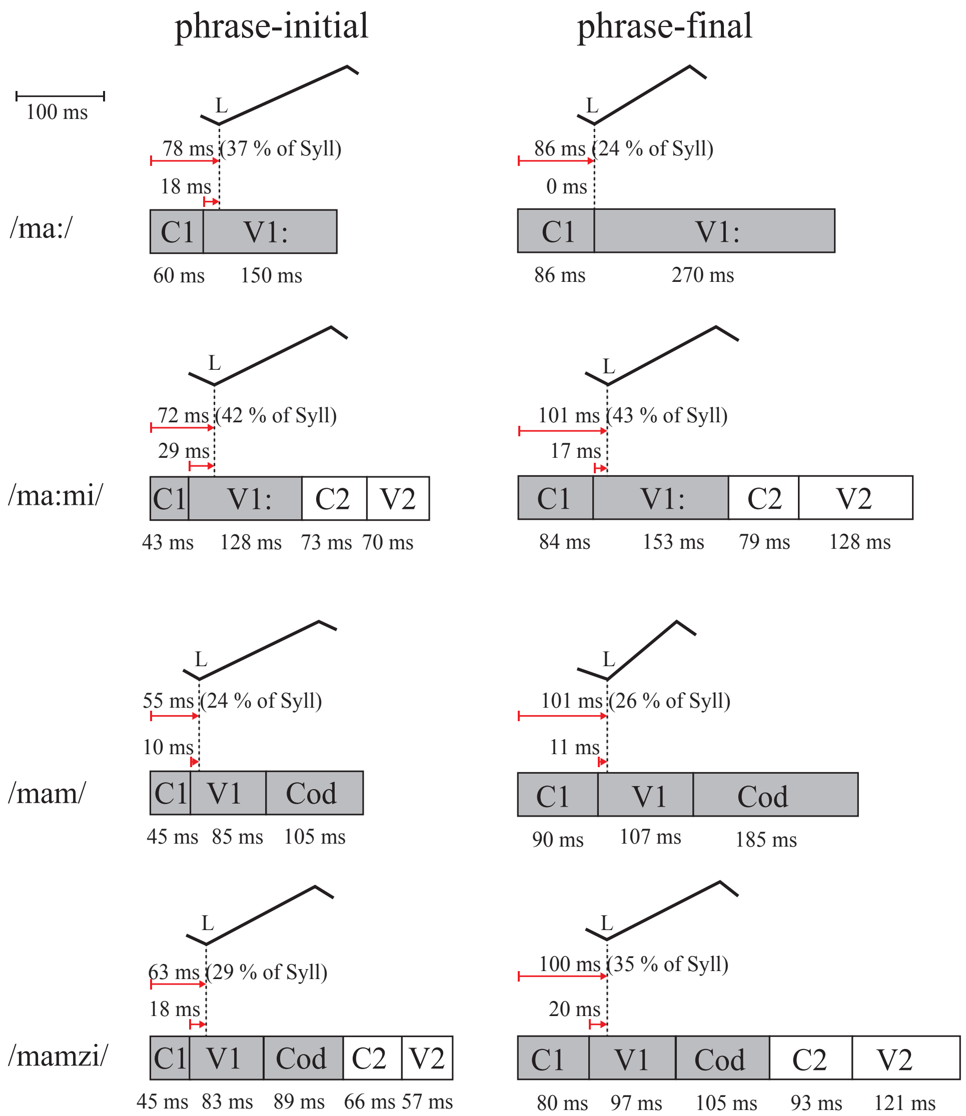


Figure 5.9: Acoustic alignment of L (data from speaker S1).

Each box corresponds to a segment of the target word; the accented syllable is shaded. Values below each box indicate the segmental duration averaged across all repetitions. The dotted line displays the alignment of L. The red arrows indicate the absolute alignment of L relative to the acoustic onset of the target word and the onset of the accented vowel. The values in parentheses indicate the proportional alignment of L relative to the total syllable duration. The alignment of L in phrase-initial[1] position is displayed on the left-hand side, the alignment in phrase-final position is shown on the right-hand side. In all target words, this speaker aligns L shortly after the acoustic onset of the stressed vowel. This alignment lag between L and the vowel onset was affected neither by the phrasal position nor by the word length, nor by the syllable structure, indicating that this landmark is an appropriate anchor point for the onset of the accentual rise.

Table 5.10 sums up the statistical analyses for the acoustic alignment of H. All factors under investigation had an effect on the alignment of H. More specifically, H aligns earlier in phrase-final position as compared to the nonfinal positions. With respect to the syllable boundary, it aligns earlier in closed as compared to open syllables, and it aligns earlier in monosyllables than in disyllables.

Table 5.10: Overview of the statistical analyses for the acoustic alignment of H.

	HtoEndSyll	HtoSyllDur
Position	$\eta^2 = 0.86$	$\eta^2 = 0.77$
Position x Syllable Structure	n.s.	$\eta^2 = 0.16$
Position x Word length	$\eta^2 = 0.69$	$\eta^2 = 0.50$
Syllable structure	$\eta^2 = 0.66$	$\eta^2 = 0.67$
Word length	$\eta^2 = 0.76$	$\eta^2 = 0.68$

Figure 5.10 exemplifies the highly variable timing for H relative to acoustic landmarks, again using data from speaker S1. Here, the dotted line represents the timing of the accentual peak. The red arrows show the absolute peak alignment relative to the onset of the target word and relative to end of the accented vowel. In addition, alignment lags between the F0 peak and the end of the accented syllable are presented for target words

[1]The results from phrase-noninitial position were not included in this figure because they did not differ significantly from the phrase-initial ones.

with a closed syllable.

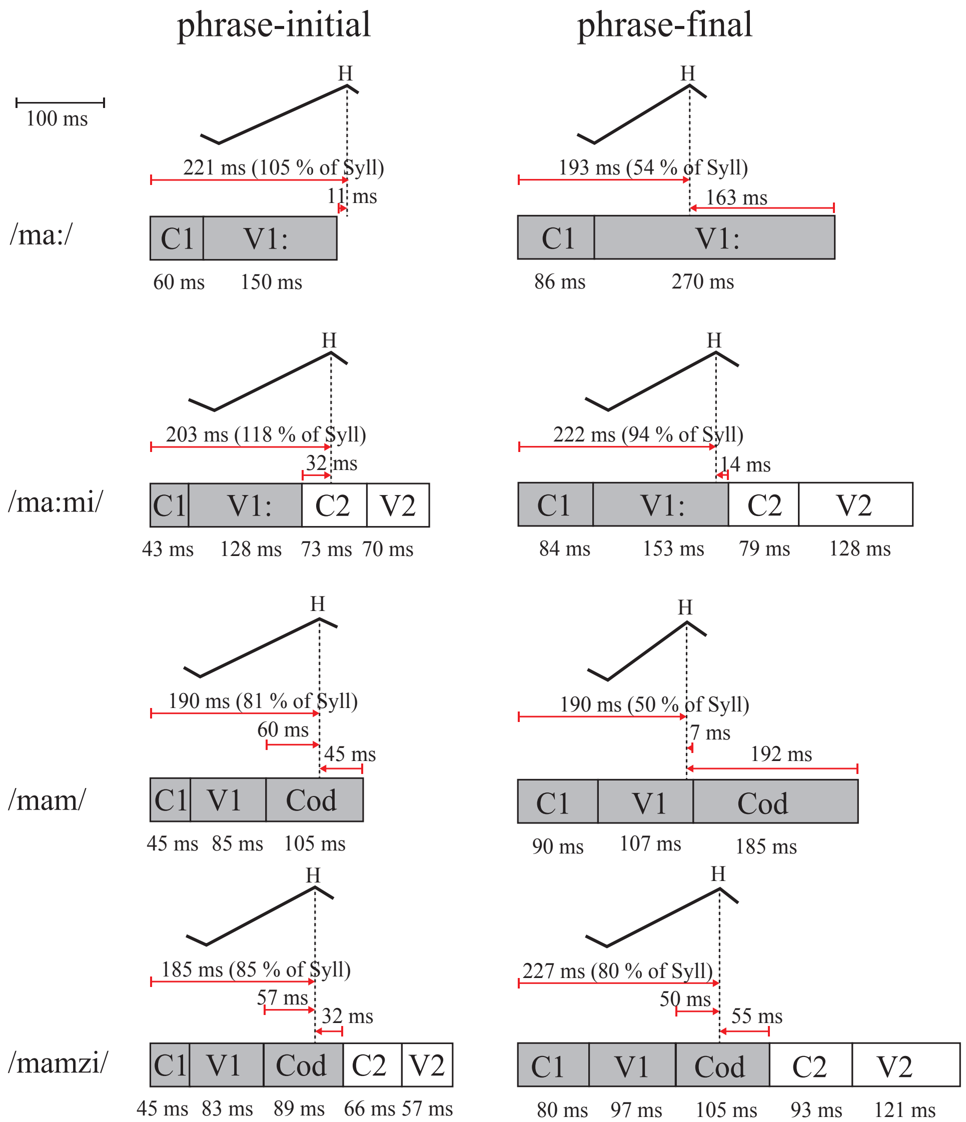


Figure 5.10: Acoustic alignment of H (data from speaker S1).

Phrase-initially, the accentual peak aligns after the stressed vowel in open syllables (11 ms after in /ma:/ and 32 ms after in /ma:mi/) while in closed syllables, it aligns in the coda consonant (60 ms after the vowel offset in /mam/ and 57 ms after in /mamzi/). Phrase-finally, F0 peaks are shifted leftwards, with monosyllables showing the largest effect. In /ma:/ and /mam/, the peak aligns 163 ms and 192 ms before the syllable offset, respectively, while in /ma:mi/ and /mamzi/ it aligns 14 ms and 55 ms, respectively, before it.

To sum up the results for the alignment in the acoustic domain, there is evidence for a stable alignment for the beginning of the accentual rise relative to the vowel duration of the accented vowel. In contrast, the end of the accentual rise is highly variable. It aligns earlier in phrase-final position than in nonfinal position, earlier in monosyllables than in disyllables and earlier in open syllables than in closed syllables. The next section summarises the findings for the alignment in the articulatory domain.

Table 5.11 sums up the statistical analyses for the articulatory alignment of L. Neither phrasal position nor syllable structure had a significant effect on the alignment of L. It was solely affected by word length such that L aligns slightly earlier in monosyllabic target words as compared to disyllabic ones.

Table 5.11: Overview of the statistical analyses for the articulatory alignment of L.

	LtotargC1	LtorelC1pvel
Position	n.s.	n.s.
Syllable structure	n.s.	n.s.
Word length	$\eta^2 = 0.05$	$\eta^2 = 0.09$

Figure 5.11 illustrates the findings for L using data from speaker S1 for all four target words, in both phrase-initial and phrase-final position. The trajectory represents the lip aperture, i.e. the closing and opening gesture for the word-initial /m/. The arrows indicate the interval between the maximum constriction of the lips and the beginning of the accentual rise.

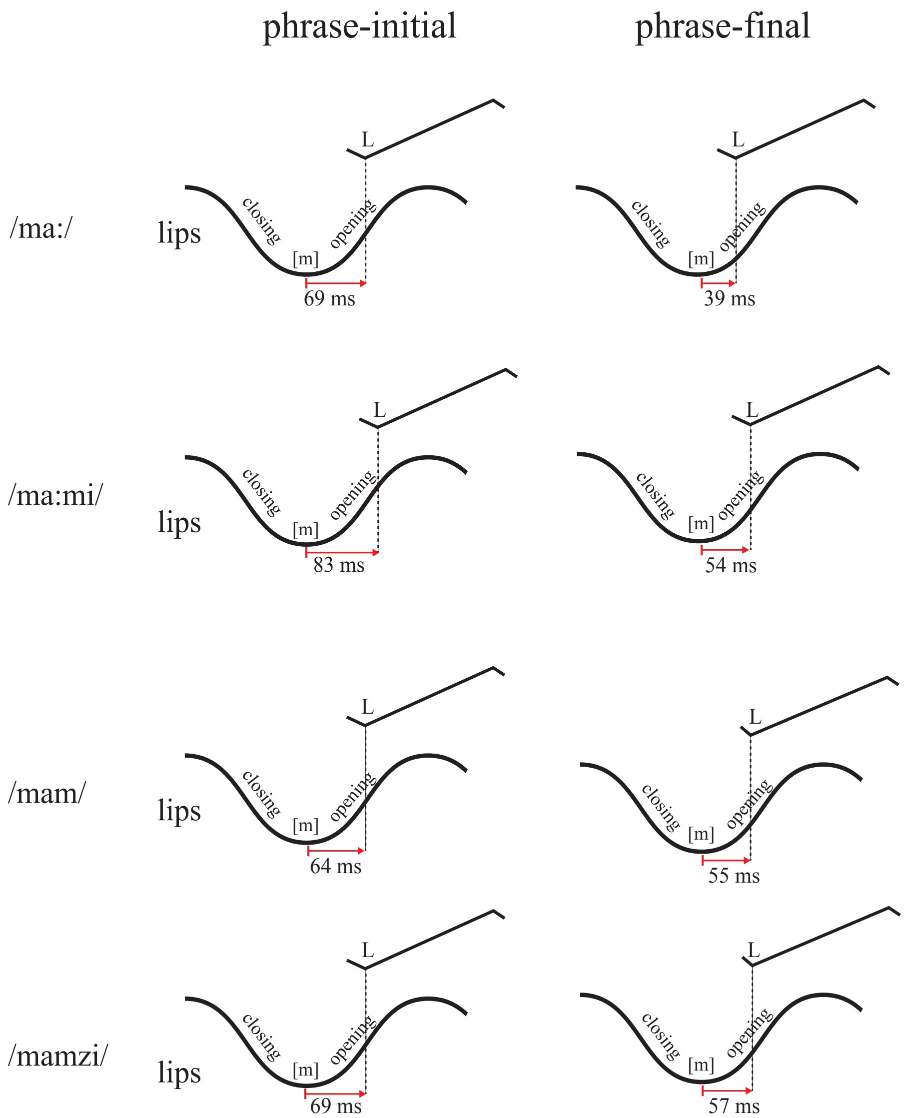


Figure 5.11: Articulatory alignment of L (data from speaker S1).

In all target words, L aligns *after* the maximum closure, i.e. within the consonant's release gesture. The alignment of L roughly coincides with the peak velocity of the release. More specifically, in the disyllables /ma:mi/ and /mamzi/, L aligns 83 ms and 69 ms after the maximum closure, respectively. It aligns slightly earlier in the monosyllables, being aligned 69 ms and 64 ms after the maximum closure in /ma:/ and /mam/, respectively.

Phrasal position, on the other hand, did not have any effect on the alignment of L.

Finally, Table 5.12 sums up the statistical analyses for the articulatory alignment of H. In contrast to the articulatory alignment of L, the articulatory alignment of H was significantly affected by position, with an interaction between position and word length. In general, the F0 peak aligns earlier in phrase-final position than in phrase-initial or noninitial position. The effect of word length was constrained to target words in phrase-final position. In this position, H aligns earlier in monosyllables as compared to disyllables.

Table 5.12: Overview of the statistical analyses for the articulatory alignment of H.

	HtotargV
Position	$\eta^2 = 0.69$
Syllable structure	n.s.
Word length	$\eta^2 = 0.10$
Position x Word length	$\eta^2 = 0.26$

Figure 5.12 illustrates the findings for H using data from speaker S1. In this figure, the end of the accentual rise was related to the articulatory target of the vowel gesture, i.e. the maximum tongue body lowering in the vertical dimension for the production of the accented vowel (/a:/ or /a/). Arrows indicate the time interval between the articulatory vowel target and the F0 peak.

The figure clearly illustrates the positional effect on F0 peak alignment with respect to the articulatory vowel target. On average, this speaker aligns H between 91 ms and 111 ms after the vowel target in phrase-initial position, between 56 ms and 78 ms after it in phrase-noninitial position and between -20 ms and 61 ms after it in phrase-final position. In other words, the F0 peak is pushed leftwards from phrase-initial to phrase-final position.

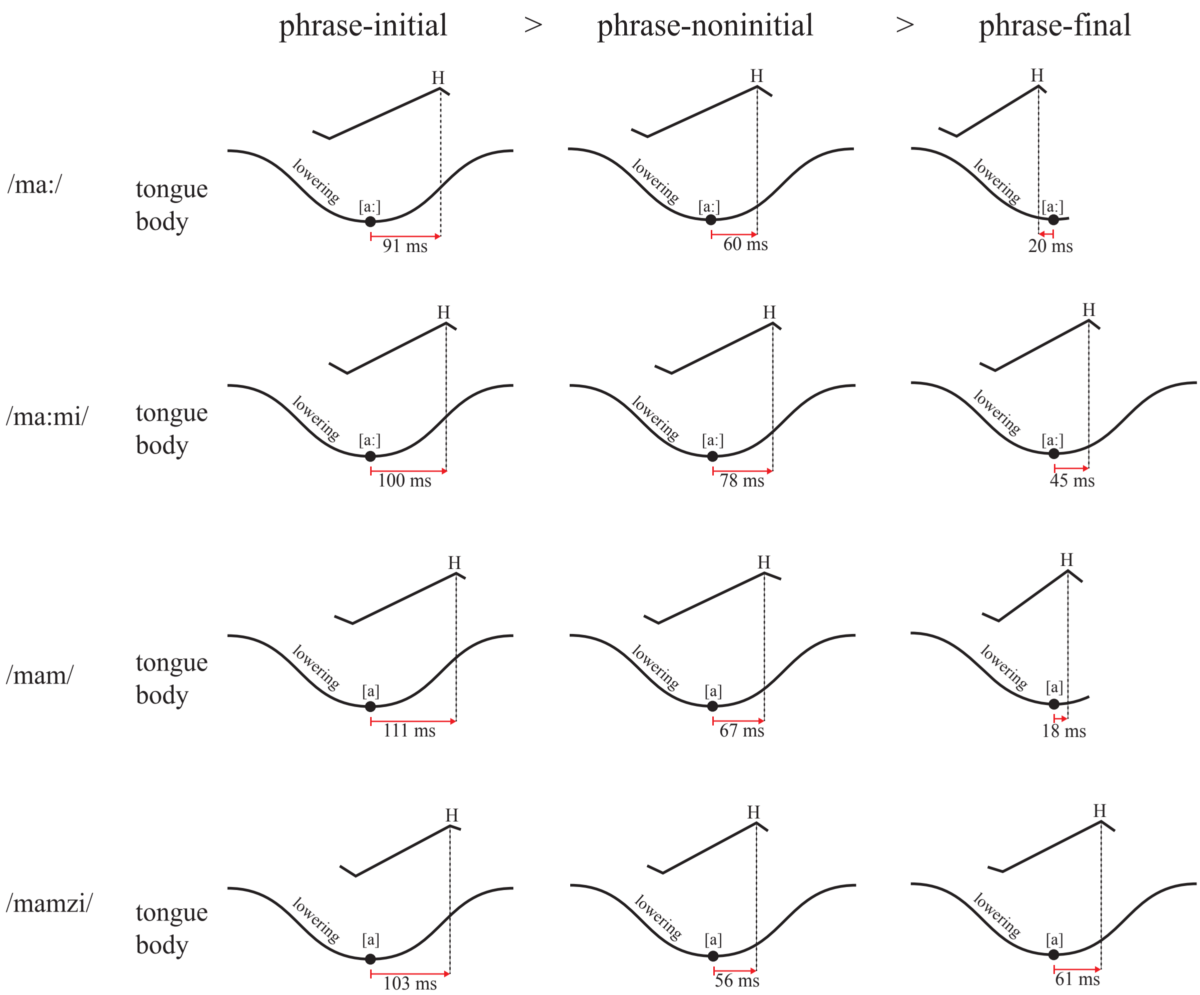


Figure 5.12: Articulatory alignment of H (data from speaker S1).

In final position, word length is a significant F0 peak alignment predictor in that H aligns earlier in monosyllables than in disyllables. On average, speaker S1 aligns H 45 ms after the gestural vowel target in /ma:mi/. In its monosyllabic counterpart /ma:/, however, H aligns 20 ms *before* the vocalic gesture reaches its target. The same applies to the closed syllables /mam/ and /mamzi/. In /mamzi/, the speaker aligns H 61 ms after the vowel target, while she aligns H 18 ms after it in /mam/.

In sum, speakers show a stable yet speaker-specific alignment for the beginning of the nuclear rise relative to the consonantal release gesture. There was neither an effect of phrasal position nor of syllable structure on the alignment of L relative to the articulatory landmarks (L to both the maximum lip constriction and the peak velocity of the release gesture). The factor word length only played a minor role, resulting in a marginally earlier alignment in monosyllables than in disyllables.

The end of the nuclear rise displays a stable alignment relative to the articulatory target of the vocalic gesture in that it was not affected by syllable structure. However, phrasal position and word length did have an effect on F0 peak alignment. It aligns earlier in phrase-final position as compared to non-final positions. Furthermore, in phrase-final position it aligns earlier in monosyllables than in disyllables.

6 Prenuclear accents

This chapter presents the results of the investigation into the alignment of prenuclear pitch accents relative to acoustic and articulatory landmarks. Section 6.1 investigates the alignment in the acoustic domain while Section 7.2 investigates the alignment in the articulatory domain. Section 6.3 summarises the findings.

6.1 Alignment relative to acoustic landmarks

This section examines the prenuclear pitch accent alignment relative to acoustic landmarks. As in the analysis of the nuclear accents, the beginning of the prenuclear rise was measured in both absolute (relative to the acoustic onset of the accented vowel) and relative terms (as a proportion of the vowel duration and as a proportion of the total syllable duration, respectively). The end of the prenuclear rise was measured relative to the end of the accented syllable and as a proportion of the total syllable duration.

6.1.1 Acoustic alignment of L

In this section, we look at the start of the prenuclear rise, first in absolute terms relative to the acoustic onset of the accented vowel, then in proportional terms relative to the vowel duration and the total syllable duration. Figure 6.1 shows the alignment lag from the beginning of the rise relative the beginning of the accented vowel. Full data are given in Table 6.1. Zero denotes the onset of the vowel. Negative values indicate that L aligns before the onset, while positive values indicate that it aligns after the onset.

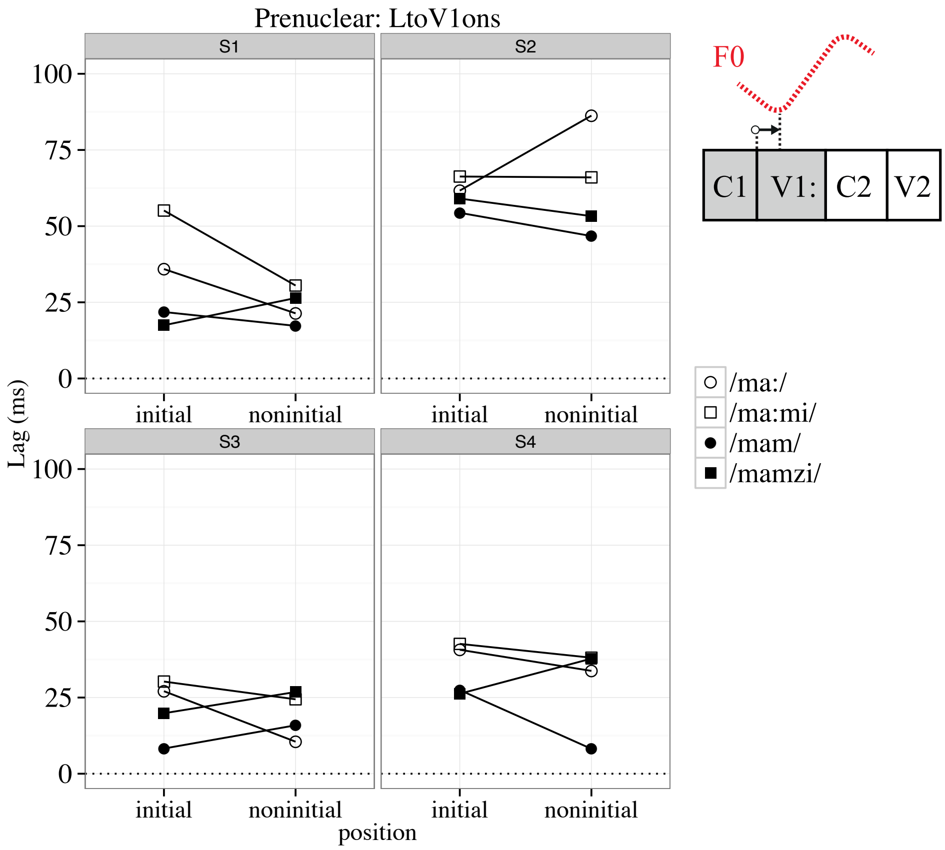


Figure 6.1: Mean alignment lags (in ms) for L relative to the acoustic vowel onset (LtoV1ons).

In general, all speakers align L after the acoustic vowel onset. An rmANOVA did not reveal any effect of WORD LENGTH [F(1,3) = 6.43, $p \geq 0.05$, η^2=0.18] nor of POSITION

[F(1,3) = 1.75, p≥0.05, η^2=0.05]. On average, L aligns 37 ms after the vowel onset in phrase-initial position and 34 ms after it in phrase-noninitial position.

Table 6.1: Mean alignment lags (in ms) for L relative to acoustic vowel onset (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	36 (20)	62 (13)	27 (12)	41 (28)	41 (22)	46 (27)	37 (27)
		/ma:mi/	55 (41)	66 (22)	30 (7)	43 (28)	50 (30)		
	closed	/mam/	22 (13)	54 (12)	8 (8)	27 (11)	28 (20)	29 (26)	
		/mamzi/	17 (7)	59 (52)	20 (13)	26 (13)	30 (31)		
noninitial	open	/ma:/	21 (6)	86 (14)	11 (24)	34 (12)	38 (34)	39 (30)	34 (26)
		/ma:mi/	31 (14)	66 (30)	24 (28)	38 (10)	39 (26)		
	closed	/mam/	17 (4)	47 (14)	16 (10)	8 (32)	23 (22)	29 (20)	
		/mamzi/	26 (10)	53 (14)	27 (9)	28 (16)	36 (16)		

Syllable structure, however, did have large effect on the alignment of L (SYLL [F(1,3) = 25.45, p<0.05, η^2=0.46]), such that L aligns earlier in closed syllables as compared to open syllables. In phrase-initial position, L aligns on average 46 ms after the vowel onset in open syllables and 29 ms after it in closed syllables. The same applies to target words in non-initial position. Here, L aligns 39 ms after the acoustic vowel onset in open syllables and 29 ms after it in closed syllables. This alignment difference between open and closed syllables could indicate that the vowel onset may not be an appropriate anchor for L. Thus, the alignment of L relative to total vowel duration was investigated.

Figure 6.2 presents the alignment of L relative to the acoustic vowel onset as a proportion of the accented vowel duration. Data are provided in Table 6.2. Zero denotes the vowel onset while, one denotes the vowel offset. In general, all speakers align L well within the accented vowel.

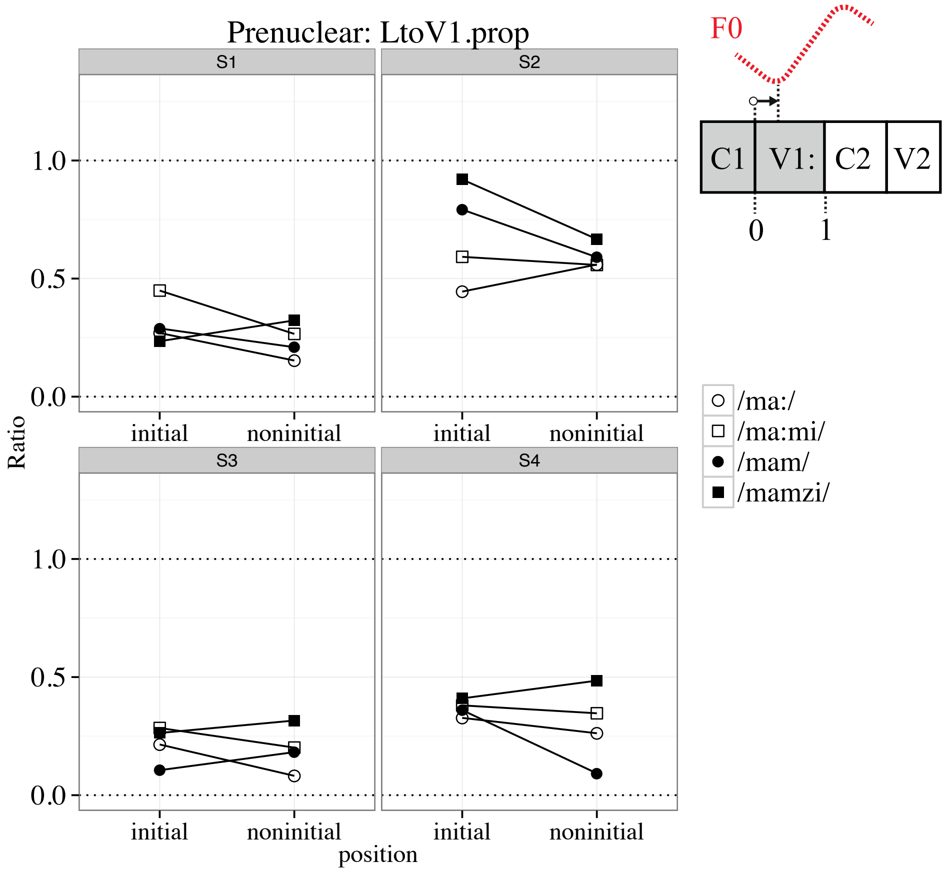


Figure 6.2: Mean alignment lags for L as a proportion of the vowel duration (LtoV1.prop).

This time, the rmANOVA showed no effect of SYLL [F(1,3) = 1.08, p≥0.05, η^2=0.11] but did show effects WORD LENGTH [F(1,3) = 62.67, p<0.01, η^2=0.34] and POSITION [F(1,3) = 18.11, p<0.05, η^2=0.15]. Overall, L aligns slightly earlier in monosyllables as compared to disyllables, and slightly earlier in phrase-noninitial position as compared to phrase-initial position. On average, it aligns at 40 % of the vowel duration in phrase-initial position and at 33 % of the vowel duration in phrase-noninitial position.

The effect of WORD LENGTH is evident in the earlier alignment of L in monosyllables as compared to disyllables; speakers align L earlier in monosyllables irrespective of phrasal position. For example, phrase-initially, L aligns at % 31 of the vowel duration in /ma:/ while it aligns 44 % of the vowel duration in /ma:mi/. The same applies to closed syllables; for example, in phrase-noninitial position, L aligns at 28 % of the vowel duration in /mam/ at 44 % of the vowel duration in /mamzi/.

Table 6.2: Means alignment lags for L as a proportion of the vowel duration (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	0.27 (0.14)	0.44 (0.09)	0.21 (0.10)	0.33 (0.22)	0.31 (0.16)	0.38 (0.21)	0.40 (0.32)
		/ma:mi/	0.45 (0.29)	0.59 (0.17)	0.28 (0.06)	0.38 (0.23)	0.44 (0.23)		
	closed	/mam/	0.29 (0.17)	0.79 (0.17)	0.11 (0.10)	0.36 (0.15)	0.38 (0.29)	0.41 (0.40)	
		/mamzi/	0.23 (0.10)	0.92 (0.79)	0.26 (0.18)	0.41 (0.24)	0.44 (0.48)		
noninitial	open	/ma:/	0.15 (0.04)	0.56 (0.07)	0.08 (0.19)	0.26 (0.10)	0.26 (0.22)	0.30 (0.22)	0.33 (0.24)
		/ma:mi/	0.27 (0.13)	0.56 (0.23)	0.20 (0.23)	0.35 (0.07)	0.34 (0.21)		
	closed	/mam/	0.21 (0.05)	0.59 (0.16)	0.18 (0.11)	0.09 (0.36)	0.28 (0.27)	0.36 (0.25)	
		/mamzi/	0.32 (0.12)	0.67 (0.19)	0.32 (0.11)	0.48 (0.19)	0.44 (0.21)		

The exact timing of L as a proportion of the accented vowel duration seems, however, highly speaker-specific and not systematic. As can be seen in Figure 6.2, speakers S1, S3 and S4 generally align L in the first half of the accented vowel (as indicated by values between 0 and 0.5). In contrast, speaker S2 generally aligns L in the second half of the vowel (as indicated by values above 0.5). As for the effect of position and word length, speakers S1 and S3 only show subtle alignment differences, while speakers S2 and S4 produce somewhat clearer differences. More specifically, speaker S2 produces a clear alignment distinction between all target words in the phrase-initial condition while this distinction disappears in phrase-noninitial position. The opposite is true for S4, who does not differentiate between the target words in phrase-initial but does in phrase-noninitial position.

To sum up, there are conflicting results for the alignment of L relative to the acoustic onset of the accented vowel. In absolute terms, L was not affected by position or word length but it was found to align earlier in closed syllables than in open syllables. This difference disappears when taking the vowel duration into account. This proportional measure of alignment shows, however speaker-specific effects of position and word length: L aligns slightly earlier in noninitial position as compared to initial position and earlier in monosyllables than in disyllables.

A considerably clearer picture can be drawn from the evaluation of the beginning of the prenuclear rise relative to the syllable onset when it is considered in relation to the total syllable duration (Figure 6.3, data in Table 6.3). Here, zero denotes the acoustic onset of the syllable; one indicates the end of the syllable, which refers either to the vowel offset in open syllables or the end of the coda consonant in closed syllables. The rmANOVA revealed large effects of all factors (POSITION [$F(1,3) = 10.44$, $p<0.05$, $\eta^2=0.44$], SYLL [$F(1,3) = 497.66$, $p<0.001$, $\eta^2=0.87$] and of WORD LENGTH [$F(1,3) = 242.84$, $p<0.001$, $\eta^2=0.40$]).

In general, speakers align L earlier in phrase-initial position. On average, it aligns at 45 % of the syllable duration, while it aligns significantly later (at 51 % of the syllable duration) in phrase-noninitial position. Also, the prenuclear rise starts earlier in closed

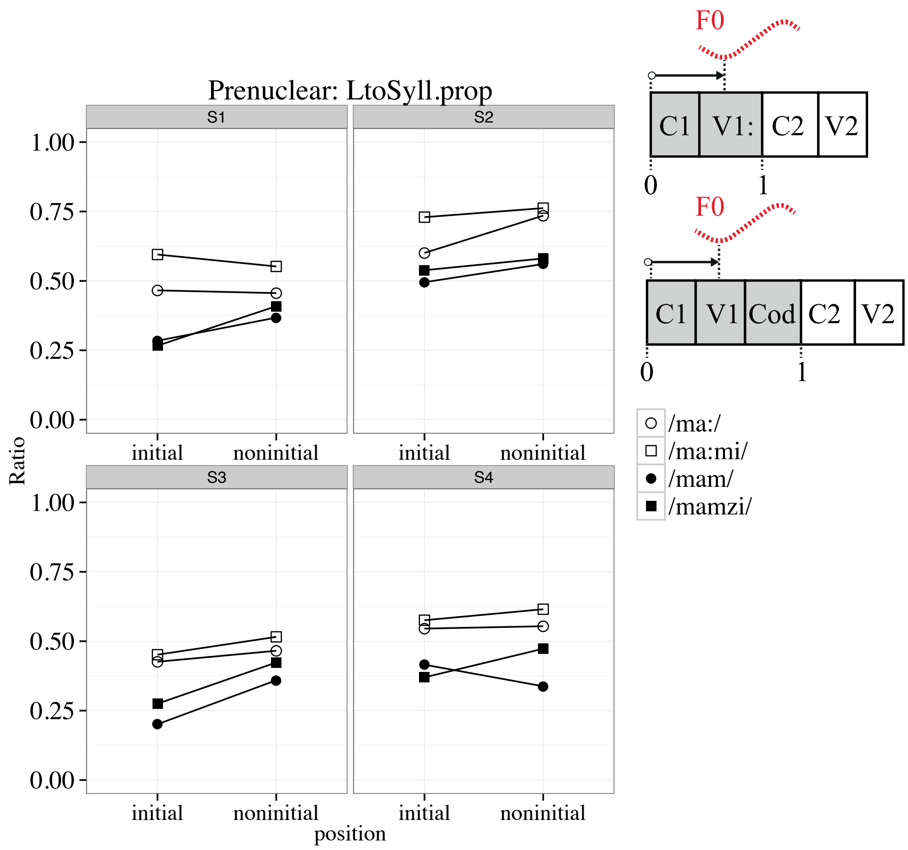


Figure 6.3: Mean alignment lags for L as a proportion of the syllable duration (LtoSyll.prop).

syllables than in open syllables. Phrase-initially, it starts at 55 % of the syllable duration in open and at 35 % of the syllable duration in closed syllables. Phrase-noninitially, the rise starts at 58 % of the syllable duration in open and (earlier) at 44 % of the syllable duration in closed syllables. Word length also played a significant role in determining the beginning of the accentual rise in that the rise starts earlier in monosyllables than in disyllables.

Furthermore, there are some speaker-specific patterns. For example, phrase-initially, speaker S3 aligns L roughly at the same time both in the monosyllabic /mam/ (28 % of syllable duration) and the disyllabic /mamzi/ (27 % of syllable duration). Another

Table 6.3: Mean alignment lags for L as a proportion of the syllable duration (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	0.47 (0.11)	0.60 (0.07)	0.43 (0.09)	0.55 (0.15)	0.51 (0.12)	0.55 (0.15)	0.45 (0.18)
		/ma:mi/	0.60 (0.22)	0.73 (0.10)	0.45 (0.06)	0.58 (0.14)	0.60 (0.17)		
	closed	/mam/	0.28 (0.06)	0.50 (0.06)	0.20 (0.03)	0.42 (0.07)	0.35 (0.13)	0.35 (0.16)	
		/mamzi/	0.27 (0.03)	0.54 (0.28)	0.27 (0.08)	0.37 (0.12)	0.36 (0.18)		
noninitial	open	/ma:/	0.46 (0.03)	0.74 (0.04)	0.47 (0.11)	0.55 (0.07)	0.55 (0.13)	0.58 (0.14)	0.51 (0.14)
		/ma:mi/	0.55 (0.09)	0.76 (0.13)	0.52 (0.15)	0.62 (0.05)	0.61 (0.14)		
	closed	/mam/	0.37 (0.02)	0.56 (0.05)	0.36 (0.03)	0.34 (0.14)	0.41 (0.11)	0.44 (0.10)	
		/mamzi/	0.41 (0.05)	0.58 (0.05)	0.42 (0.03)	0.47 (0.07)	0.47 (0.08)		

exception is speaker S4 who, phrase-initially, aligns L earlier in /mamzi/ (37 % of syllable duration) than its counterpart /mam/ (42 % of syllable duration).

Even though speakers align L later in open than in closed syllables in relation to the total syllable duration, there is evidence for a *rather* stable alignment with the accented vowel. More specifically, for three of the four speakers, it consistently aligns in the first half of the accented vowel.

6.1.2 Acoustic alignment of H

This section investigates the alignment of the prenuclear F0 peak relative to acoustic landmarks. This was measured from the acoustic offset of the syllable and as function of the total syllable duration.

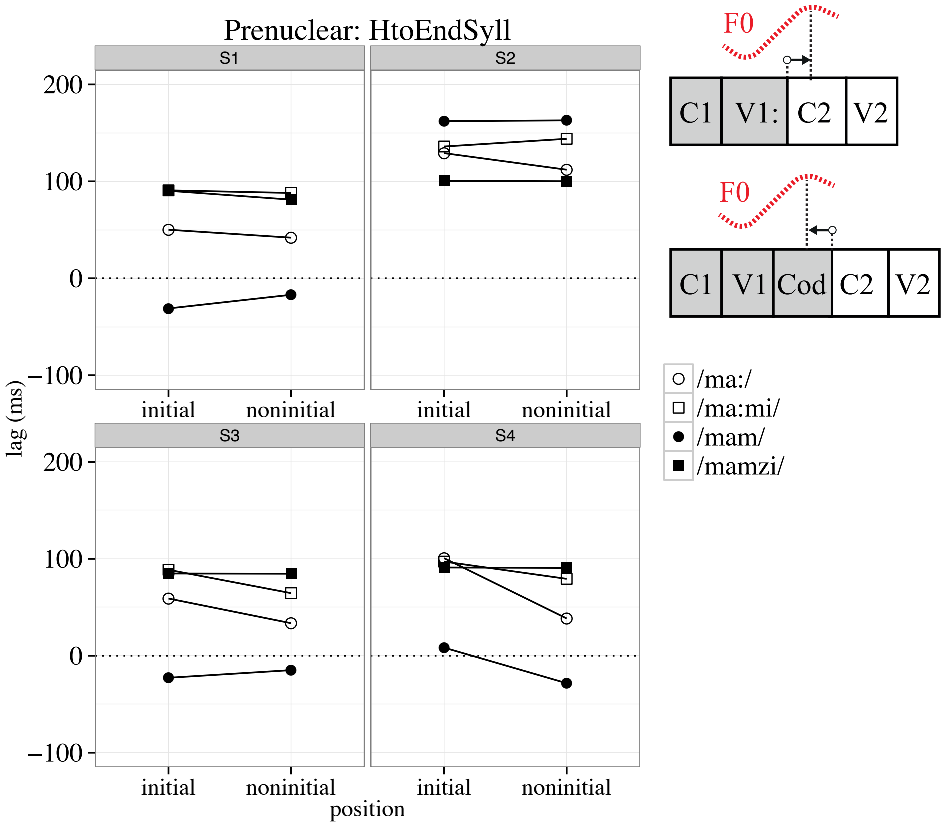


Figure 6.4: Mean alignment lags (in ms) for H relative to the syllable offset (HtoEndSyll).

Figure 6.4 show the F0 peak alignment relative to the end of the accented syllable. Data are given in Table 6.4. Zero denotes the syllable boundary, i.e. the vowel offset in target words with an open syllable or the end of the coda consonant in target words with a closed syllable.

With the exception of /mamzi/, all speakers align the prenuclear F0 peak *after* the

Table 6.4: Mean alignment lags (in ms) for H relative to acoustic syllable offset (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	50 (22)	129 (31)	59 (43)	101 (51)	85 (48)	95 (42)	77 (60)
		/ma:mi/	91 (21)	136 (23)	89 (17)	97 (42)	104 (33)		
	closed	/mam/	-31 (8)	162 (30)	-23 (50)	8 (31)	27 (85)	59 (69)	
		/mamzi/	90 (11)	101 (37)	85 (9)	91 (23)	91 (22)		
noninitial	open	/ma:/	42 (16)	112 (17)	34 (15)	38 (21)	57 (37)	75 (40)	67 (57)
		/ma:mi/	88 (10)	144 (34)	65 (24)	79 (14)	94 (36)		
	closed	/mam/	-17 (13)	163 (34)	-15 (40)	-28 (22)	28 (86)	59 (68)	
		/mamzi/	81 (15)	100 (24)	85 (6)	91 (13)	89 (17)		

syllable boundary. The rmANOVA showed no effect of POSITION [$F(1,3) = 2.81$, $p \geq 0.05$, $\eta^2=0.04$] nor of SYLL [$F(1,3) = 7.77$, $p \geq 0.05$, $\eta^2=0.17$] nor of WORD LENGTH [$F(1,3) = 4.06$, $p \geq 0.05$, $\eta^2=0.39$].[1]

On average, H aligns 77 ms after the syllable boundary in phrase-initial position and 67 ms after it in phrase-noninitial position. In this position, the F0 peak aligns on average 75 ms after the syllable boundary in open syllables as compared to 59 ms after it in closed syllables. The same pattern applies to the F0 peak alignment in phrase-initial position. Here, H aligns on average 95 ms after the syllable boundary in open syllables as compared to 59 ms in closed syllables. As none of the factors reached significance, there

[1]However, the statistical analysis did reveal a small three-way interaction between the three factors under investigation [$F(1,3) = 11.07$, $p<0.05$, $\eta^2<0.01$]. L aligns earlier in phrase-initial /mam/ than in phrase-nonitial /ma:mi/.

is only a (non-significant) trend for the F0 peak to align earlier in closed syllables.

Rather, the alignment of the prenuclear F0 peak appears to be speaker-specific. On the one hand, Speaker S2 aligns H consistently after the syllable boundary with alignment lags always greater than 100 ms. On the other hand, speakers S1, S3 and S4 align the F0 peak shortly before the syllable boundary in /mam/ but after the boundary, within a time span of 100 ms, in /ma:/, /ma:mi/ and /mamzi/. In other words, H aligns roughly with the end of the coda consonant in /mam/ while it aligns *after* the coda consonant in /mamzi/. In /ma:/ and /ma:mi/, these speakers align H after the vowel offset.

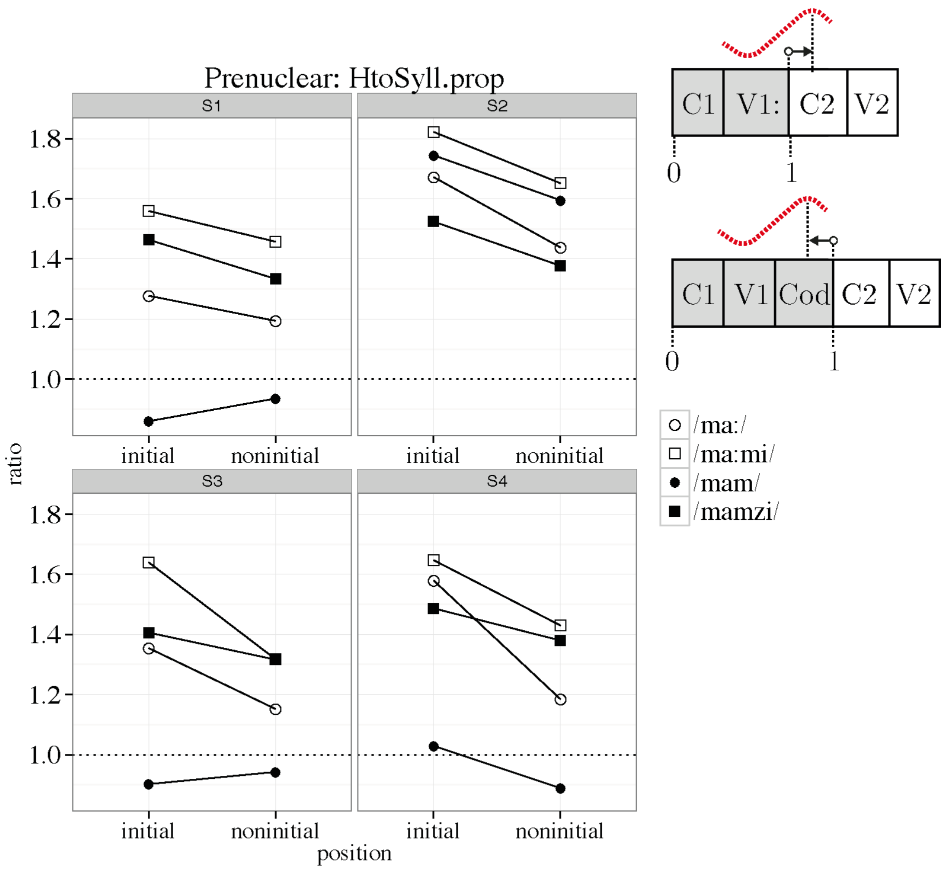


Figure 6.5: Mean alignment lags for H as a function of the syllable duration (HtoSyll.prop).

Figure 6.5 shows mean alignment lags of H as a function of syllable duration. Values below 1.0 indicate that it aligns before the syllable boundary, and values higher then 1.0 indicate that H aligns after the syllable boundary. Full data are given in Table 6.5.

Table 6.5: Mean alignment lags for H as a function of the syllable duration (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	1.28 (0.11)	1.67 (0.18)	1.35 (0.28)	1.58 (0.33)	1.47 (0.28)	1.57 (0.27)	1.43 (0.33)
		/ma:mi/	1.56 (0.14)	1.82 (0.11)	1.64 (0.10)	1.65 (0.37)	1.67 (0.22)		
	closed	/mam/	0.86 (0.03)	1.47 (0.14)	0.90 (0.23)	1.03 (0.13)	1.12 (0.39)	1.30 (0.33)	
		/mamzi/	1.46 (0.07)	1.52 (0.20)	1.41 (0.06)	1.49 (0.10)	1.47 (0.12)		
noninitial	open	/ma:/	1.19 (0.07)	1.44 (0.05)	1.15 (0.07)	1.18 (0.12)	1.24 (0.14)	1.35 (0.18)	1.29 (0.23)
		/ma:mi/	1.46 (0.07)	1.65 (0.15)	1.32 (0.12)	1.43 (0.12)	1.46 (0.15)		
	closed	/mam/	0.94 (0.05)	1.60 (0.12)	0.94 (0.14)	0.89 (0.09)	1.10 (0.32)	1.23 (0.26)	
		/mamzi/	1.33 (0.06)	1.38 (0.09)	1.32 (0.03)	1.38 (0.04)	1.35 (0.06)		

The rmANOVA revealed large effects of POSITION [$F(1,3) = 20.36$, $p<0.05$, $\eta^2=0.28$] and SYLL [$F(1,3) = 26.11$, $p<0.05$, $\eta^2=0.41$]. The factor WORD LENGTH did not reach significance [$F(1,3) = 7.59$, $p=0.07$, $\eta^2=0.53$]. In general, H aligns earlier in phrase-noninitial than in phrase-initial position, and it aligns earlier in closed than in open syllables. Speakers S1, S3 and S4 show roughly the same pattern: With the exception of /mam/, the F0 peak always aligns after the syllable boundary. Speaker S2, in contrast, aligns H after the syllable boundary in all target words.

For speakers S1 and S3, H aligns consistently earlier in the monosyllables /ma:/ and /mam/ than in the disyllables /ma:mi/ and /mamzi/. Speaker S2 and S4 show a different pattern. For speaker S2, H aligns later in the monosyllabic /mam/ than its disyllabic counterpart /mamzi/. Speaker S3 shows an interaction between position and syllable structure. Phrase-initially, H aligns later in /ma:mi/ than in /mamzi/. In phrase-noninitial position, however, it aligns earlier.

Summarising the data, the end of the prenuclear rise aligns (with the exception of /mamzi/ for three speakers) consistently after the syllable boundary. In relation to this landmark, the F0 is not affected by the factors under investigation. However, in relation to the syllable duration, the F0 peaks tend to earlier in closed than in open syllables. Moreover, the F0 peak tends to align earlier in phrase-noninitial than in phrase-initial position.

6.2 Alignment relative to articulatory landmarks

This section examines the alignment of the prenuclear rise relative to landmarks in the articulatory domain. As in the analysis of the nuclear rise, the beginning of the rise was related to both the articulatory target for the word-initial /m/ and the peak velocity of its release. The end of the rise was measured relative to the articulatory vowel target corresponding to the maximum tongue lowering for the accented vowel.

6.2.1 Articulatory alignment of L

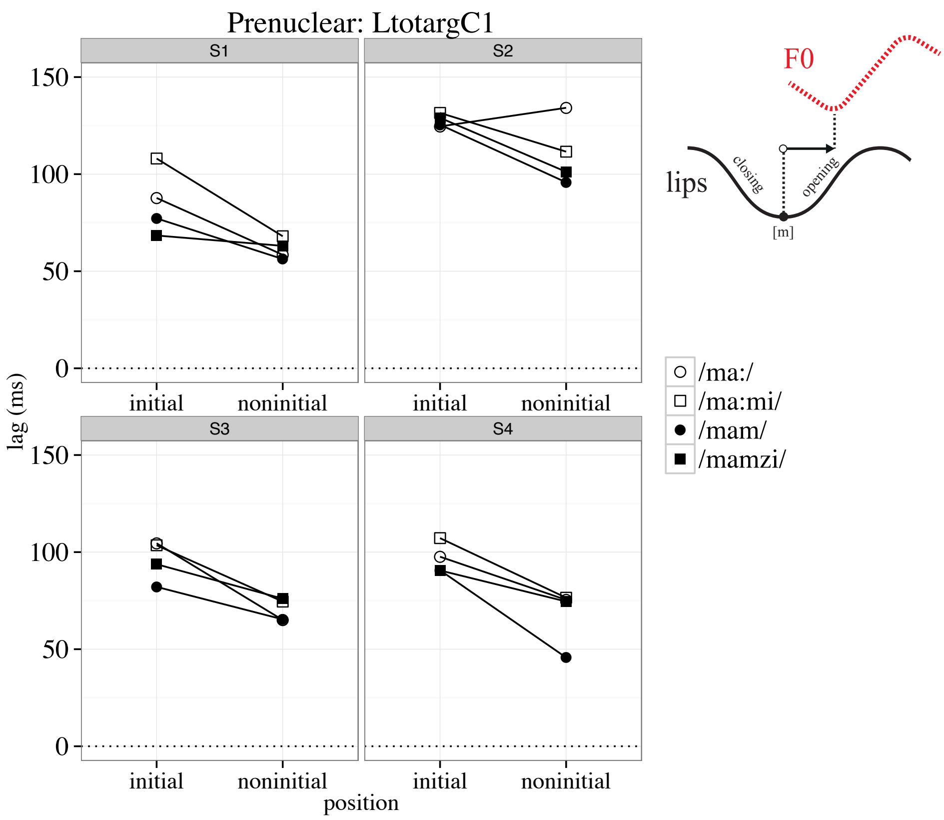


Figure 6.6: Mean alignment lags (in ms) for L relative to the maximum closure in /m/ (LtotargC1).

Figure 6.6 displays the alignment lags from the beginning of the rise relative to the articulatory target of the initial consonant, i.e. the maximum lip closure in /m/. Full Data are given in Table 6.6.

Table 6.6: Mean alignment lags (in ms) for L relative to the maximum closure in /m/ (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	88 (22)	125 (17)	105 (14)	98 (14)	104 (23)	109 (28)	101 (29)
		/ma:mi/	108 (43)	132 (22)	104 (7)	107 (34)	113 (31)		
	closed	/mam/	77 (16)	125 (11)	82 (9)	91 (15)	93 (23)	94 (29)	
		/mamzi/	68 (5)	129 (51)	94 (10)	91 (18)	95 (34)		
noninitial	open	/ma:/	58 5)	134 (15)	65 (22)	75 (12)	84 (34)	73 (22)	77 (27)
		/ma:mi/	68 (13)	112 (37)	75 (26)	77 (14)	82 (28)		
	closed	/mam/	56 (4)	96 (15)	65 (6)	46 (31)	66 (25)	83 (31)	
		/mamzi/	63 (10)	101 (12)	76 (9)	75 (15)	79 (18)		

All speakers align L considerably *after* the lip closure, i.e. the articulatory target for /m/. An overall rmANOVA revealed large effects of SYLL [$F(1,3) = 61.77$, $p<0.01$, $\eta^2=0.42$] and POSITION [$F(1,3) = 93.95$, $p<0.01$, $\eta^2=0.74$] with WORD LENGTH [$F(1,3) = 5.08$, $p\geq0.05$, $\eta^2=0.14$] having no significant effect. In general, speakers align L slightly earlier in phrase-noninitial position than in phrase-initial position. Phrase-initially, L aligns, on average, 101 ms after the maximum lip closure, while phrase-noninitially, it aligns, on average, only 77 ms after it.

The role of syllable structure expresses itself in different forms: Phrase-initially, L aligns slightly earlier in closed syllables (94 ms after the maximum lip closure) than in open syllables (109 ms after it). Phrase-noninitially, it aligns *earlier* in open syllables (83 ms after the maximum closure) than in closed ones syllables (73 ms after it).

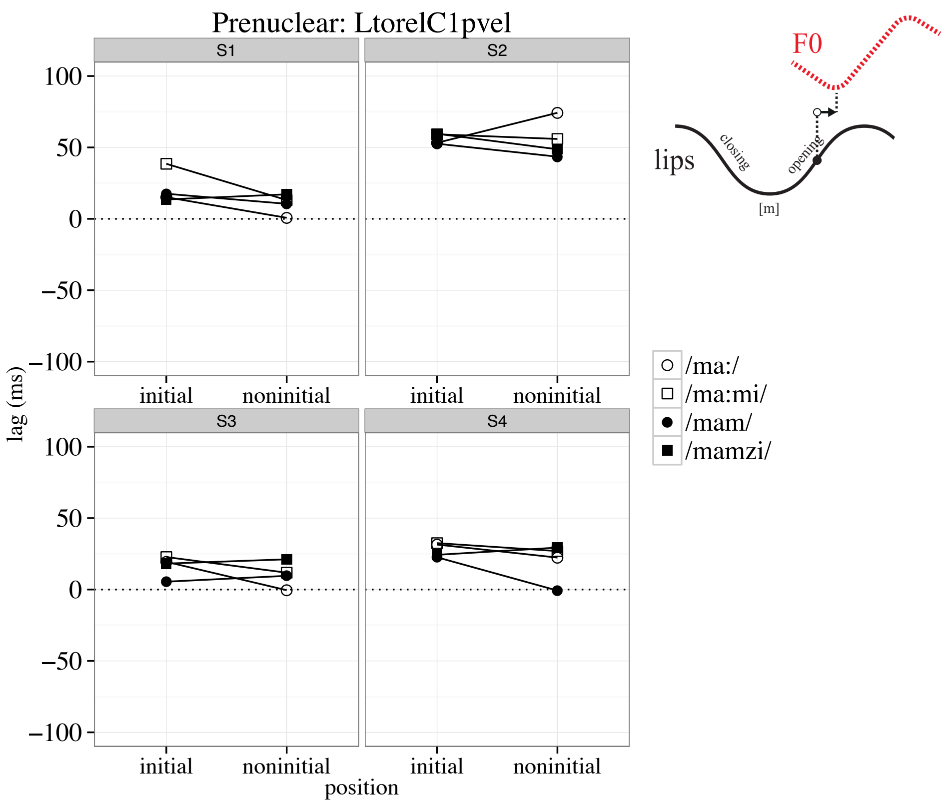


Figure 6.7: Mean alignment lags (in ms) for L relative to the peak velocity of the opening gesture from /m/ to /a/ (LtorelC1pvel).

These effects, however, disappear when investigating the onset of the accentual rise relative to the peak velocity of the consonantal release gesture, i.e. the point in time where the lip opening movement reaches its maximum velocity. Figure 6.7 displays the mean alignment of L relative to the peak velocity of the release of /m/. Data are given in Table 6.7.

The rmANOVA did not find any effects to be significant at all. (SYLL [F(1,3) = 4.37, p≥0.05, η^2=0.13]; POSITION [F(1,3) = 8.53, p≥0.05, η^2=0.18]; WORD LENGTH [F(1,3) = 8.86, p≥0.05, η^2=0.22]). On average, speakers align L, on average, 30 ms after the peak velocity of the consonant's release gesture in phrase-initial position. In phrase-nonitial position, speakers align it, on average, 34 ms after this landmark.

Table 6.7: Mean alignment lags (in ms) for L relative to the peak velocity of the release in /m/ (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	15 (17)	53 (12)	19 (13)	31 (27)	30 (23)	35 (27)	30 (27)
		/ma:mi/	39 (41)	59 (22)	23 (6)	32 (27)	40 (30)		
	closed	/mam/	18 (13)	53 (11)	5 (8)	23 (10)	24 (20)	26 (27)	
		/mamzi/	14 (7)	60 (53)	18 (13)	24 (15)	28 (22)		
noninitial	open	/ma:/	1 (8)	74 (12)	0 (22)	22 (12)	24 (35)	25 (31)	24 (26)
		/ma:mi/	13 (15)	56 (30)	12 (28)	27 (10)	26 (27)		
	closed	/mam/	11 (4)	44 (12)	10 (9)	-1 (31)	16 (23)	23 (21)	
		/mamzi/	17 (8)	49 (12)	21 (10)	29 (17)	29 (17)		

Overall, speakers show a consistent alignment pattern in that none of the speakers show different alignment strategies. With the exception of speaker S2, all speakers align L within a time interval of, on average, 30 ms after the peak velocity of the release gesture. Speaker S2, in contrast, aligns it later, between 44 ms and 74 ms after this landmark.

Summarising the results for the alignment of L relative to landmarks in the articulatory domain, there is evidence of a rather stable alignment for L shortly after peak velocity of the release gesture. None of the investigated factors had an effect on the alignment with respect to the peak velocity, indicating this landmark to be an appropriate anchor point.

6.2.2 Articulatory alignment of H

Figure 6.8 shows the mean alignment lags for the prenuclear peak relative to the articulatory target for /a/, i.e. the maximum tongue body lowering during the vocalic opening. Data are given in Table 6.8.

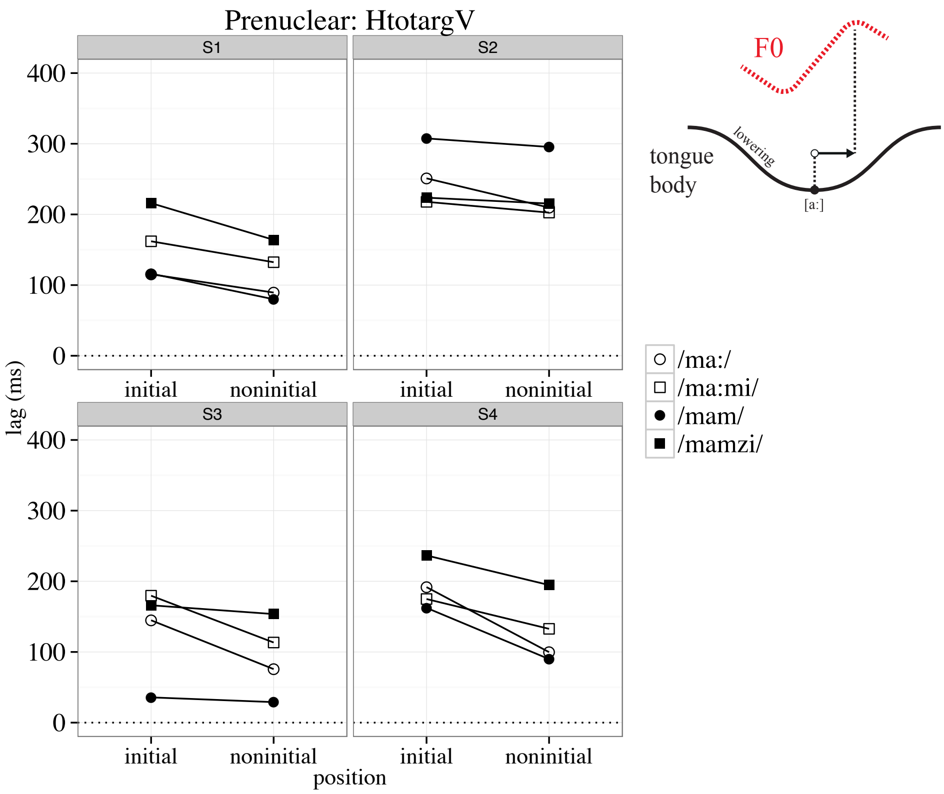


Figure 6.8: Mean alignment lags (in ms) for H relative to the articulatory vowel target in /a/ (HtotargV).

Similar to the results for the nuclear peak, the rmANOVA **did not** reveal effects of SYLL ([F(1,3) = 0.61, p≥0.05], η^2=0.03) or WORD LENGTH ([F(1,3) = 1.51, p<0.05], η^2=0.22). Instead, it showed an effect of POSITION ([F(1,3) = 19.50, p<0.05], η^2=0.22) and a three-way interaction between WORD LENGTH, (SYLL and POSITION ([F(1,3) =

23.57, $p<0.05$], $\eta^2=0.00$).[2]

Table 6.8: Mean alignment lags (in ms) for H relative to the articulatory vowel target in /a/ (standard deviation in parentheses).

position	syllable structure	target word	S1	S2	S3	S4	mean	mean	mean
initial	open	/ma:/	115 (27)	251 (27)	145 (61)	192 (42)	176 (65)	180 (52)	181 (68)
		/ma:mi/	162 (27)	218 (23)	180 (37)	175 (32)	184 (36)		
	closed	/mam/	116 (9)	308 (28)	36 (41)	162 (23)	154 (102)	181 (82)	
		/mamzi/	216 (10)	224 (36)	166 (30)	237 (44)	208 (40)		
noninitial	open	/ma:/	89 (18)	210 (24)	76 (14)	100 (17)	119 (58)	132 (51)	143 (70)
		/ma:mi/	132 (11)	203 (36)	113 (26)	133 (14)	144 (40)		
	closed	/mam/	80 (18)	295 (35)	29 (43)	90 (8)	127 (108)	154 (83)	
		/mamzi/	164 (29)	215 (31)	154 (14)	195 (14)	180 (35)		

In general, H aligns earlier in phrase-noninitial position than in phrase-initial position. Phrase-initially, H aligns 181 ms after the articulatory vowel target and phrase-noninitially, 143 ms after it. Even though the rmANOVA did not find any effect of word length, there was a trend for the F0 peak to align earlier in monosyllables than in disyllables. For example, phrase-noninitially, speakers align H, on average, 119 ms after the vowel target in /ma:/ while they align it, on average, 144 ms after it in /ma:mi/. The same pattern applies to target words with a closed syllable. In /mam/, H aligns 127 ms after the vowel target while in /mamzi/, it aligns 180 ms after it.

It is worth noting that once again, syllable structure *did not* play a role in determining

[2] As the effect size was too small, a post-hoc test could not detect any statistically significant difference between the conditions.

F0 peak alignment. Phrase-initially, the F0 peak aligns, on average, 180 ms after the articulatory vowel target in open syllables and 181 ms after it in closed syllables. Phrase-noninitially, it aligns 132 ms after the vowel target in open syllables and 154 ms after it in closed syllables.

The next section summarises the findings for the alignment of the prenuclear pitch accents relative to landmarks in both the acoustic and articulatory domain.

6.3 Summary (acoustics and articulation)

To sum up the results, there are landmark-specific effects on the alignment of L. All investigated alignment lags were either affected by phrasal position, syllable structure or word length. Table 6.9 sums up the statistical analyses for the acoustic alignment of L. A significant effect is indicated by the effect size η^2 (n.s. = not significant).

In general, L aligns earlier in closed than in open syllables and earlier in monosyllables than in disyllables. The positional effect depends on the investigated alignment lag. Within the vowel, L aligns a little earlier in noninitial position. Across the entire syllabe, L aligns *later* in that position. Despite these effects, the overall impression is that L aligns well within the accented vowel. For speakers S1, S2 and S4, L aligns well within the first half of the vowel while speaker S2 shows a greater degree of variability in her alignment.

Table 6.9: Overview of the statistical analyses for the acoustic alignment of L.

	LtoV1ons	LtoV1ons.prop	LtoSyll
Position	n.s.	$\eta^2 = 0.15$	$\eta^2 = 0.44$
Syllable structure	$\eta^2 = 0.46$	n.s.	$\eta^2 = 0.87$
Word length	n.s.	$\eta^2 = 0.34$	$\eta^2 = 0.40$

Figure 6.9 illustrates the findings for the acoustic alignment of L using data from speaker S1. The red arrows indicate the absolute alignment of L relative to the acoustic onset of the target word and the onset of the accented vowel. The values in parentheses indicate the proportional alignment of L relative to the total syllable duration.

The absolute alignment of L relative to the acoustic vowel onset is only affected by the syllable structure of the target word. L aligns slightly later in open syllables than in

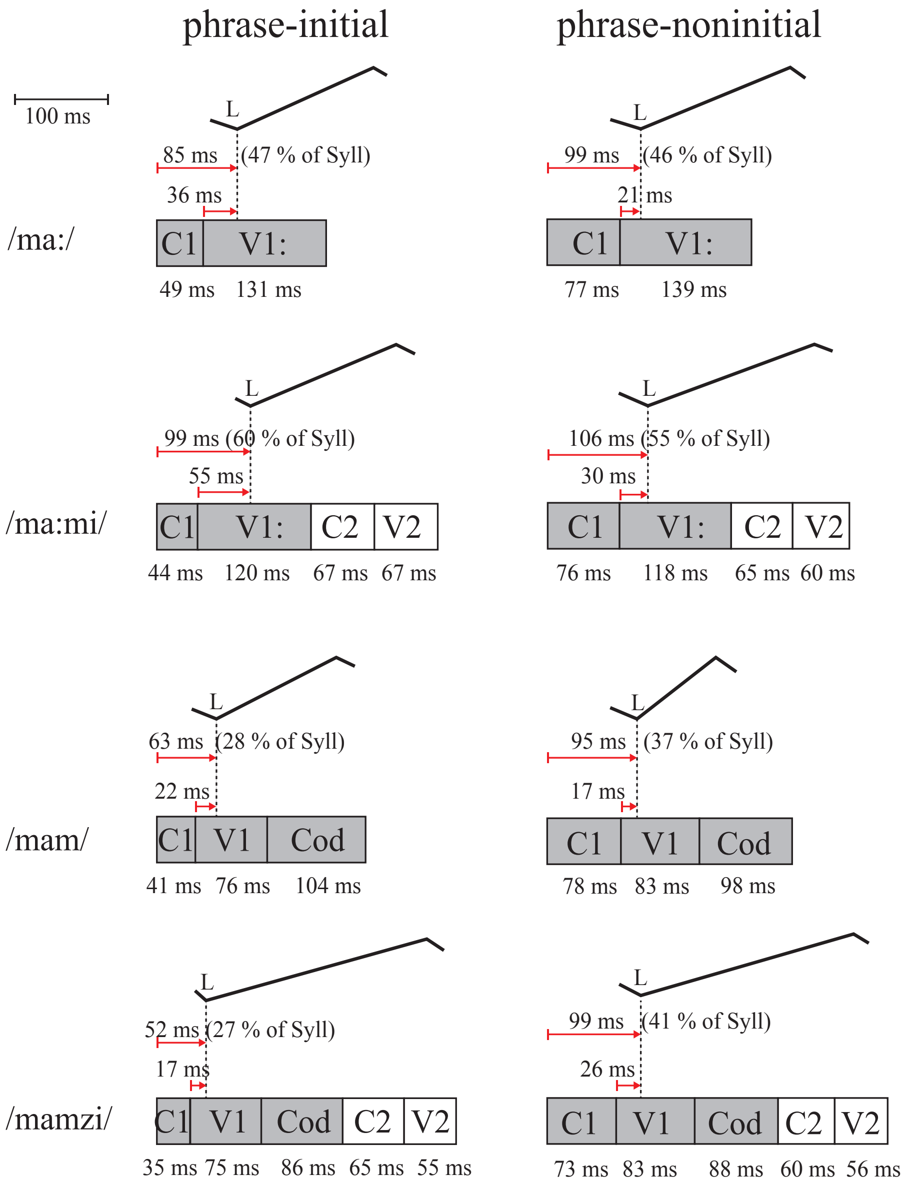


Figure 6.9: Acoust alignment of L (data from S1).

closed syllables. Phrase-initially, L aligns 36 ms after the vowel onset in /ma:/ and 22 ms after it in /mam/. In /ma:mi/, L aligns 55 ms after the vowel onset, while it aligns 17 ms after it in /mamzi/. Phrase-noninitially, L aligns 21 ms after the vowel onset in /ma:/ and 17 ms after it in /mam/. In /ma:mi/, L aligns 30 ms after it while it aligns

26 ms after it in /mamzi/.

The alignment of L in relation to the entire syllable duration was affected by all investigated prosodic factors. L aligns earlier in phrase-initial than in phrase-noninitial position and earlier in closed than in open syllables. It is worth noting, however, that these effects seem to be highly speaker and even target word specific. For speaker S1 (see Figure 6.9), an earlier alignment in phrase-initial position can ony be observed for closed syllables. Specifically, in /mam/, L aligns at 37 % of the syllable duration in phrase-noninitial position and earlier (at 28 % of the syllable duration) in phrase-initial position. The same pattern applies to /mamzi/ where this speaker aligns L at 41 % and 27 % of the syllable duration in phrase-noninitial and phrase-initial position, respectively. Also, an earlier alignment of L in monosyllables as compared to disyllables can only be found in target words with an open syllable: In the disyllabic word /ma:mi/ L aligns at 60 % of the syllable duration while in /ma:/ it aligns earlier at 47 % of the syllable. There is no such difference between the closed syllables /mam/ and /mamzi/ in phrase-initial position. Here, L aligns at 28 % and 27 % of total syllable duration, respectively.

However, given these rather small differences between the conditions, the overall impression is that there is a fairly fixed alignment for the onset of the prenuclear rise. In all investigated target words, L aligns within the accented vowel. Three of the four speakers align L well within the first half of the accented vowel.

The alignment of H, in contrast, shows a different picture. Table 6.10 summarizes the findings for the acoustic alignment of H.[3]

[3]The three-way interaction between position, syllable structure and word length was excluded from this table as the effect size was lower than 0.01.

Table 6.10: Overview of the statistical analyses for the acoustic alignment of H.

	HtoEndSyll	HtoSyll
Position	n.s.	$\eta^2 = 0.28$
Syllable structure	n.s.	$\eta^2 = 0.41$
Word length	n.s.	n.s.

The absolute alignment of H relative to the syllable boundary was not affected by the investigated prosodic factors. However, speakers S1, S3 and S4 do make a clear distinction between the peak alignment in the monosyllabic /mam/ and its disyllabic counterpart /mamzi/. In /mam/, H aligns well within the accented syllable (in the coda consonant /m/), whereas in /mamzi/, H aligns *after* the syllable boundary (in the following unstressed syllable). Figure 6.10 illustrates this effect with data from speaker S1.

The overall impression is that there is no F0 peak alignment difference between target words in phrase-initial and phrase-noninitial position. In both positions, however, there are three different alignment patterns for H. In the monosyllabic /ma:/, the F0 peak crosses the word boundary in that it aligns within the next word. In the disyllables /ma:mi/ and /mamzi/, the peak aligns well within the postaccented vowel V2. Finally, in /mam/, the F0 peak aligns within the coda consonant.

In sum, the onset of the prenuclear rise is only marginally affected by the investigated factors. It is rather stably aligned with the accented vowel. The accentual peak, on the other hand, is sensitive to both syllable structure and word length. In all target words, it occurs *after* the accented vowel. In monosyllables, it aligns either with the onset of the next word (as in the case of /ma:/) or within the coda consonant (as in /mam/). In disyllables, it aligns well within the postaccented vowel (as in /ma:mi/ and /mamzi/).

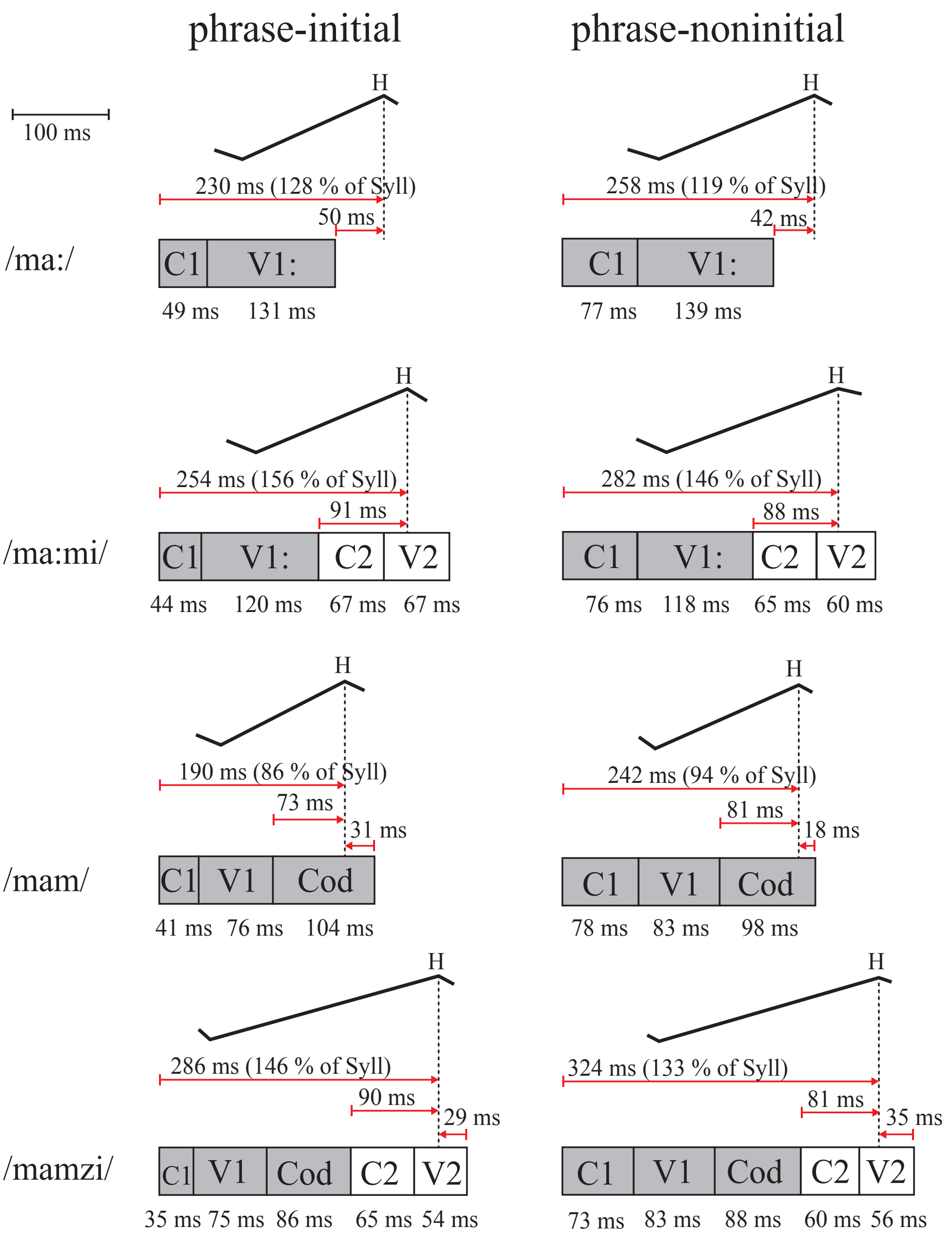


Figure 6.10: Acoust alignment of H (data from S1).

Table 6.11 sums up the statistical analyses for the articulatory alignment of L. With respect to the maximum lip closure, the onset aligns later in phrase-initial than in phrase-noninitial position. In phrase-initial position, the onset was aligned slightly later in open as compared to closed syllables. In phrase-noninitial position, in contrast, the onset was aligned slightly earlier in open as compared to closed syllables. These effects, however, disappear when relating the onset of the prenuclear rise to the peak velocity of the consonant's release gesture. This lag was consistent across all investigated prosodic factors. In other words, the onset of the prenuclear rise was not affected by phrasal position, syllable structure or by word length.

Table 6.11: Overview of the statistical analyses for the alignment of L.

	LtotargC1	LtorelC1pvel
Position	$\eta^2 = 0.74$	n.s.
Syllable structure	$\eta^2 = 0.42$	n.s.
Word length	n.s.	n.s.

Figure 6.11 illustrates the findings for L using data from speaker S1 for all four target words in phrase-initial and phrase-noninitial position. The trajectory represents the lip aperture, i.e. the closing and opening movement for the word-initial /m/. The black dot in the release gesture represents the occurence of the peak velocity of the movement. The arrows indicate the interval between the peak velocity and the beginning of the prenuclear rise. In all target words, L aligns shortly after the peak velocity of the release. It aligns between 1 ms (/ma:/ phrase-noninitially) and 39 ms (/ma:mi/, phrase-initially) after the release.

Table 6.12 sums up the statistical analyses for the articulatory alignment of H. Like for the nuclear F0 peaks, H was significantly affected by phrasal position only. In general, the F0 peak aligns earlier in phrase-noninitial position than in phrase-initial position. Furthermore, the statistical analysis revealed a three-way interaction between

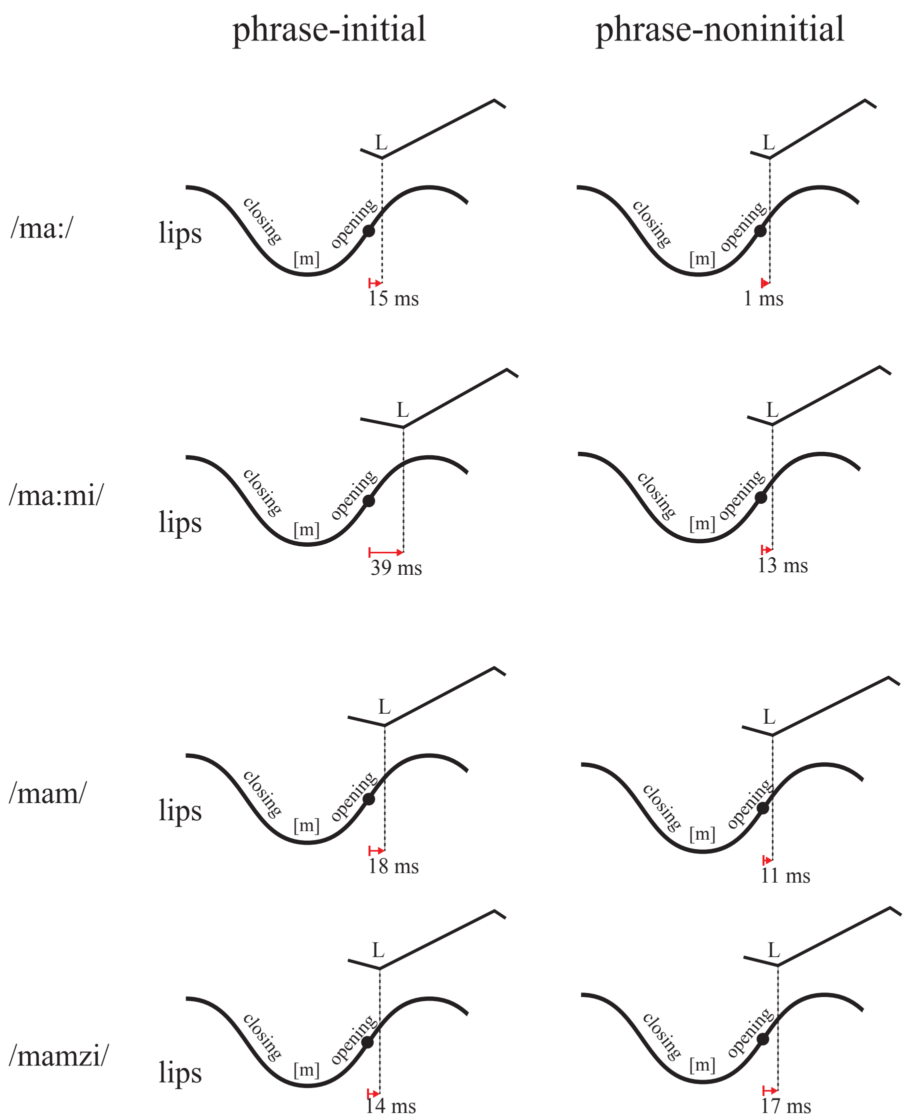


Figure 6.11: Articulatory alignment of L (data from speaker S1).

all investigated factors.[4] There was no evidence of an effect of syllable structure or word length; the alignment of the prenuclear F0 peaks did not change between target words with open or closed syllables or between mono- or disyllabic target words.

Figure 6.12 illustrates the findings for H using data from speaker S1. In this figure,

[4]The effect size η^2, however, was 0.00.

Table 6.12: Overview of the statistical analysis for the articulatory alignment of H.

	HtotargV
Position	$\eta^2 = 0.52$
Syllable structure	n.s.
Word length	n.s.
Position x Syllable structure x Word length	$\eta^2 = 0.00$

the end of the prenuclear rise was measured relative to the articulatory target of the vowel gesture, i.e. the maximum tongue body lowering in the vertical dimension for the production of the accented vowel (/a/ or /a:/). Arrows indicate the time interval between the articulatory vowel target and the F0 peak.

In all target words, H aligns earlier in phrase-noninitial position. For example, phrase-noninitially, it aligns 164 ms after the vowel target in /mamzi/ while phrase-initially, it aligns 216 ms after it. Importantly, the rmANOVA revealed no effect of syllable structure or word length.

Like in the nuclear accents, speakers show a stable yet speaker-specific alignment for both the beginning and the end of the prenuclear rise relative to articulatory landmarks in prenuclear accents. L is stably aligned shortly after the peak velocity of the consonantal release gesture while H displays a stable alignment with the articulatory vowel target which is only affected by phrasal position.

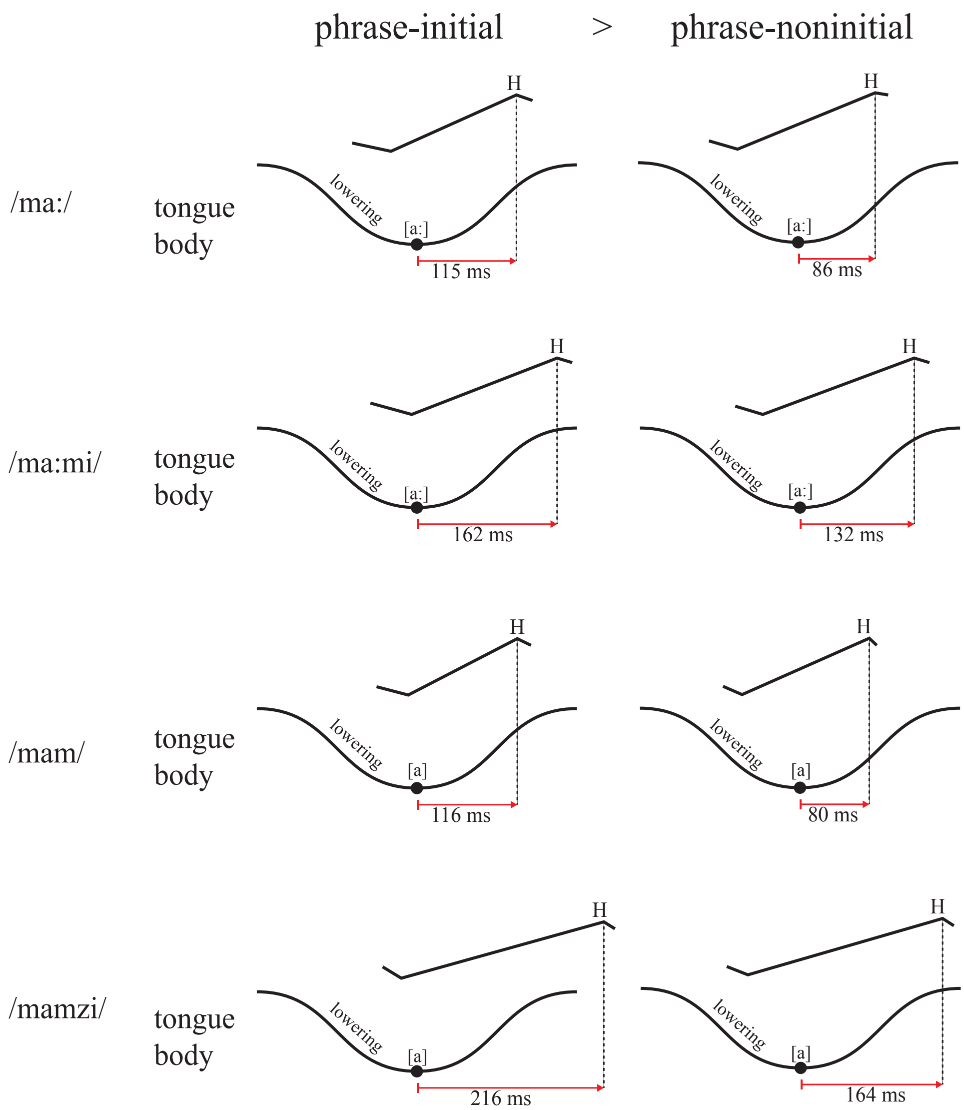


Figure 6.12: Articulatory Alignment of H (data from speaker S1).

7 Modelling nuclear and prenuclear accents

This chapter compares the coordination of nuclear and prenuclear rising pitch accents relative to landmarks in the articulatory domain. As pointed out in section 5.2 and section 7.2, the onset of the high tone gesture, L, is consistently aligned with an articulatory landmark, namely the point in time where the release gesture of the word-initial consonant /m/ reaches its peak velocity. In contrast, the target of the high tone gesture, H, shows a stable coordination pattern with the articulatory target of the accented vowel (the maximum tongue body lowering for the production of the vowel /a/ or /a:/). In what follows, the coordination of the high tone gesture will be modeled in the articulatory domain by using both a π-gesture and a μ-gesture in order to account for the attested effects of phrasal position and word length.

7.1 Onset of high tone gesture (L)

Figure 7.1 displays the coordination of the onset of the high tone gesture, L, relative to the peak velocity of the opening gesture for C1 pooled across target words. Each facet displays data from one speaker. Filled circles represent the nuclear data while empty circles represent the prenuclear data. Zero denotes the time point at which the peak velocity of the consonant's release gesture occurs (dotted line). Full data are given in Table 7.1.

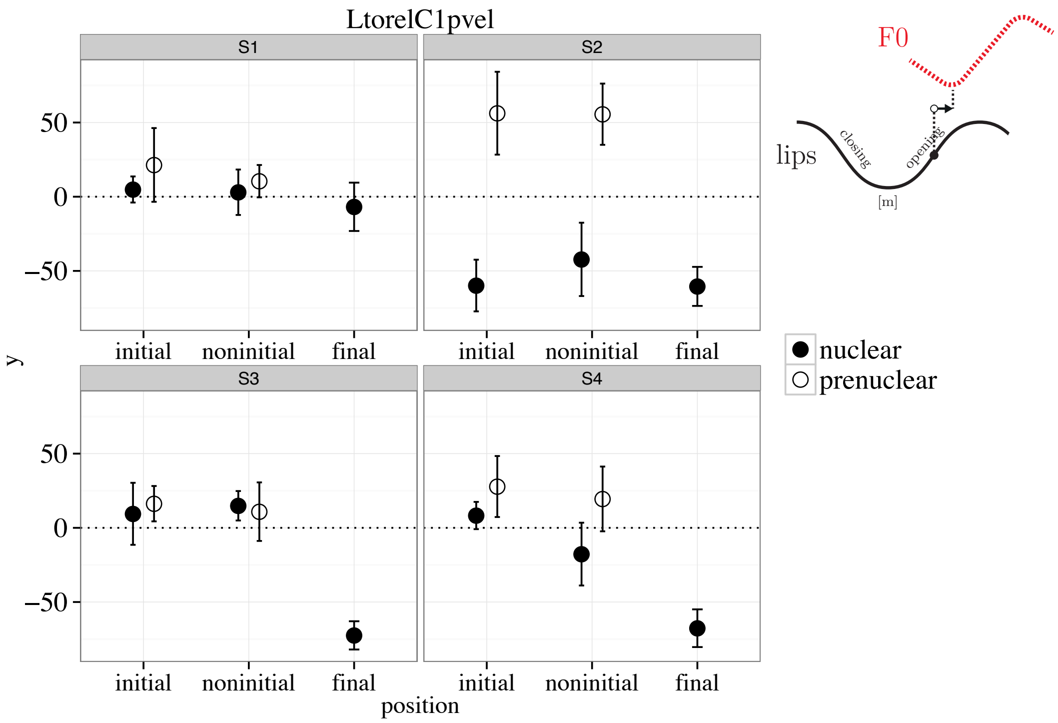


Figure 7.1: Mean alignment lags (in ms) for L relative to the peak velocity of the consonant's release gesture C1.

As data are unbalanced (only nuclear accents are produced in phrase-final position), a generalized linear mixed effects model was performed to test for the effects of position and accent on this alignment lag. Speaker was included as random factor, and by speaker random slopes were included for the factors accent and position. The model was tested against models without the factors accent and position as well as against a null model without any fixed effects to determine significant effects using likelihood ratio tests.

Table 7.1: Mean alignment lags (in ms) for the onset of the high tone gesture relative to peak velocity of the consonant's release gesture.

accent	position	S1	S2	S3	S4	mean	mean
nuclear	initial	5 (9)	-60 (17)	9 (21)	8 (9)	-11 (33)	-22 (35)
	noninitial	3 (15)	-42 (25)	15 (10)	-18 (21)	-11 (28)	
	final	-7 (16)	-60 (13)	-72 (10)	-68 (13)	-48 (32)	
prenuclear	initial	21 (25)	56 (28)	16 (20)	28 (21)	30 (27)	27 (27)
	noninitial	10 (11)	56 (21)	11 (20)	19 (22)	24 (26)	

The comparison of the full model against the null model revealed no significant difference ($\chi^2(4) = 8.85, p > 0.05$). Statistically, neither accent nor position had an effect on the onset of the high tone gesture relative to the peak velocity of the opening gesture for C1. In both nuclear and prenuclear accents the onset of the tone gesture is stably coordinated with the peak velocity. However, individual speakers show a high degree of variability.

While for speakers S1 and S3 the onset of the nuclear and the prenuclear tone gestures start almost at the same time in phrase-initial and phrase-noninitial position, for speakers S2 and S4, the prenuclear tone gesture starts later than the nuclear one. For example, for speaker S1, the prenuclear tone gesture starts, on average, 5 ms after the release's peak velocity in phrase-initial position. The *nuclear* tone gesture, starts only 16 ms later, namely 21 ms after the release's peak velocity. In contrast, speaker S2 shows the greatest difference between the nuclear and prenuclear data. For this speaker, the prenuclear tone gesture starts, on average, 60 ms *before* the peak velocity in phrase-initial position. The nuclear tone gesture is initiated, on average, 56 ms after this articulatory landmark. Another striking difference is in the coordination of the nuclear tone gesture in phrase-final position. Speakers S2, S3 and S4 start the tone gesture considerably before the peak velocity of the release, while speaker S1 starts the tone gesture almost simultaneously with it.

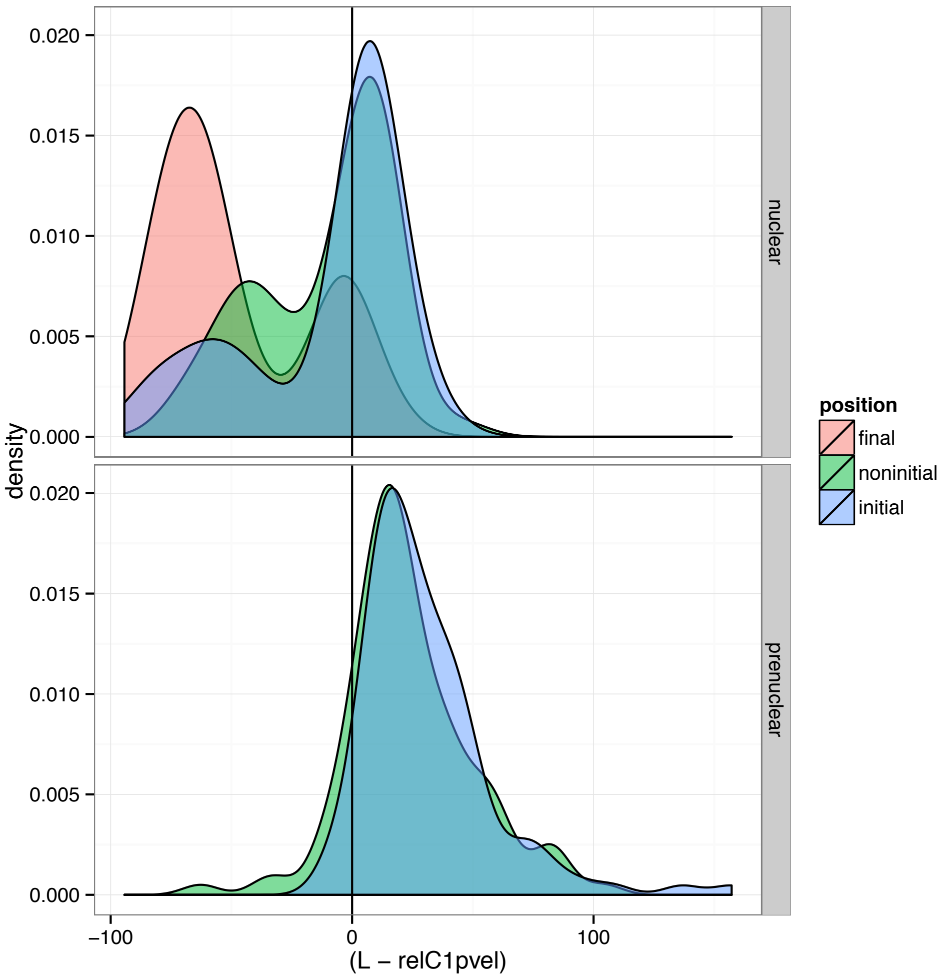


Figure 7.2: Density curve (averaged across all speakers and target words) for the coordination of the onset of the tone gesture relative to the peak velocity of the consonant's release gesture (upper panel: nuclear data, lower panel: prenuclear data).

In order to better understand the (speaker) variability, kernel density curves were calculated. Kernel density curves estimate the underlying distribution of the data by means of their frequency of occurence of different values of the investigated variable. Figure 7.2 presents the coordination of the onset of the tone gesture, L, relative to the peak velocity of the consonant's release gesture showing the coordination of the onset of the high tone gesture relative to the peak velocity of the consonant's release gesture. The upper panel shows the nuclear data, the lower panel shows the prenuclear data. The solid line at 0 ms marks the occurrence of the peak velocity of consonant's release gesture. Phrasal position is coded for by different colors with red showing target words in phrase-final position (only nuclear), while green shows target words in phrase-noninitial position, and blue shows target words in phrase-initial position.

The prenuclear data (lower panel) consistently show a unimodal and almost symmetrical distribution, with a clear peak occuring shortly after the peak velocity of the opening gesture. Data from noninitial (green) and initial position (blue) almost completely overlap indicating that speakers coordinate the onset of the high tone gesture, L, of the prenuclear rise stably in relation to the peak velocity, irrespective of phrasal position, word length or syllable structure. In contrast, the nuclear data (upper panel) show a bimodal distribution with two density peaks. Data from phrase-initial (blue) and phrase-noninitial (green) position overlap to a large degree, and their peaks roughly occur with the peak velocity of the consonant's release gesture. However, both show a second distributional peak *before* this alignment point indicating that in some cases the tone gesture starts before this landmark. The data from phrase-final position (red) show an even earlier and higher initial peak, indicating that in most phrase-final target words the high tone gesture starts considerably before the peak velocity of the consonant's release gesture. The question arises as to which factor contributes to this bimodal distribution in nuclear accents. To determine this, data were separated by both speaker and accent as shown in the density curves in Figure 7.3.

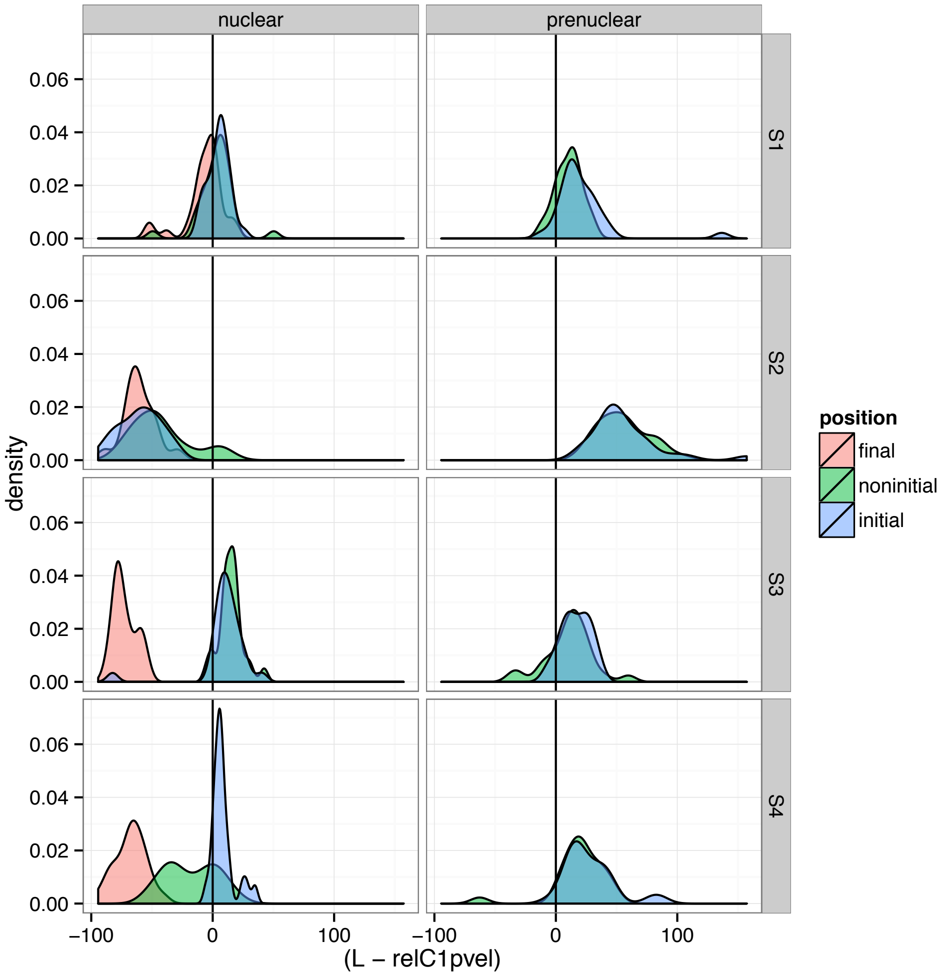


Figure 7.3: Density curve (averaged across target words) for the coordination of the onset of the tone gesture relative to the peak velocity of the consonant's release gesture.

The left panel shows the nuclear data, while the right panel shows the prenuclear data. Each row displays data from one speaker. As expected, the density curves for the prenuclear data overlap completely for all speakers. Looking at the nuclear accents, it becomes clear that speakers employ different strategies for the coordination of the onset of the high tone gesture. Specifically, for speakers S1 and S2, data from all phrasal positions overlap to a high degree indicating that these speakers do not change the onset the of the high tone gesture, L, as a function of phrasal position. The crucial difference between these speakers, however, is that speaker S1 consistently starts the high tone gesture with the peak velocity of the consonant's release, while speaker S2 starts the high tone gesture earlier (at around 50 ms before the peak velocity of the consonant's release). In contrast, data from speakers S3 and S4 show a different pattern. For speaker S3, density curves from phrase-initial and phrase-noninitial position overlap and show a peak shortly after the peak velocity of the consonant's release. For this speaker, however, the tone gesture is initiated earlier in phrase-final position. An earlier start of the tone gesture in phrase-final position can be observed for speaker S4, too. Data from this speaker additionally show a somewhat distinct (and wider) density curve in phrase-noninitial position indicating that in this position the high tone gesture is initiated intermediate between the onset of the high tone gesture, L, in phrase-final and phrase-initial position. In sum, the bimodal distribution for the (nuclear) density curves in Figure 7.2 result from speaker-specific strategies. While half of the speakers do not adjust the onset of the high tone gesture as a function of phrasal position, the other half does. Specifically, for two speakers, the high tone gesture starts earlier in phrase-final position.

7.2 Target of high tone gesture (H)

Figure 7.4 shows the coordination for the target of the high tone gesture, H, relative to the articulatory vowel target, i.e. the maximum tongue lowering for the accented vowel (/a/ or /a:/), by speaker and phrasal position. Zero marks the occurrence of the articulatory vowel target. Positive values indicate that the target of the high tone gesture, H, occurs after the articulatory vowel target. Full data are given in Table 7.2.

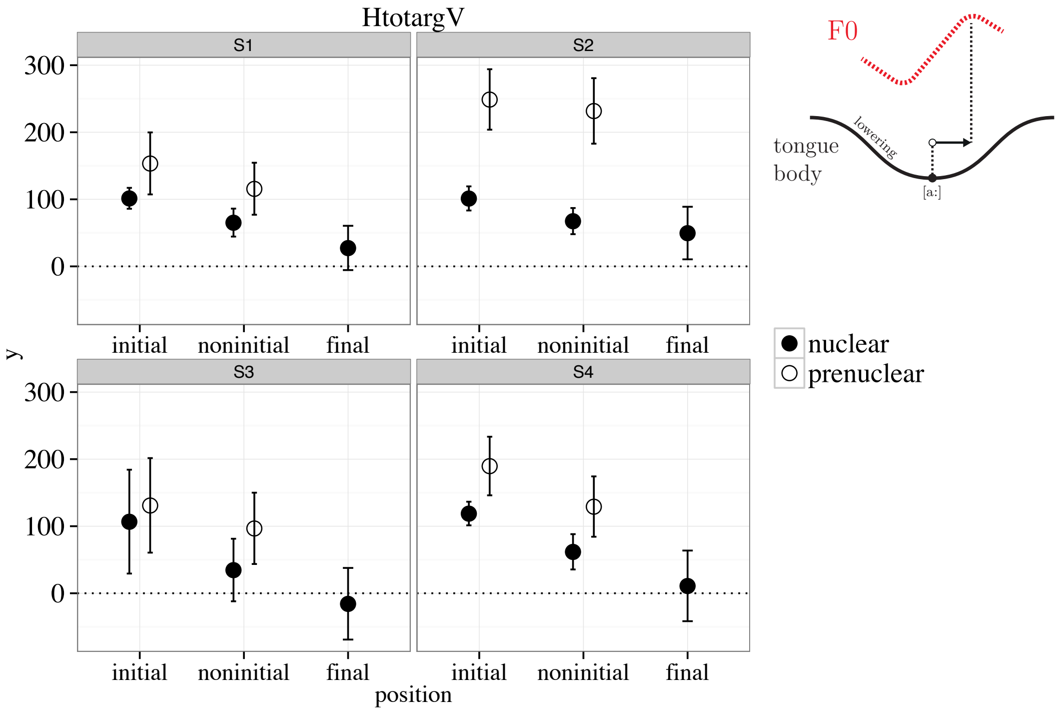


Figure 7.4: Mean alignment lags (in ms) for the target of the high tone gesture relative to the articulatory vowel target.

A generalized linear mixed effects model was performed to test for the effects of position and accent on the coordination of the accentual F0 peak with the articulatory vowel target. The model had the same structure as the one used to test for the effects on the coordination of the tone gesture onset. A comparison of the model including position and accent as fixed effects (full model) and the model without any fixed effects (null model) showed a significant difference ($\chi^2(4) = 18.12, p < 0.01$). Post-hoc testing confirmed that all but one pairwise comparisons were significant. The only non-significant difference found was between prenuclear accents in noninitial position and nuclear accents in initial position ($\beta = 36.6, SE = 26.8, z = 1.37, p \geq 0.05$).

The nuclear tone gesture reaches its target, on average, 62 ms after the articulatory

Table 7.2: Mean alignment lags (in ms) for the target of the high tone gesture relative to the articulatory vowel target.

accent	position	S1	S2	S3	S4	mean	mean
nuclear	initial	102 (16)	101 (18)	107 (78)	119 (18)	107 (41)	62 (54)
	noninitial	65 (21)	68 (20)	35 (47)	62 (26)	58 (32)	
	final	27 (33)	50 (39)	-16 (53)	11 (53)	16 (50)	
prenuclear	initial	154 (46)	249 (45)	131 (70)	190 (43)	181 (68)	162 (72)
	noninitial	116 (39)	232 (49)	97 (53)	129 (45)	143 (70)	

vowel target, while the prenuclear tone gesture reaches its target, on average, 100 ms later. Moreover, both nuclear and prenuclear tone gestures reach their targets earlier in phrase-initial position than in phrase-noninitial position. Phrase-finally (only nuclear), it reaches its target, H, even earlier. Even though all speakers show the similar pattern, the rather large standard deviations (as compared to the onset of the tone gesture) is indicative of a high degree of variability.

Figure 7.5 presents density curves for the coordination of the target of the nuclear (upper panel) and prenuclear (lower panel) high tone gesture. The solid line at 0 ms marks the occurrence of the articulatory vowel target. When comparing the nuclear data with the prenuclear data, the prencular data show wider distributions, with density plateaus for both phrase-initial and phrase-noninitial position, indicating a high degree of variability. The two density curves overlap to a large degree, too, reflecting only a subtle difference between the coordination of the target of the high tone gesture (H) in the two phrasal positions. In contrast, the density curves in the nuclear data show a more narrow

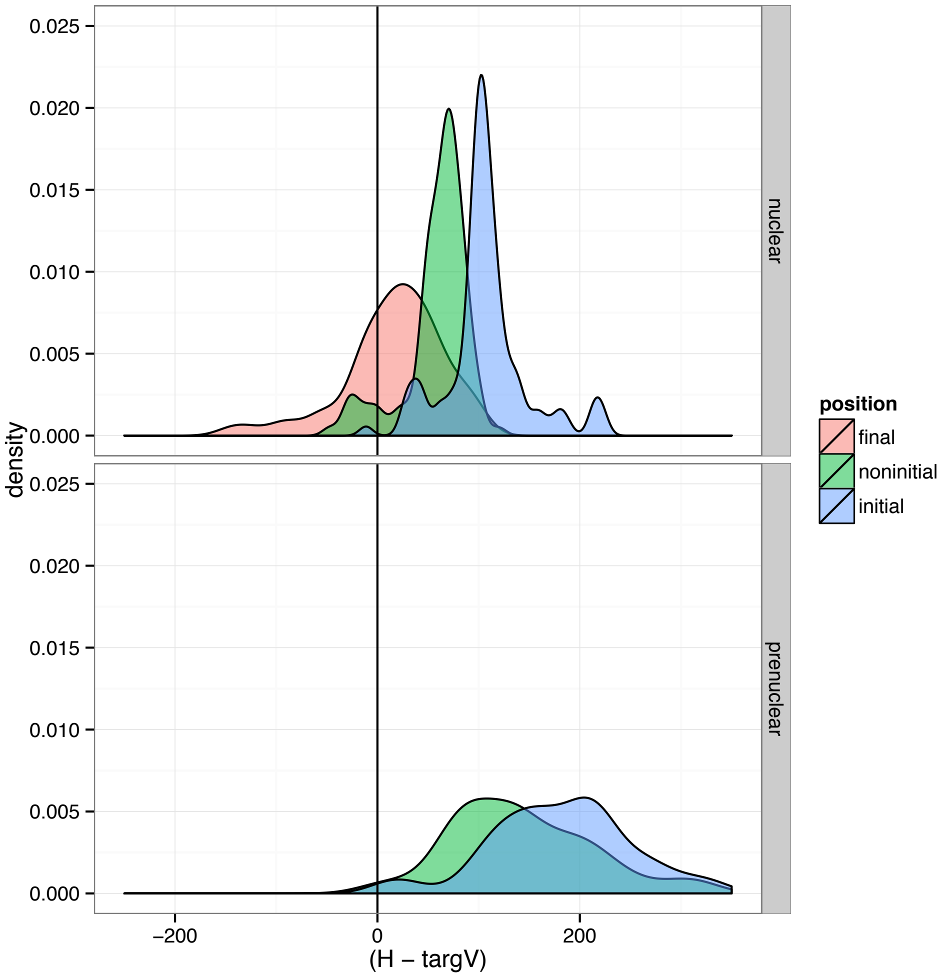


Figure 7.5: Density curve (averaged across all speakers and target words) for the coordination of the target of the tone gesture relative to the articulatory vowel target.

distribution with a distinguishable peak for each of the three phrasal positions. More specifically, the peak for the phrase-final data occurs shortly after the articulatory vowel target, followed by the peaks for phrase-noninitial and phrase-final position. That is, the nuclear tone gesture reaches its target, H, earlier than in phrase-noninitial position and earlier than in phrase-initial position. Section has already shown that the coordination of the target of the prenuclear high tone gesture is only affected by phrasal position. The target of the textitnuclear high tone gesture, H, is, however, affected by both phrasal position and word length. That is, phrase-finally, the tone gesture reaches its target earlier in monosyllables than in disyllables.

Figure 7.6 shows density curves for the target of the nuclear high tone gesture relative to the articulatory vowel target as a function of both phrasal position and target word.

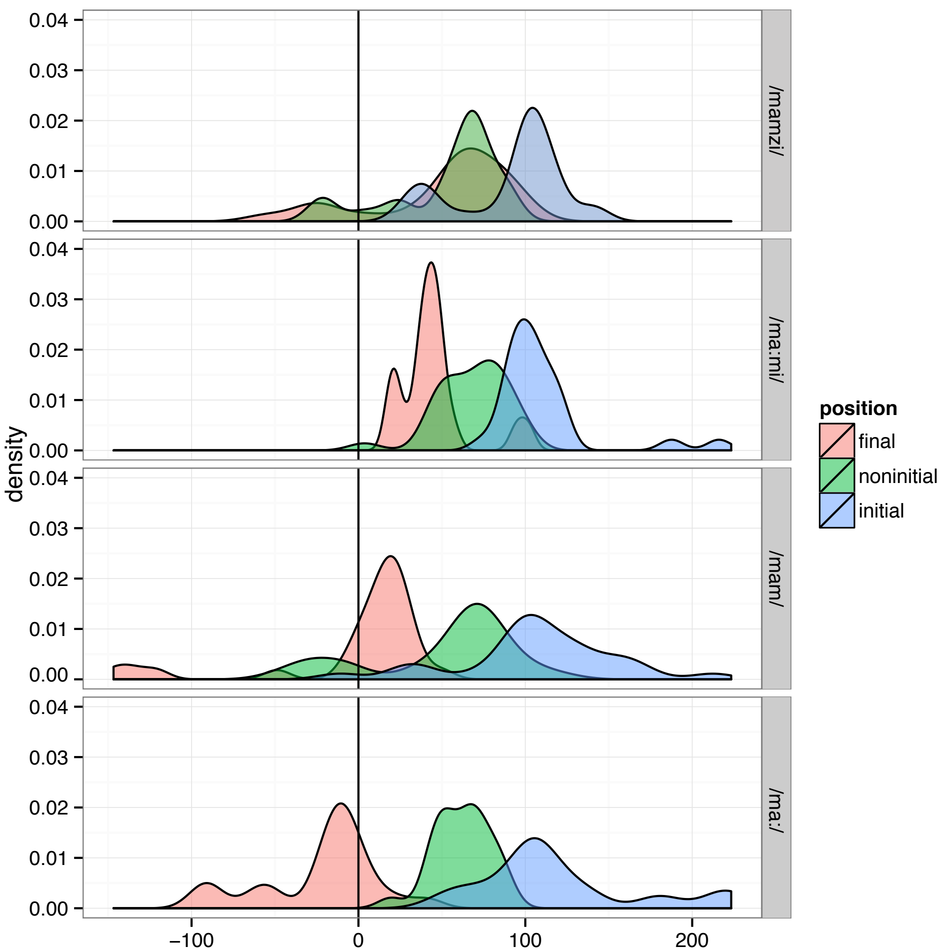


Figure 7.6: Density curves (averaged across all speakers) for the coordination of the target of the tone gesture relative to the articulatory vowel target (nuclear data only).

Zero denotes the articulatory vowel target. While the coordination of the target of the high tone gesture is not affected by the structure of the target words in the phrase-noninitial (green) and phrase-initial (blue) data, it changes as a function of word length in phrase-final position (red). For example, in /mamzi/, the density curve representing the phrase-final data (red) occurs after the articulatory vowel target and overlaps with the phrase-noninitial curve (green) to a high degree. In /ma:mi/ and /mam/, it "moves" towards the articulatory vowel target, and in /mam/ it even shows a peak before the articulatory vowel target. In other words, the target of the nuclear high tone gesture is reached earlier when there is less segmental material available as in /ma:/ as compared to /mamzi/ where there is a second (unstressed) syllable on which the tone gesture can reach its target.

Summarizing the data, the comparison of prenuclear and nuclear high tone gestures has revealed the following patterns: The onset of both the nuclear and prenuclear high tone gesture, L, tends to be coordinated with the peak velocity of the consonant's release gesture. However, there is a certain amount of speaker-specific variability in that some speakers start the nuclear high tone gesture earlier in phrase-final position. The target of the high tone gesture, H, is highly sensitive to accent status: High tone gestures in prenuclear position show a high degree of variance in their coordination with the articulatory vowel target and tend to reach their targets later than in nuclear position. In contrast, nuclear high tone gestures show a tighter coordination pattern and reach their targets with respect to their position in the phrase in the following order: phrase-final $<$ phrase-noninitial $<$ phrase-initial. Moreover, in phrase-final position, the nuclear high tone gesture reaches its target earlier in monosyllables as compared to disyllables. The next section models these effects employing the π-gesture and the μ-gesture in the coupled oscillator model of syllable structure.

7.3 Coupled oscillators

This section summarizes the key results of the present study and presents three coupling structures that account for the attested effects by employing prosodic gestures in the framework of coupled oscillators. Figure 7.7 provides schematized scores and corresponding coupling structures for prenuclear (upper panel) and nuclear (lower panel) high tone gestures.

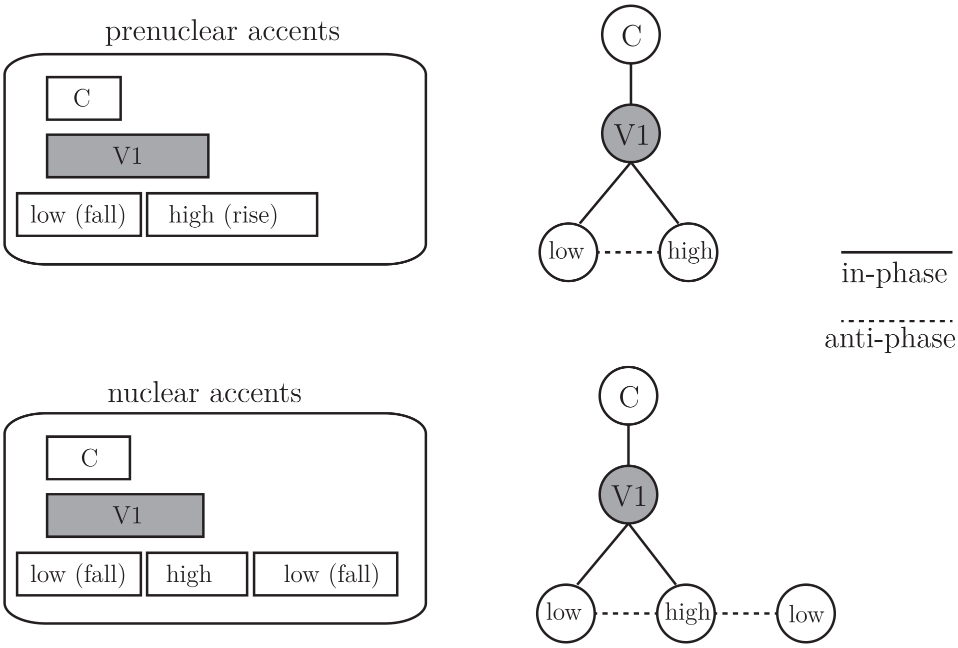


Figure 7.7: Schematized gestural scores and proposed coupling structures for nuclear (upper panel) and prenuclear pitch accents (lower panel).

Mücke et al. (2012) and Niemann et al. (2011) (see section 3.4) have already modeled rising pitch accents in German with two tone gestures, a low and a high tone gesture. The target of the low tone gesture, L, is at at the same time the onset of the high tone gesture. The coupling structures in Figure 7.7 extend this modeling to prenuclear accents. As known from consonantal clusters in complex onsets, where both consonants are coupled in-phase with the vocalic gesture, but are coupled in anti-phase with each other, a similar pattern is assumed for the rising pitch accent. Both nuclear and prenuclear rising pitch accents consist of a high tone gesture and a preceding low tone gesture. Both are coupled in-phase with the vocalic gesture and anti-phase with each each other leading to a competitive structure. However, there is a crucial difference between nuclear and prenuclear accents in that in nuclear accents an *additional* low tone gesture follows the high tone gesture. The onset of this additional low tone gesture corresponds to the target of the preceding high tone gesture. This additional tone gesture is coupled in anti-phase with the preceding high tone gesture. It thus forces the high tone gesture to reach its target, the F0 peak, earlier. In prenuclear accents, on the other hand, the high tone gesture can be realized without any constraints as no low tone gesture follows, i.e. the F0 peak can align later as compared to nuclear accents.

The proposal of an additional low tone gesture following the high tone gesture in nuclear accents is similar to the concept of the phrase accent used in autosegmental-metrical analyses (see section 2.1). Evidence for a low tone gesture following the high tone gesture comes from visual inspection of nuclear and prenuclear accents. Usually, nuclear F0 peaks are immediately followed by a sharp fall in F0 with an evident low "elbow", whereas the prenuclear accents show more variability. The overall impression is that the prenuclear F0 peak is followed by a sagging transition without a clearly identifiable low tone following it.

Figure 7.8 presents gestural scores based on means for each of the four speakers (S1-S4) producing the target word /ma:.mi/ in phrase-noninitial position. Each score consists of two tiers: The "TB" (tongue body) tier indicates the vocalic gestures while the "Tone" tier indicates the tone gestures. The box labeled "V1" represents the interval from the onset to the target of the vocalic gesture, i.e. the tongue body lowering for the production

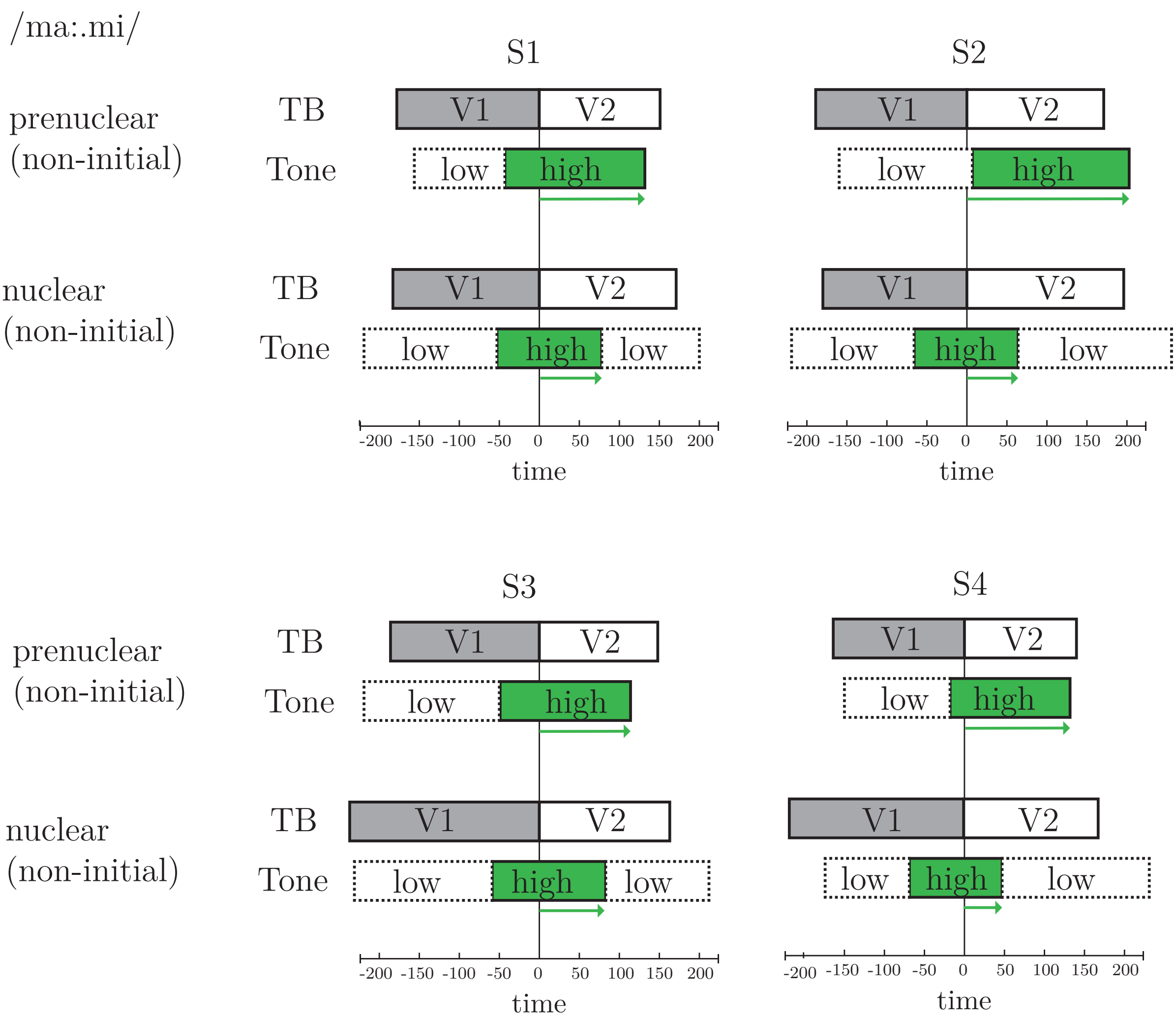


Figure 7.8: Gestural scores for the four speakers producing the target word /ma:.mi/ in phrase-noninitial position.

of the accented vowel /a:/. The box labeled "V2" represents the vocalic gesture (tongue body raising) for the following syllable, which contains the vowel /i/. The green boxes represent the activation of the high tone gesture. Note that the onset of the low tone gesture preceding the high tone gesture is not detectable in the signal. Zero denotes the articulatory vowel target for the production of the vowel /a:/ (V1). The green arrow indicates the time between the articulatory target and the end of the high tone gesture. All speakers show a similar pattern: The nuclear high tone gestures reach their targets, the F0 peaks, earlier than the prenuclear ones. This is caused by the following low tone gesture coupled anti-phase with the high tone gesture in nuclear accents, inducing time pressure on the realisation of the preceding high tone gesture.

The following coupling structures account for the positional effects found in both prenuclear and nuclear rising pitch accents.

Figure 7.9 shows schematized scores and the proposed coupling structures for (nuclear) high tone gestures in phrase-initial (top panel), phrase-noninitial (middle panel) and phrase-final position (lower pannel). The vocalic gesture representing the accented vowel V1 is shaded. In phrase-initial position, both the π-gesture (marking the boundary) and the μ-gesture (marking stress) are coupled anti-phase and in-phase, respectively, with the vocalic gesture V1. In addition, the high tone gesture is coupled in-phase with the vocalic gesture. This coupling structure results in a rightward shift of the target of the high tone gesture; thus the F0 peak aligns later as compared to the other positions.

In phrase-noninitial position, a vocalic gesture (V0) precedes the accented vowel. This vocalic gesture is coupled anti-phase with the π-gesture marking the boundary. In this position, the π-gesture cannot exert time pressure on the high tone gesture; thus

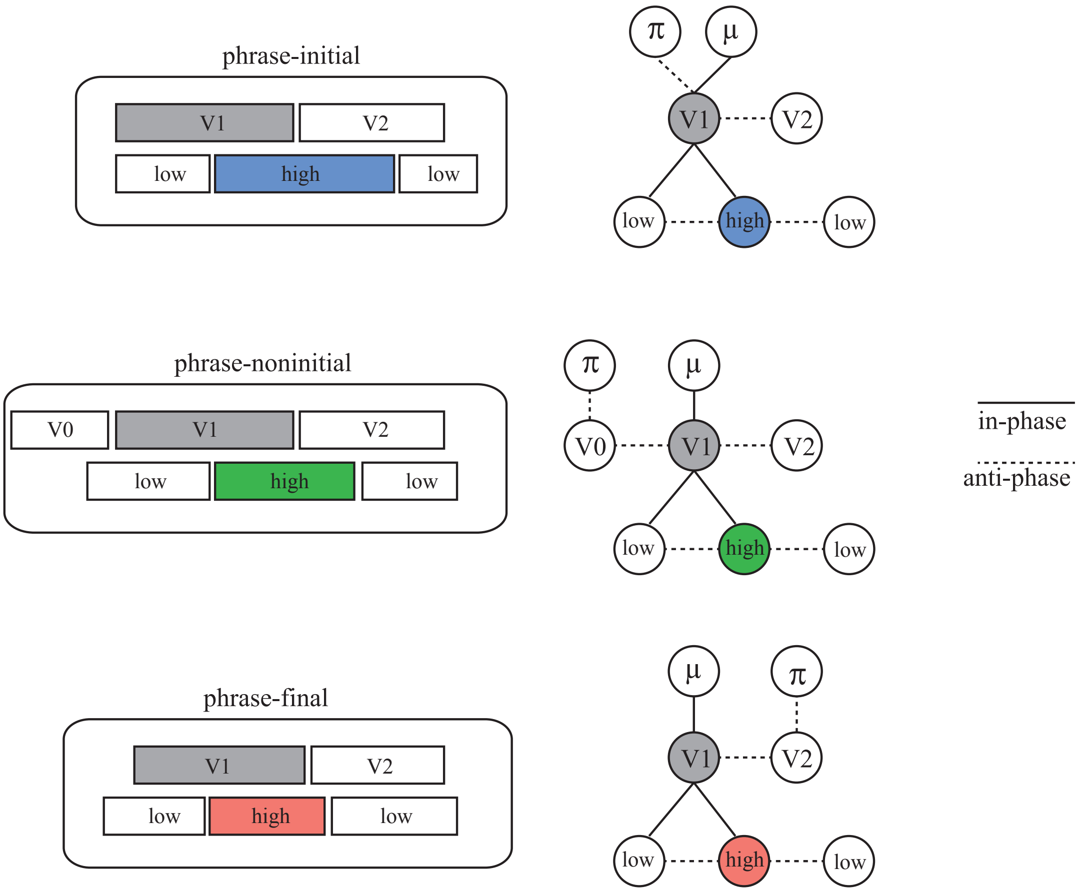


Figure 7.9: Schematized gestural scores and proposed coupling structures for the high tone gesture in phrase-initial, phrase-noninitial and phrase-final position.

the F0 peak is realized somewhat earlier in phrase-noninitial position as compared to phrase-initial position.

In phrase-final position, however, the high tone gesture needs to achieve its target earlier as compared to the non-final positions. In this position, the π-gesture is coupled anti-phase with the vowel in the following, phrase-final, unstressed syllable (V2). The time pressure resulting from the upcoming low boundary tone (the end of the low tone gesture) forces the high tone gesture to achieve its target earlier.

Figure 7.10 presents gestural scores based on means from the four speakers producing the

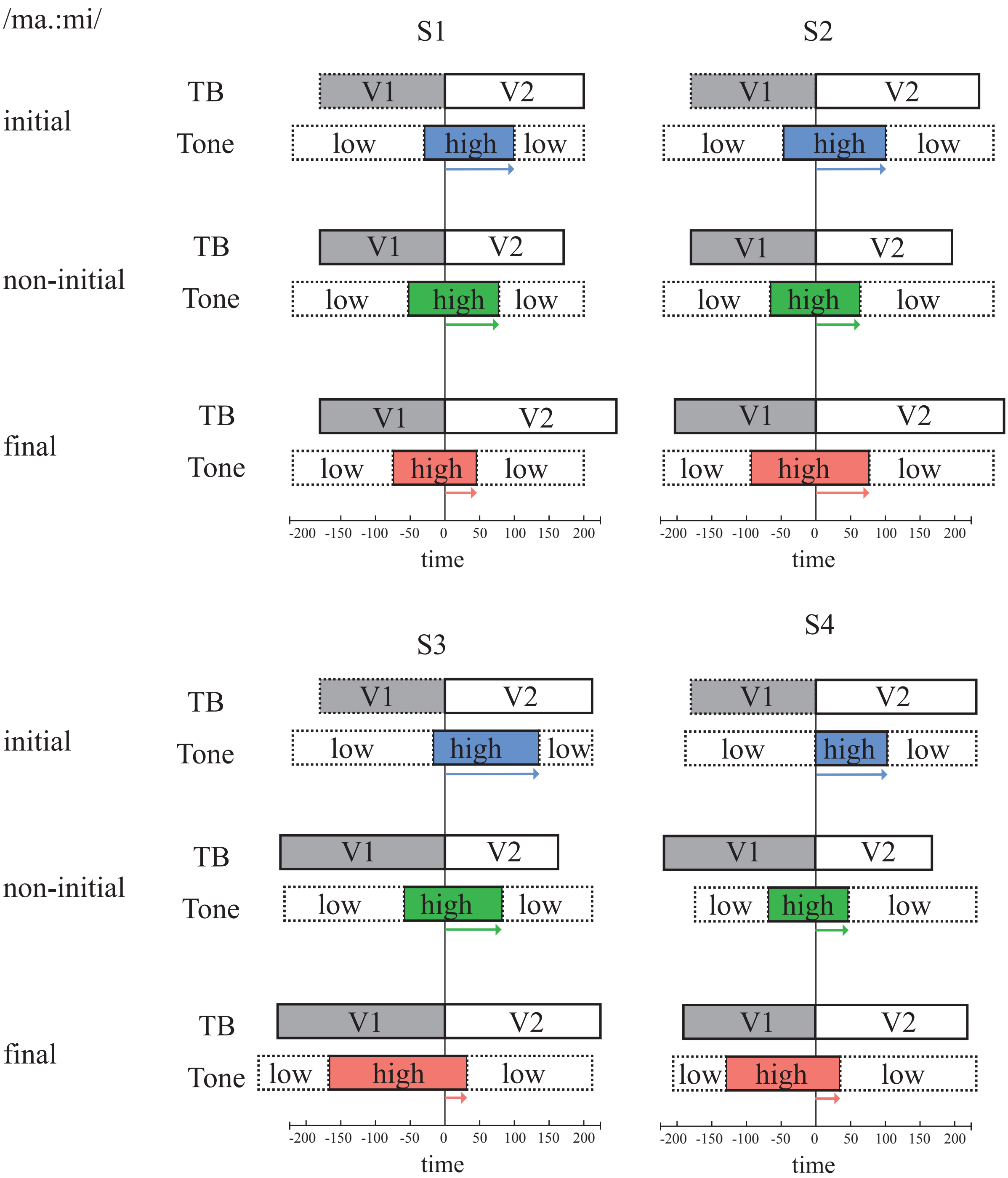


Figure 7.10: Gestural scores for the four speakers producing the target word /ma:.mi/ in phrase-initial, phrase-noninitial and phrase-final position.

target word /ma:mi/ in phrase-initial, phrase-noninitial and phrase-final position that illustrate the coordination of the high tone gesture relative to the articulatory target of the vocalic gesture for the accented vowel V1. The arrows indicate the distance between the articulatory vowel target and the target of the high tone gesture. Each speaker shows a similar pattern: The high tone gesture reaches its target earliest in phrase-final position (red), followed by the high tone gestures in phrase-noninitial (green) and phrase-initial position (blue).

The last coupling structure is concerned with the difference between monosyllables and disyllables bearing the nuclear pitch accent. In monosyllables, the high tone gesture reaches its target, H, earlier as compared to disyllables, particularly in phrase-final position.

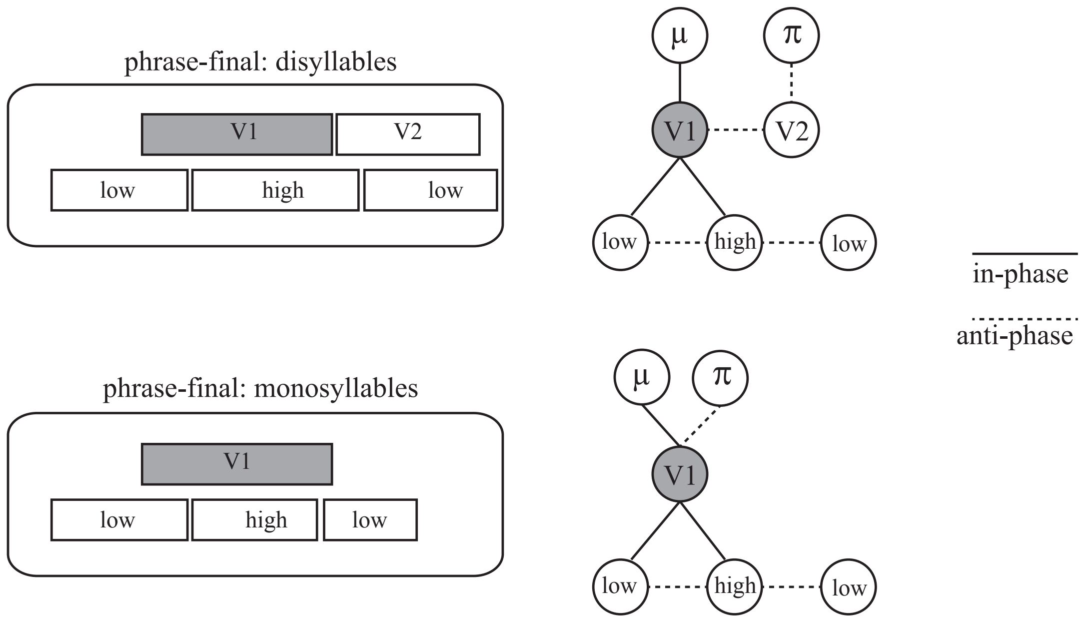


Figure 7.11: Schematized gestural scores and proposed coupling structures for phrase-final nuclear accents.

Figure 7.11 provides schematized gestural scores and corresponding coupling structures for the effect of word length in phrase-final position. The coupling structures given in Figure 7.11 account for this difference. In both disyllables and monosyllables the π-gesture is coupled anti-phase with the boundary-adjacent vowel gesture. In disyllables, the π-gesture is coupled with V2, while in monosyllabes it must be coupled with V1. In line with Katsika et al. (2014) who assume that the *pi*-gesture serve as trigger for the boundary tone gesture (here: low tone gesture), it seems plausible that the low tone gesture is initiated earlier in monosyllables than in disyllables. In turn, an earlier activation of the low tone gesture can only be achieved by superseding the high tone gesture, thus leading to an earlier F0 peak alignment.

Figure 7.12 presents gestural scores based on means for the two maximally divergent target words, the disyllabic target word /mam.zi/ and the monosyllabic target word /ma:/, in phrase-final (nuclear) position.

The red box indicates the activation interval of the nuclear high tone gesture. Zero denotes the articulatory vowel target, and the arrows indicate the time between the articulatory vowel target and the target of the high tone gesture. Except for speaker S3, the high tone gestures of all speakers reach their targets *after* the articulatory vowel target in the disyllabic target word /mam.zi/. In the monosyllable /ma:/, however, it is achieved before the articulatory vowel target (S1 and S4) or roughly co-occurs with it (S2 and S3).

In summary, the coupled oscillator model of syllable structure including prosodic gestures such as the π-gesture and μ-gesture provides suitable and sufficient tools for modeling the coordination of both nuclear and prenuclear high tone gestures. The greatest benefit of this model is that it copes with the inherint dynamical nature of gestures.

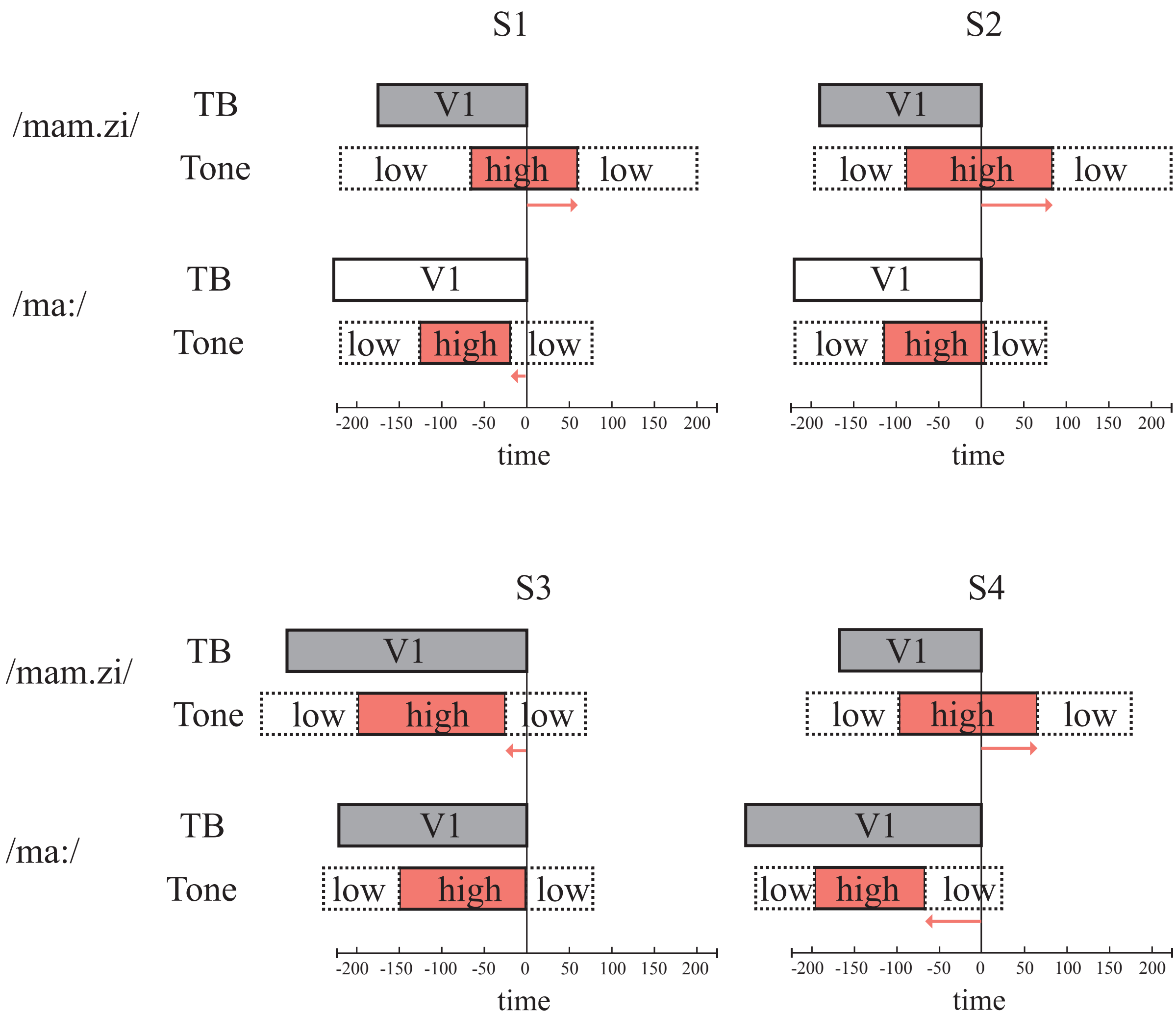


Figure 7.12: Gestural scores for the four speakers producing the target word /ma:.mi/ in phrase-initial, phrase-noninitial and phrase-final position.

8 Summary and conclusion

This thesis has investigated nuclear and prenuclear rising pitch accents by means of their alignment relative to specific landmarks in both the *acoustic* and *articulatory* domain. While in the acoustic domain the timing of accentual rises is highly variable, stable coordination patterns are found in the articulatory domain. This is likely due to the fact that within the Segmental Anchoring Hypothesis (cf. Ladd et al. 1999, Ladd 2008) tones are aligned with nearby landmarks. This approach investigates the co-occurrence of tonal targets and segmental boundaries, which is a *static* perspective of the tune-text relation. In contrast, the Articulatory Phonology approach (Browman & Goldstein 1989, 1992) defines tones as gestures that are temporally coordinated with oral constriction gestures. Rising and falling tones, whether pitch accents or lexical tones, are represented as high and low tone gesture, respectively. (Gao 2009, Mücke et al. 2012, Niemann et al. 2011, Hsieh 2011). This approach allows for modeling variability of the tune-text coordination with respect to different prosodic factors in a *dynamic* way. Within the coupled oscillator model of syllable structure (Nam & Saltzman 2003, Goldstein et al. 2008, Nam, Goldstein & Saltzman 2009) the specific timing patterns of nuclear and prenuclear high *tone gestures* with oral constriction gestures result from specific coupling structures, that account for their coordination. Furthermore, effects of higher prosodic units such as the word or the intonational phrase can be modeled by using prosodic gestures, the π-gesture and the μ-gesture, that can be attached to the tune-text couplings in order to model stress- and boundary-related effects in both the temporal and spatial domain.

Four subjects took part in this study, which employed kinematic recordings (electromagnetic articulography) in order to trace the movements of the lips and the tongue body. This study focuses on the effects of phrasal position, word length and syllable structure on the alignment of nuclear and prenuclear rising pitch accents. More specifically, the corpus included trochaic mono- and disyllabic target words each with an open and closed stressed syllable. Carrier sentences were designed such that target words bearing a nuclear or prenuclear rising accent were either placed in phrase-initial position (at the left edge of an intonation phrase) or in phrase-noninitial position (with an unstressed syllable preceding the accented syllable). In addition, target words bearing a nuclear accent were also investigated in phrase-final position (at the right edge of an intonation phrase). Both the beginning and the end of the accentual rises were measured relative to acoustically-defined landmarks such as the acoustic onset of the accented vowel, and to articulatory-defined landmarks such as the lip opening gesture and the articulatory target of the accented vowel.

In the acoustics, the beginning of both nuclear and prenuclear rises tends to be stably aligned with the accented vowel. More specifically, the beginning of the nuclear rise is aligned shortly after or just before the onset of the accented vowel. The beginning of the prenuclear rise aligns sometime later and displays subtle speaker-specific alignment differences. In contrast, the end of the accentual rise, the F0 peak, is prone to a large amount of variability. In general, *prenuclear* F0 peaks were found to align later than nuclear F0 peaks. More specifically, they tend to align after the syllable boundary in monosyllables or even later, in the postaccented vowel, in disyllables. Furthermore, prenuclear F0 peaks were less affected by the factors under investigation, i.e. the effects were not as distinctive as for the nuclear F0 peaks, and they seem to be speaker-specific. In contrast, *nuclear* F0 peaks were found to align within or shortly after the accented syllable. Their alignment is highly variable. They tend to align earlier in phrase-final position as compared to in non-final positions, earlier in open than in closed syllables and earlier in monosyllables than in disyllables. These effects, however, interact and are expressed differently, that is, the effect of phrasal position is more pronounced in monosyllables than in disyllables and interacts with syllable structure. More specifically, in nonfinal position, the F0 peak aligns with or shortly after the vowel offset in open syllables, while it aligns later, in the coda consonant, in closed syllables. In these positions, there is only a small alignment difference between monosyllabic and disyllabic target words in that the F0 peak aligns sometime earlier in monosyllables. In phrase-final position, however, a huge difference can be found between monosyllables and disyllables. In monosyllables, the F0 peak is retracted into the accented vowel to a large degree, whereas disyllables only show a small retraction in that direction.

In the articulation, the beginning of both the nuclear and prenuclear rise shows a stable coordination pattern with the peak velocity of the consonant's release gesture. This coordination is not affected by phrasal position, word length or syllable structure. The end of both nuclear and prenuclear accentual rises shows a stable coordination pattern with the articulatory target for the accented vowel. This coordination is also not affected by the syllable structure of the target words under investigation. However, there is an effect of accent status. The end of the accentual rise, the F0 peak, is achieved earlier in nuclear pitch accents than in prenuclear pitch accents. In addition, there is an effect of the upcoming phrase boundary on F0 peak alignment in nuclear accents. Due to time pressure, the F0 peak aligns earlier in phrase-final position as compared to in nonfinal positions. Furthermore, phrase-finally, the F0 peak aligns earlier in monosyllables as compared to disyllables.

The coordination between high tone gestures and articulatory gestures in nuclear and prenuclear accents can be modeled in the coupled oscillator model as follows: *Prenuclear* rising pitch accents consist of two tone gestures, a high tone gesture and a low tone gesture. As in consonant clusters, there is a competitive structure between the two tone gestures as both are coupled in-phase with the vocalic gesture. In *nuclear* rising pitch accents, the F0 peaks aligns earlier. Thus, another low tone gesture following the high tone gesture is proposed. This low tone gesture is coupled anti-phase with the preceding high tone gesture. This coupling structure exerts time pressure on the high tone gesture, resulting in an earlier F0 peak alignment relative to the articulatory vowel target.

The positional effects are modeled by employing prosodic gestures, namely the π-gesture and the μ-gesture. First, in phrase-initial position, both the π-gesture and the μ-gesture are coupled anti-phase and in-phase, respectively, with the vocalic gesture. This coupling structure leads to a rightward shift of the target of the high tone gesture resulting in a later F0 peak alignment as compared to phrase-noninitial and phrase-final position. Second, in phrase-noninitial position, an unstressed syllable precedes the accented syllable. The vocalic gesture of this syllable is coupled anti-phase with the π-gesture. Only the μ-gesture is coupled in-phase with the vocalic gesture of the accented syllable; thus there is no leftward pressure in that the F0 peak aligns earlier than in phrase-initial position. Third, in phrase-final position, the vocalic gesture of the accented syllable is coupled in-phase with the μ-gesture. In disyllabic target words, the vocalic gesture of the following syllable is coupled anti-phase with the π-gesture. This coupling structure exerts pressure on the target achievement of the high tone gesture. The result is an even earlier F0 peak alignment in this position.

Furthermore, the earlier F0 peak alignment in phrase-final monosyllables as compared to disyllables can be modelled as follows: In phrase-final monosyllables, *both* the π-gesture and the μ-gesture are coupled anti-phase and in-phase, respectively, with the vocalic gesture of that syllable; thus the high tone gesture must achieve its target earlier in this position.

A following step will be the implementation of these coupling structures in TADA (cf. Saltzman et al. 2008, Nam & Saltzman 2003), the **TA**sk-**D**ynamic **A**pplication, that generates trajectories on the basis of coupling graphs specifying interarticulatory timing, and a comparison of the results with the data obtained in this study. Another fruitful starting point for further work would be the representation of the attested effects in terms of potential functions (cf. Tuller et al. 1994, Gafos & Beňuš 2006). This would be particularly interesting regarding the speaker-specific strategies revealed in the present study and as to which control parameter can account for these.

In conclusion, this thesis provides further evidence that we obtain a better and deeper understanding of speech, particularly of the synchronisation of text and tune, if we regard it as a dynamical system involving the interplay of coordinative structures, thus extending our knowledge from purely acoustic-based approaches.

References

Arvaniti, Amalia (2011). The representation of intonation. In Marc van Oostendorp, Ewen Colin J., Elizabeth Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*, vol. 2, 757–780. John Wiley & Sons.

Arvaniti, Amalia (2012). Segment-to-tone association. In Abigail Cohn, Cécile Fougeron & Marie Huffman (eds.), *The Oxford Handbook of Laboratory Phonology*, 265–274. Oxford University Press.

Arvaniti, Amalia & D. Robert Ladd (1995). Tonal alignment and the representation of accentual targets. In *Proceedings of the 13th International Congress of Phonetic Sciences*, 220–223.

Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen (1998). Stability of tonal alignment: the case of greek prenuclear accents. *Journal of Phonetics* 26. 3–25.

Ashby, Michael G. (1978). A study of two English nuclear tones. *Language and speech* 21(4). 326–336.

Atterer, Michaela & D. Robert Ladd (2004). On the phonetics and phonology of segmental anchoring of f0: Evidence from German. *Journal of Phonetics* 32. 177–197.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Beckman, Mary E. & Janet B. Pierrehumbert (1986). Intonational structure in Japanese and English. *Phonology* 3(01). 255–309.

Bolinger, Dwight L. (1951). Intonation: levels versus configurations. *Word* 7(3). 199–210.

Bombien, Lasse, Steve Cassidy, Jonathan Harrington, Tina John & Sallyanne Palethorpe (2006). Recent developments in the Emu speech database system. In *Proceedings of the 11th SST Conference*, 313–316. Auckland.

Browman, Catherine P. & Louis Goldstein (1986). Towards an Articulatory Phonology. *Phonology Yearbook* 3. 219–252.

Browman, Catherine P. & Louis Goldstein (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica* 45. 140–155.

Browman, Catherine P. & Louis Goldstein (1989). Articulatory gestures as phonological units. *Phonology* 6. 201–251.

Browman, Catherine P. & Louis Goldstein (1990). Tiers in Articulatory Phonology, with some implications for casual speech. In John Kingston & Mary Beckman (eds.), *Papers in Laboratory Phonology I. Between the Grammar and Physics of Speech*, 341–376. Cambridge University Press.

Browman, Catherine P. & Louis Goldstein (1992). Articulatory Phonology: An Overview. *Phonetica* 49. 155–180.

Browman, Catherine P. & Louis Goldstein (1995). Dynamics and Articulatory Phonology. In Robert F. Port & Timothy Van Gelder (eds.), *Mind as motion: explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.

Bruce, Gösta (1977). *Swedish word accents in sentence perspective*. Lund University.

Büring, Daniel (2007). Semantics, intonation, and information structure. In Charles Reiss & Gillian Ramchand (eds.), *The Oxford Handbook of Linguistic Interfaces*, 445–474. Oxford University Press.

Byrd, Dani (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* 57. 3–16.

Byrd, Dani, Abigail Kaun, Shrikanth Narayanan & Elliot Saltzman (2000). Phrasal signatures in articulation. In Michael B. Broe & Janet B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V. Acquisition and the Lexicon*, vol. 5, 70–87. Cambridge University Press.

Byrd, Dani, Jelena Krivokapić & Sungbok Lee (2006). How far, how long: on the temporal scope of prosodic boundary effects. *The Journal of the Acoustical Society of America* 120(3). 1589–1599.

Byrd, Dani & Elliot Saltzman (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31(2). 149–180.

Cassidy, Steve & Jonathan Harrington (2001). Multi-level annotation in the EMU speech database management system. *Speech Communication* 33(61-77).

Cho, Taehong (2006). Manifestation of prosodic structure in articulatory variation: evidence from lip kinematics in English. In Louis Goldstein, Douglas H. Whalen & Catherine T. Best (eds.), *Laboratory phonology 8*, vol. 8, 519–548. Mouton De Gruyter.

Cho, Taehong (2016). Prosodic boundary strengthening in the phonetics-prosody interface. *Language and Linguistics Compass* 10(3). 120–141.

Cho, Taehong & Patricia A. Keating (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29(2). 155–190.

Chomsky, Noam & Morris Halle (1968). *The sound pattern of English.* New York: Harper & Row.

Crystal, David (1972). The intonation system of English. In Dwight L. Bolinger (ed.), *Intonation*, 110–136. Penguin.

D'Imperio, Mariapaola (2000). *The role of perception in defining tonal targets and their alignment*. The Ohio State University: Department of Linguistics Doctoral dissertation.

D'Imperio, Mariapaola (2002). Italian intonation: an overview and some questions. *Probus* 14. 37–69.

D'Imperio, Mariapaola (2012). Tonal alignment. In Abigail Cohn, Cécile Fougeron & Marie Huffman (eds.), *The Oxford Handbook of Laboratory Phonology*, 275–287. Oxford University Press.

D'Imperio, Mariapaola, Robert Espesser, Hélène Loevenbruck, Caroline Menezes, Noël Nguyen & Pauline Welby (2007). Are tones aligned to articulatory events? evidence from Italian and French. In Jennifer Cole & José Ignacio Hualde (eds.), *Laboratory Phonology 9*. Mouton De Gruyter.

Ellis, Paul D. (2010). *The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press.

Fougeron, Cécile (2001). Articulatory properties of initial segments in several prosodic constituents in french. *Journal of phonetics* 29(2). 109–135.

Fujisaki, Hiroya (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. In Peter F. MacNeilage (ed.), *The production of speech*, 39–55. Springer.

Gafos, Adamantios & Stefan Beňuš (2006). Dynamics of phonological cognition. *Cognitive science* 30(5). 905–943.

Gao, Man (2009). Gestural coordination among vowel, consonant and tone gestures in mandarin chinese. *Chinese Journal of Phonetics* 2. 43–50.

Goldsmith, John (1976). An overview of autosegmental phonology. *Linguistic analysis* 2. 23–68.

Goldstein, Louis (2011). Back to the past tense in english. *Representing language: essays in honor of Judith Aissen. Santa Cruz: Linguistics Research Center*. 69–88.

Goldstein, Louis, Dani Byrd & Elliot Saltzman (2006). The role of the vocal tract gestural action units in understanding the evolution of phonology. In Michael A. Arbib (ed.), *Action to language via the mirror neuron system*. Cambridge University Press.

Goldstein, Louis, Ioana Chitoran & Elisabeth Selkirk (2007). Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. In *Proceedings of the International Congress on Phonetic Sciences*, 241–242. Saarbrücken.

Goldstein, Louis, Hosung Nam, Elliot Saltzman & Ioana Chitoran (2008). Coupled oscillator planning model of speech timing and syllable structure. In *Proceedings of the 8th Phonetics Congress of China*.

Grice, Martine (2006). Intonation. In Keith Brown (ed.), *Encyclopedia of Language & Linguistics*, 2nd edn., vol. 5, 778–788. Oxford: Elsevier.

Grice, Martine, D. Robert Ladd & Amalia Arvaniti (2000). On the place of phrase accents in intonational phonology. *Phonology* 17(2). 143–185.

Gussenhoven, Carlos (2002). Phonology of intonation. *Glot International* 6(9/10). 271–284.

Haken, Hermann, J.A. Scott Kelso & Heinz Bunz (1985). A theoretical model of phase transitions in human hand movements. *Biological cybernetics* 51(5). 347–356.

Hall, T. Alan (2011). *Phonologie. Eine Einführung*. 2nd edn. Berlin/New York: de Gruyter.

Halliday, Michael A.K. (1970). *A course in spoken English: Intonation*. Oxford University Press.

Harrington, Jonathan (2010). *Phonetic analysis of speech corpora*. John Wiley & Sons.

Hawkins, Sarah (1992). An introduction to task dynamics. In G.J. Docherty & D. Robert Ladd (eds.), *Papers in Laboratory Phonology II. Gestures, Segment, Prosody*, vol. 2. Cambridge University Press.

Hermes, Anne, Doris Mücke & Martine Grice (2013). Gestural coordination of Italian word-initial clusters: the case of 'impure s'. *Phonology* 30(01). 1–25.

Hermes, Anne, Doris Mücke, Martine Grice & Henrik Niemann (2008). Articulatory indicators of syllable affiliation in word initial consonant clusters in Italian. In *Proceedings of the 8th International Seminar on Speech Production*, 433–436. Strasbourg.

Hermes, Anne, Rachid Ridouane, Doris Mücke & Martine Grice (2011). Gestural coordination in Tashlhiyt syllables. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 859–862.

Hirst, Daniel & Albert Di Cristo (1998). A survey of intonation systems. In Daniel Hirst & Albert Di Cristo (eds.), *Intonation Systems a Survey of Twenty Languages*. Cambridge University Press.

Hockett, Charles F. (1955). A manual of phonology. *Anthropology and Linguistics* 11.

Hsieh, Fang-Ying (2011). A gestural account of Mandarin tone 3 variation. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 890–893. Hong Kong.

Jones, Daniel (1972). *An outline of English phonetics*. Cambridge University Press.

Katsika, Argyro, Jelena Krivokapić, Christine Mooshammer, Mark Tiede & Louis Goldstein (2014). The coordination of boundary tones and its interaction with prominence. *Journal of phonetics* 44. 62–82.

Kelso, J.A. Scott (1981). On the oscillatory basis of movement. In *Bulletin of the Psychonomic Society*, vol. 18, 63–63.

Kelso, J.A. Scott (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 6. R1000–R1004.

Kelso, J.A. Scott & J. P. Scholz (1985). Cooperative phenomena in biological motion. In Hermann Haken (ed.), *Synergetics of complex systems: operational principles in neurobiology, physical systems, and computers*, 213–232. Berlin, Heidelberg: Springer Berlin Heidelberg.

Kelso, J.A. Scott, G. Schöner, J.P. Scholz & H. Haken (1987). Phase-locked modes, phase transitions and component oscillators in biological motion. *Physica Scripta* 35(1). 79–87.

Kleber, Felicitas & Tamara Rathcke (2008). More on the "segmental anchoring" of prenuclear rises: evidence from East Middle German. In *Proceedings of the 4th Conference on Speech Prosody*, 6–9.

Krivokapić, Jelena (2007). Prosodic planning: effects of phrasal length and complexity on pause duration. *Journal of phonetics* 35(2). 162–179.

Krivokapić, Jelena & Dani Byrd (2012). Prosodic boundary strength: an articulatory and perceptual study. *Journal of phonetics* 40(3). 430–442.

Kühnert, Barbara, Phil Hoole & Christine Mooshammer (2006). Gestural overlap and c-center in selected French consonant clusters. In *7th International Seminar on Speech Production (ISSP)*, 327–334.

Ladd, D. Robert (2006). Segmental anchoring of pitch movements: autosegmental association or gestural coordination? *Rivista di Linguistica* 18(1). 19–38.

Ladd, D. Robert (2008). *Intonational Phonology*. 2nd edn. Vol. 119 (Cambridge Studies in Lingustics). Cambridge: Cambridge University Press.

Ladd, D. Robert, Dan Faulkner, Hanneke Faulkner & Astrid Schepman (1999). Constant “segmental anchoring" of f0 movements under changes in speech rate. *Journal of the Acoustical Society of America* 106(3). 1543–1554.

Ladd, D. Robert, Ineke Mennen & Astrid Schepman (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America* 107(5). 2685–2696.

Ladd, D. Robert, Astrid Schepman, Laurence White, Louise May Quarmby & Rebekah Stackhouse (2009). Structural and dialectal effects on pitch peak alignment in two varieties of British English. *Journal of Phonetics* 37(2). 145–161.

Laver, John (1994). *Principles of phonetics*. Cambridge University Press.

Marin, Stefania (2013). The temporal organization of complex onsets and codas in Romanian: a gestural approach. *Journal of Phonetics* 41(3). 211–227.

Marin, Stefania & Marianne Pouplier (2010). Temporal organization of complex onsets and codas in American English: testing the predictions of a gestural coupling model. *Motor Control* 14(3). 380–407.

Mücke, Doris, Martine Grice, Johannes Becker & Anne Hermes (2009). Sources of variation in tonal alignment: evidence from acoustic and kinematic data. *Journal of Phonetics* 37. 321–338.

Mücke, Doris, Martine Grice & Anne Hermes (2008). The vowel triggers the tone: evidence from German. In *8th Phonetic Conference of China*. Bejing, China.

Mücke, Doris & Anne Hermes (2007). Phrase boundaries and peak alignment: an acoustic and articulatory study. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 997–1000. Saarbrücken.

Mücke, Doris, Hosung Nam, Anne Hermes & Louis 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. Mouton De Gruyter.

Mücke, Doris, Jagoda Sieczkowska, Henrik Niemann, Martine Grice & Grzegorz Dogil (2010). Sonority profiles, gestural coordination and phonological licensing: obstruent-sonorant clusters in polish. In *Proceedings of the 12th Conference on Laboratory Phonology.* Albuquerque.

Nam, Hosung, Louis Goldstein & Elliot Saltzman (2009). Self-organization of syllable structure: a coupled oscillator model. In Francois Pellegrino, Egidio Marsico, Ioana Chitoran & Christophe Coupé (eds.), *Approaches to phonological complexity*, 299–328. Mouton De Gruyter.

Nam, Hosung, Louis Goldstein, Elliot Saltzman & Dani Byrd (2004). Tada: an enhanced, portable task dynamics model in MATLAB. *The Journal of the Acoustical Society of America* 115(5). 2430–2430.

Nam, Hosung & Elliot Saltzman (2003). A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th international congress of phonetic sciences*, 2253–2256. Barcelona.

Niemann, Henrik, Martine Grice & Doris Mücke (2014). Segmental and positional effects in tonal alignment: an articulatory approach. In Susanne Fuchs, Martine Grice, Anne Hermes, Leonarda Lancia & Doris Mücke (eds.), *Proceedings of the 10th International Seminar on Speech Production*, 285–288. Cologne.

Niemann, Henrik & Doris Mücke (2015). Effects of phrasal position and metrical structure on alignment patterns of nuclear pitch accents in German: acoustics and articulation. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow.

Niemann, Henrik, Doris Mücke, Hosung Nam, Louis Goldstein & Martine Grice (2011). Tones as gestures: the case of Italian and German. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 1486–1489. Hong Kong.

O'Connor, Joseph Desmond & Gordon Frederick Arnold (1973). *Intonation of colloquial english: a practical handbook*. Longman.

Öhman, S.E.G. (1966). Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* 39(2). 151–168.

Pastätter, Manfred & Marianne Pouplier (2014). The temporal coordination of Polish onset and coda clusters containing sibilants. In Susanne Fuchs, Martine Grice, Anne Hermes, Leonarda Lancia & Doris Mücke (eds.), *Proceedings of the 10th International Seminar on Speech Production*. Cologne.

Pierrehumbert, Janet (1980). *The phonology and phonetics of English intonation*. MIT Doctoral dissertation.

Pierrehumbert, Janet & Mary Beckman (1988). *Japanese tone structure* (Linguistic Inquiry Monograph (15)). Cambridge, MA: MIT Press.

Pike, Kenneth L. (1945). *The intonation of American English*. Ann Arbor: The University of Michigan Press.

Pouplier, Marianne (2012). The gestural approach to syllable structure: universal, language-and cluster-specific aspects. In Susanne Fuchs, Melanie Weirich, Daniel Pape & Pascal Perrier (eds.), *Speech planning and dynamics*, vol. 1 (Speech Production and Perception), 63–96. Peter Lang.

Prieto, Pilar (2011). Tonal alignment. In Marc van Oostendorp, Ewen Colin J., Elizabeth Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*, 1185–1203. John Wiley & Sons.

Prieto, Pilar, Doris Mücke, Johannes Becker & Martine Grice (2007). Coordination patterns between pitch movements and oral gestures in Catalan. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 989–992. Saarbrücken.

Prieto, Pilar, Jan van Santen & Julia Hirschberg (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics* 23. 429–451.

Prieto, Pilar & Francisco Torreira (2007). The segmental anchoring hypothesis revisited: syllable structure and speech rate effects on peak alignment in Spanish. *Journal of Phonetics* 35. 473–500.

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria. `https://www.R-project.org`.

Rathcke, Tamara & Jonathan Harrington (2007). The phonetics and phonology of high and low tones in two falling f0-contours in standard German. In *Proceedings of the 8th Interspeech*, 982–985. Antwerp.

Rosenbaum, David A. (2009). *Human motor control.* Academic Press.

Saltzman, Elliot (1985). Task dynamic coordination of the speech articulators: a preliminary model. *Status Report on Speech Research SR-84, Haskins Laboratories.* 1–18.

Saltzman, Elliot (1995). Dynamics and coordinate systems in skilled sensorimotor activity. In Robert F. Port & Timothy van Gelder (eds.), *Mind as motion: explorations in the dynamics of cognition*, 149–173. Cambridge, MA: MIT Press.

Saltzman, Elliot & J.A. Scott Kelso (1987). Skilled actions: a task-dynamic approach. *Psychological review* 94(1). 84.

Saltzman, Elliot & Kevin G. Munhall (1989). A dynamical approach to gestural patterning in speech prodution. *Ecological Psychology* 7(4). 333–382.

Saltzman, Elliot, Hosung Nam, Louis Goldstein & Dani Byrd (2006). The distinctions between state, parameter and graph dynamics in sensorimotor control and coordination. In Mark L. Latash & Francis Lestienne (eds.), *Motor control and learning*. New York: Springer.

Saltzman, Elliot, Hosung Nam, Jelena Krivokapic & Louis Goldstein (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of the 4th international conference on speech prosody*, 175–184. Campinas.

Schepman, Astrid, Robin Lickley & D. Robert Ladd (2006). Effects of vowel length and "right context" on the alignment of Dutch nuclear accents. *Journal of Phonetics* 34(1). 1–28.

Shattuck-Hufnagel, Stefanie & Alice E. Turk (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247.

Shaw, Jason, Adamantios Gafos, Philip Hoole & Chakir Zeroual (2009). Syllabification in Moroccan Arabic: evidence from patterns of temporal stability. *Phonology* 26. 187–215.

Silverman, Kim E.A. & Janet B. Pierrehumbert (1990). The timing of prenuclear high accents in English. In John Kingston & Mary E. Beckman (eds.), *Papers in Laboratory Phonology I. Between the Grammar and Physics of Speech*. Cambridge University Press.

Steele, Shirley A. (1986). Nuclear accent f0 peak location: effects of rate, vowel, and number of following syllables. *Journal of the Acoustical Society of America* 80. 51.

't Hart, Johan, René Collier & Antonie Cohen (1990). *An experimental-phonetic approach to speech melody*. Cambridge University Press.

Tabain, Marija (2003). Effects of prosodic boundary on /aC/ sequences: articulatory results. *The Journal of the Acoustical Society of America* 113(5). 2834–2849.

Tilsen, Sam, Draga Zec, Christina Bjorndahl, Becky Butler, Marie-Josee L'Esperance, Alison Fisher, Linda Heimisdottir, Margaret Renwick & Chelsea Sanker (2012). A cross-linguistic investigation of articulatory coordination in word-initial consonant clusters. *Cornell Working Papers in Phonetics and Phonology.*

Trager, George L. & Henry L. Smith (1951). *An outline of English structure.* Oklahoma: Battenburg Press.

Tuller, Betty, Pamela Case, Mingzhou Ding & J.A. Scott Kelso (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human perception and performance* 20(1). 3–16.

Turvey, Michael T. (1990). Coordination. *American Psychologist* 45(8). 938–853.

Turvey, Michael T., R.C. Schmidt & L.D. Rosenblum (1989). 'Clock' and 'motor' components in absolute coordination of rhythmic movements. *Neuroscience* 33(1). 1–10.

Wichmann, Anne, Jill House & Toni Rietveld (2000). Discourse constraints on f0 peak timing in English. In Antonis Botinis (ed.), *Intonation. Analysis, Modelling and Technology*, 163–182. Springer.

«Once More, with Feeling.»