Available online at www.sciencedirect.com
ScienceDirect
Journal of Phonetics 37 (2009) 321–338
Journal of Phonetics
www.elsevier.com/locate/phonetics

Sources of variation in tonal alignment: Evidence from acoustic and kinematic data

Doris Mücke*, Martine Grice, Johannes Becker, Anne Hermes

IfL Phonetik, University of Cologne, Herbert-Lewin-Straße 6, 50931 Cologne, Germany

Received 16 February 2007; received in revised form 11 March 2009; accepted 29 March 2009

Abstract

This study is concerned with the alignment of f0 peaks in rising LH pitch accents in German, both in relation to acoustically defined segments, referred to as *segmental anchors*, as well as to dynamically defined speech gestures, referred to as *articulatory anchors*. The effects investigated were the effects of syllable structure (test words **ˈCV:**CV and **ˈCVC**V, where the test syllable is open or closed, respectively), dialectal background (the varieties of German spoken in Düsseldorf and Vienna), and accent status in the intonational hierarchy (prenuclear and nuclear accents).

As reported for related languages, peaks in closed syllables tended to be later than those in open syllables. However, it was only in nuclear accents that those differences were systematic for all four speakers. Thus only limited support can be provided for an alignment with the syllable edge.

Although there was a tendency for Southern varieties to have later peaks than Northern ones, as also found in previous studies, alignment latencies of individual speakers in the two dialectal groups overlapped. These results support a gradient view of dialectal variation in tonal alignment. In this view, dialectal differences are not represented symbolically. Rather, the rising accents used by speakers of both varieties can be adequately captured with one symbolic representation.

When comparing prenuclear and nuclear accents, by contrast, differences were found which could be interpreted as discrete. Whereas nuclear accent peaks were anchored to the intervocalic consonant, prenuclear accent peaks were anchored to the following unstressed vowel. This *anchor shift* could clearly be observed both in the acoustic and articulatory records, reflecting a difference at the symbolic level, possibly in terms of an additional tone following the LH complex.

1. Introduction

Sources of variation in tonal alignment have received a great deal of attention in recent years. The majority of studies have concentrated on the position of L and H turning points in rising pitch accents (often transcribed as L+H* or L*+H) in relation to segmental landmarks which can be identified in the acoustic record. The alignment of tones with segments or segment boundaries has been referred to as segmental anchoring, a concept dealt with in Section 1.1. As more languages are investigated, revealing contextual variation of both a phonetic and a phonological nature, difficulties have arisen with making generalisations about the alignment of tones in relation to specific acoustic segments. A number of studies have explored whether tonal alignment variation is better accounted for in relation to articulatory events, such as raising of the tongue tip to form an alveolar closure. These will be discussed in Section 1.2.

1.1. Segmental anchoring

According to the segmental anchoring hypothesis tones are aligned with, or anchored to, acoustically defined segments, such as the beginning of the onset consonant of an accented syllable, the CV boundary in that syllable, or even a CV boundary in a following syllable. The segmental anchoring hypothesis predicts that both the beginning of the rise (the L tone) and the end of the rise (H tone) are aligned with specific *anchor* segments in such a way that this alignment is unaffected by the number of segments

*Corresponding author. Tel.: +49 221 470 4256; fax: +49 221 470 5938.
E-mail address: doris.muecke@uni-koeln.de (D. Mücke).

0095-4470/$ - see front matter
doi:10.1016/j.wocn.2009.03.005

intervening between the two anchors. Furthermore, since alignment is related to specific segments, it is also unaffected by the articulation rate at which the segmental string is produced.

The most important implication of the segmental anchoring hypothesis for intonational phonology is that a dynamic pitch pattern is anchored in two places, thus supporting a levels-based rather than configurations-based approach to intonation. More specifically, it supports autosegmental-metrical models of intonation in which a rise is composed of two events, a low tone (L) and a high tone (H). However, work on other languages which has been inspired by Arvaniti, Ladd, and Mennen (1998), Arvaniti and Ladd (1995) and Ladd, Faulkner, Faulkner, and Schepman (1999) has often revealed that it is only one of the two tones in a rise which is anchored. In particular, the H tone in a LH contour has been shown to be less stable (more susceptible to contextual influences) than the L tone, for Dutch (Ladd, Mennen, & Schepman, 2000) and Spanish (Prieto & Torreira, 2007) as well as Neapolitan Italian (D'Imperio, Petrone, & Nguyen, 2007a). Furthermore, Welby and Lœvenbruck (2006) go so far as to say that the segmental anchor hypothesis might be valid for some languages, such as Greek and German, but not for others, such as French, where they found a higher degree of variability for late rises, but no systematic source for this variability. They suggest that – in some non-tone languages – tones are anchored to a region of possible timing values ("anchorage") rather than to a specific segment or syllable ("anchor").

In what follows, particular attention will be paid to studies investigating the effects of syllable structure, dialectal background and accent status in the intonational hierarchy, all of which are investigated directly in the experiments reported on in this paper. Effects of articulation rate (Caspers & van Heuven, 1993; Igarashi, 2004; Ladd et al., 1999) will not be dealt with here. For reviews of other factors affecting alignment see Wichmann, House, and Rietveld (2000), and for a general overview see Ladd (2008).

1.1.1. *Effects of dialectal background*

For German, Atterer and Ladd (2004) showed that Southern German accents are aligned consistently later than Northern ones, although the differences they found were small. They concluded that these differences should not lead to different phonological associations across the varieties, but rather should be seen as phonetic detail. In a recent experiment (Mücke, Grice, Hermes, & Becker, 2008), the Atterer & Ladd study was replicated for another Southern variety (Vienna) and a Northern one (Düsseldorf). In this study similar differences were found between Southern and Northern varieties for neutral prenuclear LH accents. These were the first accents in the phrase. Both Atterer & Ladd's and our own results are summarised in Fig. 1, providing the alignment patterns of LH accents in German varieties in prenuclear position. Even though all H peaks co-occur with the vowel of the unstressed syllable (V2), peaks were later in Southern German than in Northern German (Atterer & Ladd, 2004) and, as expected, later in the Vienna variety than in the Düsseldorf variety (Mücke et al., 2008). In both Atterer and Ladd's and our own study, syllable structure was kept constant, so that no information about the effect of syllable structure could be obtained.

In a further study on the peak alignment in initial rises in German, Braun (2007) investigated the effect of dialectal background (Münster and Munich), this time in relation to information structure (contrastive and non-contrastive themes). In non-contrastive themes, she confirmed the findings of Atterer & Ladd and of our own study that Southern peaks occur later than Northern ones. However, in contrastive themes, she could not confirm a later alignment for the Southern variety (Munich). Moreover, the Northern variety (Münster) used delayed peaks to compensate for peak height (as described in Gussenhoven, 2002; Braun & Ladd, 2003), while the Southerners did not. She concluded that delayed peaks are used to express information structural content only in Northern varieties: "It seems that Southerners, whose peaks are already late in a neutral setting, cannot delay the peak further to

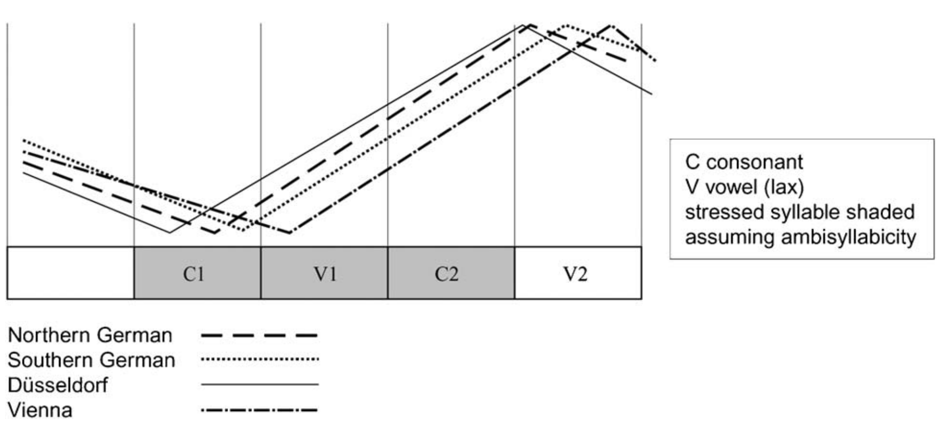


Fig. 1. Stylised summary of alignment properties of prenuclear rising accents in German. Adapted from Atterer and Ladd (2004); Vienna and Düsseldorf added by Mücke et al. (2008).

express contrast, while Northerners can." (Braun, 2007, p. 964).

1.1.2. Effects of syllable structure

In Dutch, f0 peaks were found to align with the right edge of the accented syllable rather than with a specific segment, i.e. at the end of the vowel in open syllables (CV:) and at the end of the coda consonant in closed syllables (CVC). Dutch peak alignment is thus affected by phonological vowel length (Ladd et al., 2000): if the vowel in the accented syllable is phonologically long, the peak occurs late in the vowel; if the vowel is phonologically short, the peak occurs late in the coda consonant. These results indicate that tones might be aligned with the edges of syllables, as also shown for Chinese (Xu, 1998, 2002). Moreover, since Dutch high vowels have the same phonetic duration regardless of whether they are phonologically long (/i:/) or short (/ɪ/), it is not the actual duration of the vowel which affects alignment. The alignment is thus not simply affected by time pressure (Caspers & van Heuven, 1993) or insufficient rise time (Xu & Xu, 2005) from the peak to the end of the phrase.

In a large corpus-based study, Möbius and Jilka (2007) and Jilka and Möbius (2007) investigated the influence of vowel tenseness (corresponding to phonological vowel length) on peak alignment in German falling pitch accents, H*L. They measured the f0 peak position as a percentage of the entire syllable duration, and found a systematic difference in peak alignment between syllables containing a lax (phonological short) vowel and a tense (phonological long) vowel. However, the occurrence of lax (short) vowels was restricted to open syllables in their corpus. They concluded that, instead of tenseness, it was syllable structure (open/closed) that was relevant for peak position.

Similar effects, but comparing CV.CV vs. CVC.CV syllable types, were reported for Neapolitan Italian (D'Imperio, 2000, D'Imperio et al., 2007a). In rising question pitch accents (L* + H), H peaks co-occured with the stressed vowel in open syllables such as *nono* and with the coda consonant in closed syllables such as *nonno,* such that f0 targets have different segmental anchors, depending on whether they are in open or closed syllables. Furthermore, here too, the H appears to align with the end of the accented syllable.

Effects of syllable structure were also found in Spanish. Prieto and Torreira (2007) found later peaks in closed syllables compared to open syllables. While in CV.CV (open syllables) the peak was placed at the beginning of the intervocalic consonant, C2, in CVN.CV (closed syllables) it was placed around the middle of the sonorant coda. In this case, there was an effect of syllable structure, but no clear alignment with a syllable edge.

1.1.3. Effects of accent status in the intonational hierarchy

In a study of Dutch, Schepman, Lickley, and Ladd (2006) confirmed earlier results obtained by Silverman (1987) and Silverman and Pierrehumbert (1990) for English, showing that nuclear H peaks are aligned earlier than prenuclear ones. They argue that this difference in timing is due to tonal repulsion, or a leftwards push, from an L tone following the H – a phrase accent. Schepman et al. suggest that the phrase accent has a secondary association (as proposed in Grice, Ladd, & Arvaniti, 2000), resulting in a LHL structure associated with the nuclear syllable.

A similar timing difference between nuclear and prenuclear accents was also found for Mexican Spanish (Prieto, van Santen, & Hirschberg, 1995), as well as for German (Möbius & Jilka, 2007; investigated for falling accents H*L).

1.2. Articulatory anchoring

Recent work has reported a close alignment of f0 valleys and peaks with articulatory movements, suggesting that tones are coordinated with dynamically defined speech gestures, or articulatory gestures. In a number of languages evidence for this coordination has been found by examining the co-occurrence of f0 events with landmarks in the kinematic contours of the tongue tip and lips.

D'Imperio, Nguyen, and Munhall (2003) and D'Imperio et al. (2007b) found a high co-occurrence of f0 peaks and landmarks corresponding to the lowering and raising of the tongue tip and opening and closing of the lips. They found smaller latencies between the f0 peak and articulatory anchors than between the peak and acoustic anchors. In nuclear question rises (L* + H), f0 peaks co-occurred with the maximum closure (e.g. maximum lip closure for [p] in 'mapa'), and in nuclear statement rises (L + H*) they co-occurred with the peak velocity of the closure movement for the same consonant. However, despite the fact that there was a shorter latency between f0 peaks and articulatory anchors, these anchors were as variable in their timing as the acoustic anchors. That is, they were unable to confirm their hypothesis that articulatory anchors are more stable than acoustic ones.

Smaller latencies between f0 peaks and articulatory anchors were also found in German by Mücke, Grice, Becker, Hermes, and Baumann (2006) and Mücke & Grice (2006).E.g. in prenuclear rises H was more closely aligned with the vocalic target (tongue body raising during the production of the unstressed vowel, [i]) and the transvocalic target (tongue tip lowering and lower lip opening during [i]) than with the acoustic landmark (onset of the vowel).[1] However, they did not investigate whether articulatory anchors were more stable than acoustic ones, in that they did not look at variance homogeneity within each condition.

A further study on Catalan, Prieto, Mücke, Becker, and Grice (2007), found a similar pattern of shorter latencies

[1]They also found that the transvocalic and the vocalic targets were well synchronised in symmetrical CVC contacts, where both C's are identical (in line with Browman & Goldstein, 1988 for English).

for articulatory anchors, this time in both rising and falling accents. Again, stability was not compared across the anchor types.

In the present study we compare acoustic and articulatory anchors, although our primary aim is not to compare anchor types, but rather to investigate a number of *effects* on the alignment of f0 peaks with these anchors. We concentrate on the effects discussed in Section 1.1 above: dialectal background, syllable structure and accent status.

2. Experiment 1: Prenuclear accents

2.1. Method

2.1.1. Speakers

We recorded two native speakers for each of the varieties under investigation: for Düsseldorf, a male (Gu) and a female speaker (Gi), both students in their mid-20s who had spent their first 20 years in the low Franconian area near to Düsseldorf; for Viennese, two female speakers, one student (JS) and one make-up artist (JR), both in their mid-20s, and both of whom grew up in Vienna.

2.1.2. Speech materials

Meaningful sentences were constructed in which the target word was designed to carry the first accent of the phrase, i.e. a prenuclear accent. Rising (L + H) accents were obtained by placing the target word in a contrastive context, following Braun and Ladd (2003), who showed that prenuclear contrastive topic accents typically involve rises. Eight target words were constructed, varying phonological vowel length and therefore syllable structure (CV: .CV vs. CVCV) and place of articulation in the consonants (labial, alveolar) as well as vowel height of the first vowel [a, ɛ]. Each target word was placed in a sequence as follows, e.g. 'Nahni nah(m)' [na:ni na:], 'Nahni took'[2]:

C1 V1 C2 V2 C3 V3, (1)

C = *nasal* [n, m]
V1, 3 = *either low or mid-low vowel* [a, ɛ]
V2 = *high vowel* [i]

V1 contained either a phonologically long or short vowel. Since in German, short vowels do not usually occur in open syllables but in closed syllables, ambisyllabicity is assumed for the intervocalic consonant following the short vowel. Even though the syllable affiliation of an intervocalic consonant following a short vowel is controversial in German phonology, psycholinguistic research in Dutch has shown that speakers generally tend to close syllables with a short vowel (investigated for bisyllabic nouns; Schiller, Meyer, & Levelt, 1997).

[2]Orthographic 'h' in German serves to indicate vowel length, e.g. Nahni [na:ni].

Sentence pairs were designed so that the target words would be realised as contrastive topics, as in (2a) and (2b), where Nanni is contrasted with Nahni:

Mit der Nanni nahm sie den SECHS-Uhr-Bus.
-*With the Nanni took she the SIX o'clock-bus.*
(-*With Nanni she took the SIX o'clock-bus.*) (2a)

Mit der Nahni nahm sie den EIN-Uhr-Bus.
-*With the Nahni took she the ONE o'clock-bus.*
(-*With Nahni she took the ONE o'clock-bus.*) (2b)

However, this contrast does not mean that the focus is on this constituent. The focus of the sentence (and therefore the nuclear accent) is on the last constituent of the sentence, e.g. 'den SECHS Uhr Bus' (the six o'clock bus), which is in turn contrasted with 'den EIN Uhr Bus' (the one o'clock bus). A list of sentence types is in the Appendix.

Since the analysis involved articulatory recordings, the recording time had to be kept to a minimum, precluding the use of fillers. Three hundred and twenty stimuli were recorded in total (4 speakers × 8 target sentences × 10 repetitions). Since the accents were contrastive, the lack of fillers was expected to support the contrast, if it had any effect at all. Speakers were instructed to speak at a rate which they considered to be normal. For a separate study (the results of which are reported elsewhere, Mücke et al., 2006), the same set of data were recorded at a fast rate. The fast rate tokens were recorded in the second half of the session for all speakers and should not therefore have affected the recordings made for this study.

2.1.3. Recordings

All recordings took place at the IfL-Phonetics laboratory in Cologne (Institut für Linguistik). Speech materials were displayed on a computer monitor. Target words were produced in pseudo-randomised order with 10 repetitions each. Acoustic and physiological data were recorded simultaneously. The acoustic signal was recorded with a DAT-recorder (TASCAM DA-P1) using a condenser microphone (AKG C420 head set) and digitised at 44.1 kHz/16 bit. The physiological data were recorded with a 2D Electromagnetic Midsagittal Articulograph (EMMA; Carstens AG100 with 5 channels). Sensors were placed at the mid-point of the lower lip (on the vermilion border) and the tongue blade (1 cm behind the tip) for capturing the movements of the primary constrictors. Additionally, one sensor was placed on the tongue body (4 cm behind the tip) for capturing the vowel-to-vowel articulation, and two reference sensors (on the bridge of the nose and on the gums of the upper incisors for calculating dynamic helmet corrections). An acoustic signal was used to allow for synchronisation of the high-quality acoustic and EMMA signals. At the end of each recording session a bite plate was used to rotate the occlusal plane. All physiological data were sampled at 400 Hz, downsampled to 200 Hz and smoothed with a low-pass filter at 40 Hz (MPC software). For displaying and labelling data, all acoustic and physiological data were converted to SSFF-format to enable the data to be analysed and annotated in the EMU Speech Database System (Cassidy & Harrington, 2001).[3]

2.1.4. Labelling procedures

All data were displayed and labelled by hand in EMU. Articulatory landmarks were placed without access to the f0 contour, and labelling of f0 was carried out independently—only after labelling was complete were the articulatory and f0 labels brought together.

F0 landmarks

F0 time stamps were extracted with a 7.5 ms correlation window and a 3 ms frame spacing and displayed for hand labelling. Local turning points were identified for the rising and rising–falling pitch accents in the vicinity of the target words. The f0 minimum and maximum were annotated at the start and the end of the rise (LH). If no local turning point could be identified in the contour, which was often the case for H, a clear change of the slope from a steep rise to a plateau or shoulder was used as the point in time where the rise starts or ends. In cases where the rise contour was too shallow, and therefore no reliable start or end of the rise could be identified at all, no f0 label was placed. The following f0 labels were identified:

L: Low valley at the beginning of the rise.
H: High peak at the end of the rise.

Segmental landmarks

Segmental boundaries were identified in the acoustic waveform. For annotation, an oscillogram and wide-band spectrogram were displayed simultaneously. Segmental boundaries for the combinations of nasals and vowels were identified at the abrupt change in the spectra of the nasals at the time the oral closure was formed or released. The following boundaries for segmental landmarks were identified:

C1onset Start of the onset consonant in the accented syllable.
V1onset Start of the vowel in the accented syllable.
C2onset Start of the consonant following the vowel in the accented syllable.
V2onset Start of the vowel in the syllable following the accented one.
V2offset End of the vowel in the syllable following the accented one.

Articulatory landmarks

Articulatory landmark events were identified solely for movements in the vertical plane (y-position), capturing the time of maximum constriction or openness of the vocal tract through consonant–vowel articulation. Local maxima and minima in the articulatory trajectory were labelled at the respective zero-crossings in the velocity curve. Peak velocities of the movements were labelled at respective zero-crossings in the acceleration curve. Targets and peak velocities were labelled for labial and lingual movements corresponding to consonantal, vocalic and transvocalic gestures (Browman & Goldstein, 1988, Byrd, Kaun, Narayanan, & Saltzman, 2000).

Consonantal gestures: The constriction maximum (lower lip and tongue tip) in C1 and C2 and the respective peak velocities to form and release the oral closure.
Vocalic gestures: Maximum targets for the V1-to-V2 ([a] or [ɛ] to [i]) tongue body raising movement.
Transvocalic gestures: The constriction minimum (lower lip and tongue tip) in V1 and V2.

The following specific landmarks are used for evaluation and discussion in the present study.

Lipmax2/ Tipmax2	Consonantal target in C2, i.e. the constriction maximum of the lower lip or tongue tip.
Lipmin2/ Tipmin2	Transvocalic target in V2, i.e. the constriction minimum of the lower lip or tongue tip.

The annotation of transvocalic targets is motivated by results from previous work in which it was shown for similar speech materials that transvocalic and vocalic targets coincide in symmetrical nasal–vowel–nasal production (Becker, 2005; Mücke et al., 2006).

Fig. 2 shows measurement points for segmental labels, f0 labels and kinematic labels (constriction maximum in the vertical position). Examples of labels are given for the opening and closing movements of the primary constrictor tongue tip for the test word Nahni [na:ni].

2.2. Results

A total of 317 stimuli were included into the statistical analysis: for the Düsseldorf speakers, all tokens were usable for speaker Gi (80) and Gu (80). For the Vienna speakers, one token was eliminated from speaker JR (79) and two from speaker JS (78).

Latencies of H were calculated relative to the nearest segmental and articulatory anchors in the accented or unaccented syllable. That is, in a token such as ['na:ni], the nearby line-up point for the H peak would be in the unstressed syllable [ni]. This is the syllable on which the peak occurs.

Section 2.2.1 reports on acoustic H anchors and Section 2.2.2 on articulatory H anchors. Two types of

[3](a) EMU Speech Database System; URL: http://emu.sourceforge.net (b) MPC software and EMMA2SSFF converter; URL: http://www.uni-koeln.de/phil-fak/phonetik

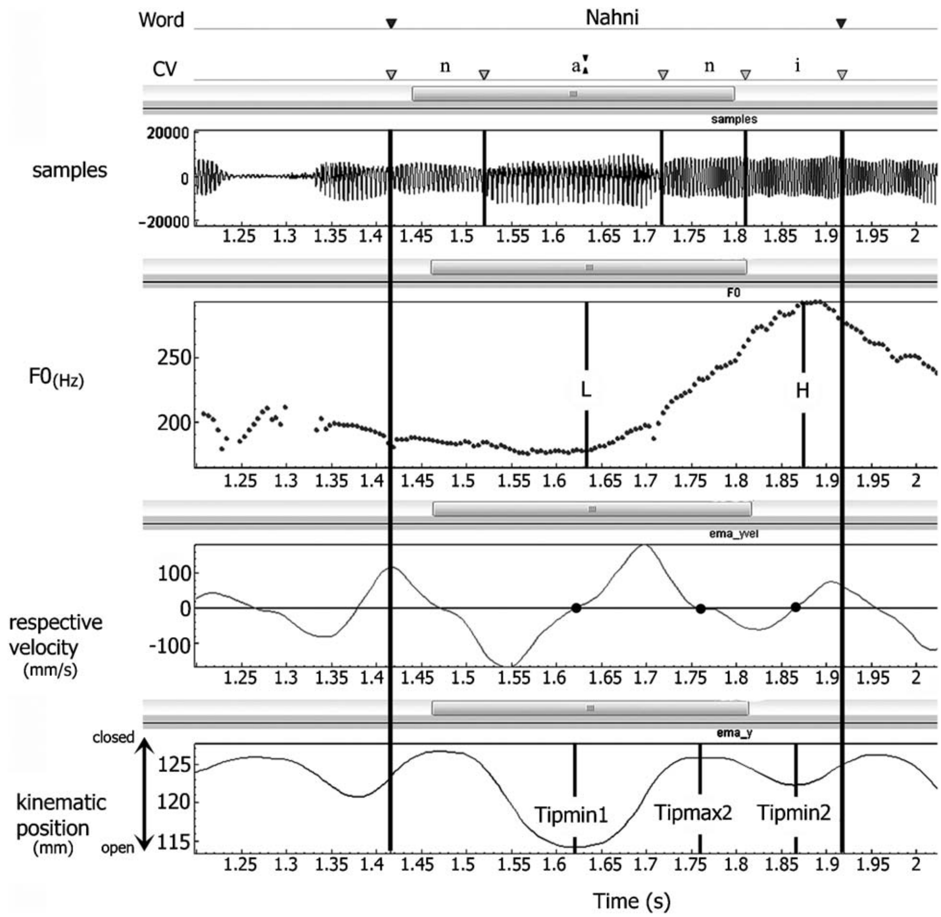


Fig. 2. Examples of landmarks for speaker JR, prenuclear accent. From top to bottom: acoustic waveform, f0 contour, tongue tip velocity trace and tongue tip position for vertical movements.

analysis were carried out:

(i) *Across-speaker comparison*: Overall ANOVAs were performed across all speakers. The entailed repeated measures ANOVAs based on the cell means for each speaker. The between-subject factor was Dialectal Background (Vienna/Düsseldorf) and within-subject factors were Syllable Structure (open/closed) and Place of Articulation (labial/alveolar). Speaker was treated as the random factor, with dependent variables as H-V2onset (acoustic anchor) or H-Lip/Tipmin2 (articulatory anchor).
(ii) *Within-speaker comparison*: ANOVAs were conducted for each speaker separately to describe the speaker-dependent strategies. Therefore, two-way ANOVAs (2×2) were calculated, including Syllable Structure (open/closed) and Place of Articulation (labial/alveolar) as the independent variables, and H-V2onset (acoustic anchor) and H-Lip/Tipmin2 (articulatory anchor) as the dependent variables.

In Section 2.2.3 the two anchor types were compared directly (acoustic and articulatory) to test for their variability and vicinity to H.

2.2.1. Acoustic anchors (prenuclear accents)

In both varieties, H co-occurred with the vowel in the unstressed syllable, V2. The boxplots in Fig. 3 provide latencies for H relative to the beginning (left figure) and the end (right figure) of V2 for all speakers and for open and closed syllables separately. The zero line marks the beginning (left figure) and the end (right figure) of V2. Positive values indicate that the H peak occurs after the V2 landmark; negative values indicate that H occurs before it. All latencies for H relative to the beginning and the end of V2 were rather large in both varieties (on average 71 ms in Vienna and 38 ms in Düsseldorf relative to the beginning).

When comparing the two dialect groups, Vienna and Düsseldorf, peaks tended to be later in the Vienna group. In Vienna, H occurred *on average 33 ms later* when measuring the distance between H and the beginning of V2. However, in the repeated measures ANOVA (Dialectal Background [$F(1, 317) = 13.776$, $p = 0.06$ ns]), that difference marginally failed to reach significance.

Furthermore, the effect of Syllable Structure [$F(1, 317) = 2.3795$, $p = 0.22$ ns] did not reach significance in the repeated measures ANOVA. This finding is confirmed by the individual analysis: three out of four speakers show no effect of Syllable Structure ($p > 0.05$), e.g. speaker Gu aligned H peaks 42 ms after the V2onset in open syllables

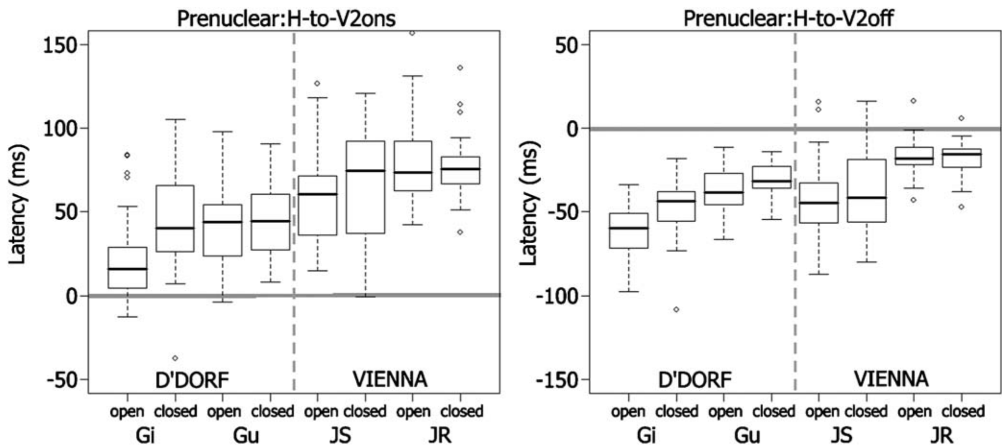


Fig. 3. Latencies for the H peak relative to the beginning (left figure) and end (right figure) of the unstressed vowel, V2. Acoustic anchors, prenuclear data.

and 46 ms in closed syllables. Only the Düsseldorf speaker Gi shows the expected Syllable Structure effects [$F(1, 80) = 25.878$, $p < 0.001$]. She aligns H peaks on average 21 ms after the onset of V2 in open syllables and 45 ms in closed syllables (later peaks in closed syllables).

The main effect Place of Articulation was significant in the repeated measures ANOVA [$F(1, 317) = 32.188$, $p < 0.05$], and did not interact with Syllable Structure [$F(1, 317) = 1.3776$, $p = 0.32$ ns]. Compared to target words involving a labial closure, all peaks were later in target words involving an alveolar closure (on average 32 ms later in the alveolar condition). This finding is also confirmed by the individual analysis (Place of Articulation, $p < 0.001$): all speakers placed peaks later in the alveolar condition when H is measured relative to the onset of V2 (speaker Gi, $\Delta = 35$ ms; speaker Gu, $\Delta = 34$ ms; speaker JS, $\Delta = 44$ ms; speaker JR, $\Delta = 17$ ms).

2.2.2. *Articulatory anchors (prenuclear accents)*

Latencies for H were calculated relative to the nearest landmark in the articulatory dimension for prenuclear accents. This was the lowering of the primary constrictor in the unstressed syllable (lower lip or tongue tip). Fig. 4 shows latencies for H relative to the lower lip minimum (left figure) and tongue tip minimum (right figure) during the production of the unstressed vowel, V2. H was closely phased with the lower lip minimum (on average 11 ms for Vienna and −16 ms for Düsseldorf) and with the tongue tip minimum (on average 21 ms for Vienna and 4 ms for Düsseldorf). H peaks were closely phased with articulatory targets, such that the alignment latencies converged around zero.

When comparing the dialect groups, H was timed slightly later (on average 10 ms) in the Vienna group than in the Düsseldorf group (differences were smaller than in the acoustic measure). However, the differences between the groups did not reach significance in the repeated measures ANOVA (Dialectal Background [$F(1, 317) = 4.3694$, $p = 0.17$ ns]).

In the repeated measures ANOVA across all speakers, no effect of Syllable Structure was found on the articulatory anchor type (H-Lipmin2 and H-Tipmin2; [$F(1, 317) = 0.9111$, $p = 0.41$ ns]). There were no later peaks in closed syllables compared to open syllables when comparing all speakers. This result is reflected by the individual analysis: three out of four speakers show no reliable effect of Syllable Structure ($p > 0.05$) on the peak placement, e.g. speaker Gu aligns H peaks on average 1 ms after the gestural target in open syllables and 5 ms after the target in closed syllables (labial and alveolar consonants together). Only the Düsseldorf speaker, Gi, shows effects of peak placement induced by Syllable Structure [$F(1, 80) = 18.527$, $p < 0.001$]. She aligns peaks on average 23 ms before the gestural target in open syllables, and 7 ms after the target in closed syllables (labial and alveolar consonants together).

Contrary to the acoustic measure, the factor Place of Articulation did not reach significance in the repeated measures ANOVA [$F(1, 317) = 4.4999$, $p = 0.41$ ns]. However, there was an alignment difference of, on average, 15 ms across all speakers (with later peaks in the alveolar condition). The individual analysis revealed that three out of four speakers showed the same tendencies: for those three speakers the main factor Place of Articulation ($p < 0.001$) reached significance with later alignment in the alveolar condition (speaker Gi, $\Delta = 18$ ms; speaker Gu, $\Delta = 24$ ms; speaker JS, $\Delta = 29$ ms). Only speaker JR showed no systematic effect between target words involving a labial and alveolar closure (speaker JR, $\Delta = 5$ ms, $p > 0.05$). For none of the speakers, an interaction between the main factors Place of Articulation and Syllable Structure was found ($p > 0.05$).

2.2.3. *Acoustic and articulatory anchors compared (prenuclear accents)*

The acoustic and articulatory anchor types were compared directly. Means and standard deviations (in parenthesis) are reported in Table 1.

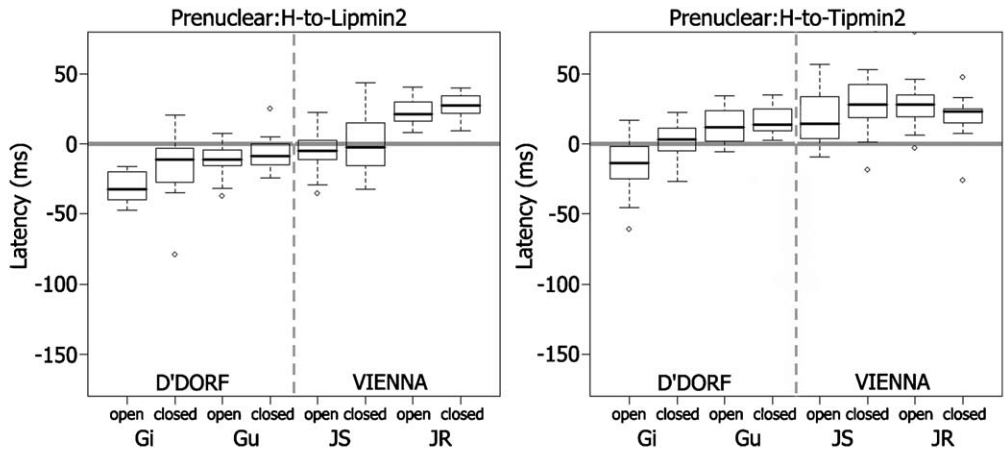


Fig. 4. Latencies for the H peak relative to the minimum of the lower lip (left figure) and tongue tip (right figure) during the production of the unstressed vowel, V2. Articulatory anchors, prenuclear data.

Table 1
Mean alignment latencies (ms) with standard deviations in parenthesis, *p*-values (comparison of the means and comparison of the standard deviations) for pooled acoustic (H-V2onset) and articulatory data (H-Lipmin2 and H-Tipmin2).

Prenuclear	Bilabial				Alveolar			
	H-segment (H-V2ons)	H-gesture (H-Lipmin2)	*p*-Value (mean)	*p*-Value (sd)	H-segment (H-V2ons)	H-gesture (H-Tipmin2)	*p*-Value (mean)	*p*-Value (sd)
Düsseldorf								
Gi open	8 (10)	−31 (10)	***	ns	34 (28)	−14 (20)	***	ns
Gi closed	23 (20)	−16 (29)	***	ns	66 (22)	2 (12)	***	0.018
Gu open	25 (14)	−11 (11)	***	ns	58 (16)	13 (13)	***	ns
Gu closed	28 (13)	−7 (11)	***	ns	62 (13)	16 (10)	***	ns
Grand mean	21 (16)	−16 (16)			55 (24)	4 (18)		
Vienna								
JS open	39 (17)	−5 (15)	***	ns	80 (24)	19 (19)	***	ns
JS closed	45 (27)	1 (21)	***	ns	91 (19)	28 (18)	***	ns
JR open	70 (17)	23 (10)	***	0.03	88 (26)	29 (17)	***	ns
JR closed	69 (13)	26 (9)	***	ns	84 (18)	10 (49)	***	***
Grand mean	56 (23)	11 (20)			86 (22)	21 (30)		

Prenuclear data.

First, we compared the means for the pooled acoustic and articulatory alignment values. We calculated *p*-values in an ANOVA for each data set. In the articulatory dimension, latencies were relatively small: H occurred 6 ms before the gestural anchor in Düsseldorf and 16 ms after it in Vienna. In the acoustic dimension, by contrast, latencies were greater ($p<0.001$): H was 38 ms after the acoustic anchor in Düsseldorf and 71 ms after it in Vienna (one exception to this pattern was found for one Düsseldorf speaker (Gi), but only in the open syllable condition, where latencies were smaller in the acoustic dimension).

Then the standard deviations for the pooled acoustic and articulatory data were compared by calculating *F*-tests (respective *p*-values are reported in Table 1). The *F*-values reached significance only in 3 out of 16 conditions (with $p>0.05$ ns). No systematic difference in dispersion was found between pooled acoustic and articulatory anchor types.

2.3. Discussion

2.3.1. Acoustic anchors (prenuclear accents)

Fig. 5 summarises the findings for the acoustic data. The figure is based on statistical means of one speaker per variety. In all conditions (dialectal background and syllable structure), H was aligned with the vowel of the unstressed syllable (nearest anchor in the acoustic dimension).

In line with Atterer and Ladd (2004), the Southern variety (Vienna) had later peaks than the Northern one (Düsseldorf). In fact, as in their study, the difference across the dialectal groups appeared to be gradient, with an overlap between the two groups. The overlap is probably due to speaker-dependent strategies since, e.g. in the prenuclear accents the Düsseldorf speaker Gu and the two Viennese speakers show similar alignment values.

Furthermore, since contrastive accents were investigated in this study, our results are in line with those reported by

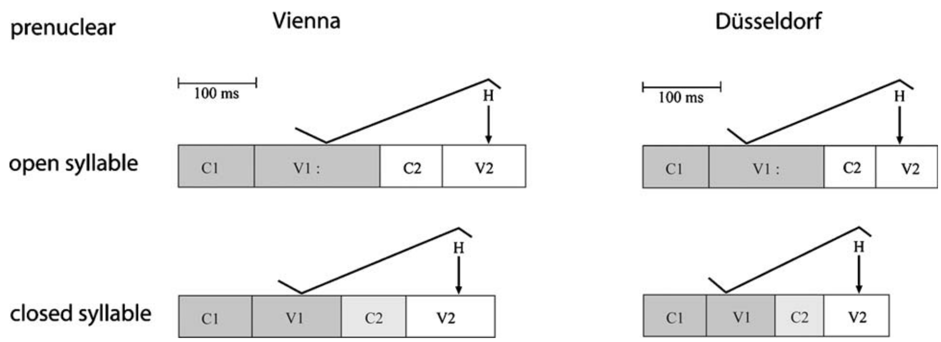


Fig. 5. Prenuclear accents: acoustic alignment patterns in open and closed accented syllables in the varieties of Vienna (speaker JR) and Düsseldorf (speaker GU). Figure to scale.

Braun (2007), discussed in Section 1.1 above. Recall that in her study, the effect of dialectal background across Northern and Southern speakers was present only in non-contrastive accents. Her explanation was that contrastive accents are later than non-contrastive ones, and that dialectal background was partly overridden by contrastive context (especially in the Northern variety, where the contrastive peaks begin to resemble the Southern ones). Although in our previous work we found significant differences between Vienna and Düsseldorf speakers in neutral, non-contrastive contexts (Mücke et al., 2008), here the tendency for Southern peaks to be later did not reach significance.

In line with Greek (Arvaniti et al., 1998), and unlike Dutch (Schepman et al., 2006), Spanish (Prieto & Torreira, 2007) or Neapolitan Italian (D'Imperio et al., 2007a), no consistent effect of syllable structure was found for three out of four speakers. That is, regardless of the affiliation of the intervocalic consonant [n], the H peak aligns with the vowel in the unstressed syllable for all speakers except for the Düsseldorf speaker Gi. However, it is not possible to draw a conclusion whether the syllable structure present in the data of Gi is due to speaker-dependent strategies or to the dialectal background.

The influence of manner of articulation in the coda of the accented syllable has been reported by Prieto and Torreira (2007) for laterals versus nasals in Spanish, and by Jilka and Möbius (2007) and Möbius and Jilka (2007) for sonorants versus obstruents in German. Our data show an influence of place of articulation. Peaks were later in target words containing an alveolar closure compared to those with a labial closure. Since the tongue tip and the tongue body effectively share two articulators (the jaw and the tongue body), whereas the lips only share the jaw with the other two, a higher degree of coarticulation can be expected for the coronal consonants.

2.3.2. *Articulatory anchors (prenuclear accents)*

In the articulatory dimension, the nearest landmark for H was the transvocalic opening during the production of the vowel V2. Fig. 6 summarises the findings for H relative to the movement of the primary constrictor (lower lip). The figure shows averaged contours for the movement of the lower lip in prenuclear accents separately for open and closed syllables. The lip contours are plotted for the vertical position (lip opening and lip closing movements) for the Vienna speakers (above) and the Düsseldorf speakers (below). High displacements indicate that the lips are closed for the nasal production, while low values indicate that the lips are open during the vowels (3 cm displacement range for each panel). The arrows are based on statistical means and mark the timing of the f0 peaks relative to the articulatory trajectories. For all four speakers the H peak appears to be closely phased with the maximal lip opening during V2 production (the transvocalic lip target) in prenuclear accents.

Across the varieties, the articulatory alignment patterns are strikingly similar. H is timed slightly later in the Vienna group than in the Düsseldorf one. The differences between the groups in the articulatory domain are smaller than in the acoustic domain. They are probably due to fine phonetic detail corresponding to different consonant–vowel articulation across the varieties. Since in the Vienna variety vowel durations usually appeared to be greater than in the Düsseldorf ones, there might be a tendency for later alignment strategies in the Vienna group for H relative to the articulatory anchor.

As in the acoustic signal, no systematic effect of syllable structure was observed in the articulatory signal for three out of four speakers. Except for the Düsseldorf speaker Gi (open syllable condition), H converged around zero with the transvocalic minimum in the long and short vowel conditions.

Furthermore, the effect of place of articulation of the intervocalic consonants is weaker in the articulatory than in the acoustic data. In the overall analysis, it failed to reach significance, even though three out of four speakers clearly show the tendencies to align H peaks later in consonants formed with a tongue tip closure compared to those with a labial closure. As pointed out in Section 2.3.1, we interpret those tendencies as being due to different vowel–consonant production (coarticulation), since in the oral system the lips are more independent from the tongue body than the tongue tip (Browman & Goldstein, 1990).

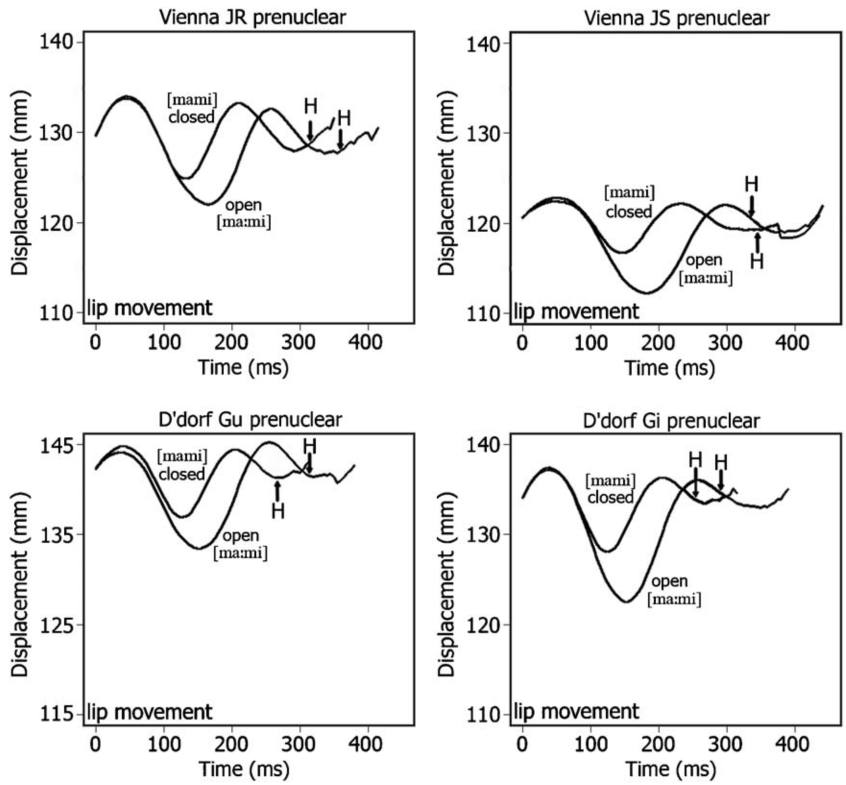


Fig. 6. Averaged contours of vertical lower lip movements for open and closed syllables in the (prenuclear) target words Mahmi and Mammi, separately for Vienna speakers JR and JS (above), and Düsseldorf speakers Gu and Gi (below). Prenuclear accents. The arrows mark the timing of the f0 peaks (H) relative to the articulatory trajectories.

2.3.3. Acoustic and articulatory anchors compared (prenuclear accents)

Even smaller alignment latencies were found in the articulatory dimension than in the acoustics, an *F*-test (pooled acoustic and articulatory data) revealed that the dispersion was the same for both anchor types. These findings (smaller latencies for articulation but similar variability) are in line with D'Imperio et al. (2003, 2007b) for Neapolitan Italian peak alignment and Prieto et al. (2007) for Catalan.

3. Experiment 2: Nuclear accents

We now turn to the results for nuclear accents. At the end of the sections on acoustic anchors (Sections 3.2.1 and 3.3.1) and articulatory anchors (Sections 3.2.2 and 3.3.2) the nuclear anchors will briefly be compared with the prenuclear ones from Experiment 1.

3.1. Method

The same four speakers were recorded as in Experiment 1 (see Section 2.1). The experimental set-up was also identical. For the speech materials the same test words were used, sentences were constructed so as to place the tokens in nuclear position. Again, a contrast was forced on the target word, since Baumann, Grice, and Steindamm (2006) have shown for *nuclear* accents that contrastive accents in declaratives typically involve rises. Mini-dialogues were designed in which the test words would carry contrastive focus, as in the answer (A) in (3). A list of sentences is in the Appendix:

Q : Hat sie die Mammi oder die Mahmi bestohlen?
– *Has she the Mammi or the Mahmi robbed?*
(*Has she robbed Mammi or Mahmi?*)
A : Sie hat die Mahmi bestohlen.
– *She has the Mahmi robbed.*
(*She has robbed Mahmi.*) (3)

Three hundred and twenty stimuli were recorded in total (4 speakers × 8 target sentences × 10 repetitions). Labels were placed as in Experiment 1. In addition to the L and H labels which were placed in the f0 contour for experiment 1, a label was also placed for L2 (since the rise was followed by an immediate fall). L2 was identified at the end of the fall. Fig. 7 shows measurement points for segmental, f0 and kinematic labels. Examples of labels are given for the opening and closing movement of the primary constrictor lower lip for the test word Maehmi [mɛ:mi].

3.2. Results (nuclear accents)

The following set of utterances were included in the statistical analysis: for the Düsseldorf speakers all tokens

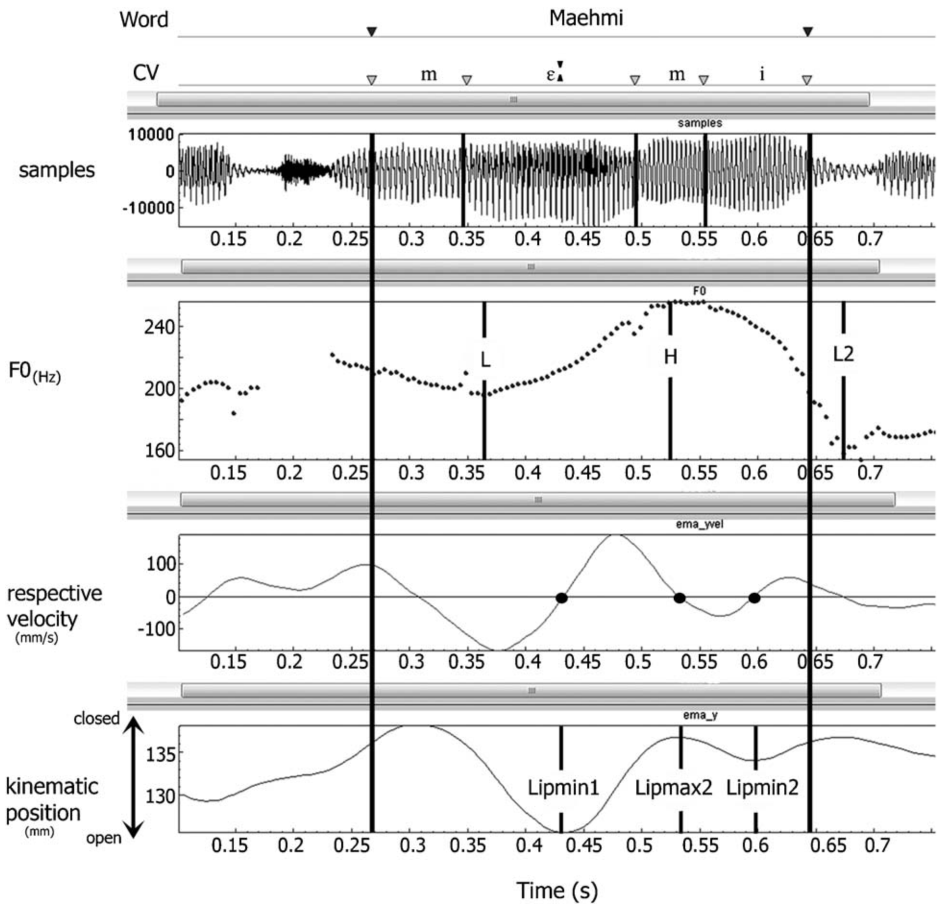


Fig. 7. Examples of landmarks for speaker Gi, nuclear accent. From top to bottom: acoustic waveform, f0 contour, articulatory lower lip velocity trace and lower lip vertical position.

minus 1 were usable for speaker Gi (79) and all tokens minus 27 (mostly rises which were too shallow to label reliably) for speaker Gu (53). For the Vienna speakers all renditions went into the analysis for both speakers JR (80) and JS (80). That makes a total of 292 stimuli.

As in Experiment 1 (Section 2.2), the segmental and articulatory landmarks co-occurring with H were identified. Latencies were calculated for acoustic anchors (f0 turning points relative to segmental boundaries) and for articulatory anchors (f0 turning points relative to articulatory targets).

As described in Section 2.2, two types of ANOVAs were performed for (i) across-speaker comparison and (ii) within-speaker comparison: (i) across all speakers, a repeated measures ANOVA was calculated based on cell means (across all speakers), including the between-subject factor Dialectal Background (Vienna/Düsseldorf), the within-subject factors Syllable Structure (open/closed) and Place of Articulation (labial/alveolar), Speaker as the random factor, and the dependent variables H-V2onset (acoustic anchor) and H-Lip/Tipmin2 (articulatory anchor). (ii) For each speaker separately, two-way (2×2) ANOVAs were conducted, including the independent variables Syllable Structure (open/closed) and Place of Articulation (labial/alveolar), and the dependent variables H-V2onset (acoustic anchor) and H-Lip/Tipmin2 (articulatory anchor).

3.2.1. *Acoustic anchors (nuclear accents)*

In nuclear accents, the landmark for H was the intervocalic consonant, C2. Fig. 8 provides alignment values for H relative to the beginning (left figure) and the end (right figure) of C2. Latencies for H relative to the beginning of the C2 segment were rather large in the Düsseldorf group (on average 13 ms, open and closed syllables together), and even larger in the Vienna group (on average 41 ms).

In terms of dialectal variation, H was aligned early in C2 in the Düsseldorf variety and late in C2 in the Vienna variety; the difference across the varieties amounted to on average 28 ms for H-C2onset (later in Vienna). However, as in the prenuclear accents, the factor Dialectal Background marginally failed to reach significance [$F(1, 292) = 11.388$, $p = 0.07$ ns] in the repeated measures ANOVA.

On the other hand, the factor Syllable Structure reached significance in the repeated measures ANOVA [$F(1, 292) = 34.657$, $p < 0.01$]. In contrast to the prenuclear accents, H

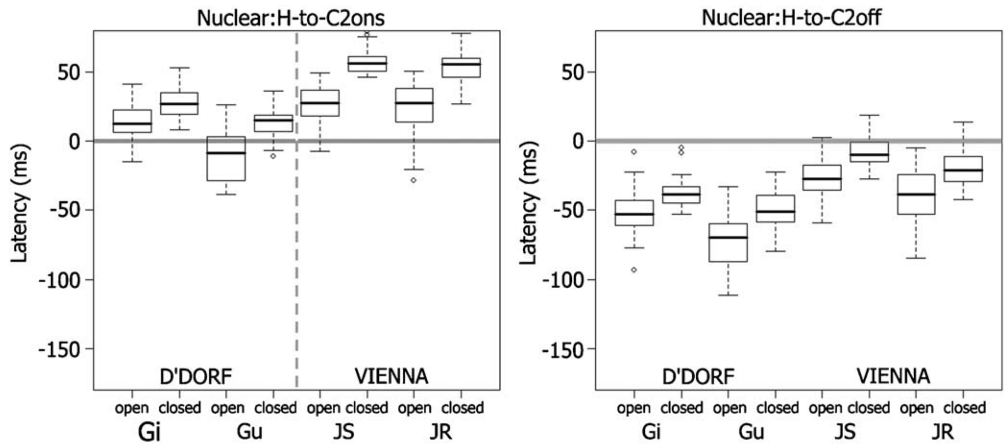


Fig. 8. Latencies for the H peaks relative to the beginning (left) and the end (right figure) of the second target consonant, C2. Acoustic anchors, nuclear data.

was *systematically* aligned later in closed syllables compared to those in open syllables, on average 18 ms later peaks in closed syllables in the Düsseldorf group, and 31 ms in the Vienna group. That finding is confirmed by the individual analysis, where Syllable Structure induced later peaks in closed syllables ($p<0.001$) in the realisations of all of the four speakers (later peaks in closed syllables: speaker Gi, $\Delta = 13$ ms; speaker Gu, $\Delta = 23$ ms; speaker JS, $\Delta = 30$ ms; speaker JR, $\Delta = 32$ ms).

Furthermore, the factor Place of Articulation reached significance [$F(1,\ 292) = 25.866$, $p<0.05$], and did not interact with Syllable Structure. Interestingly, all speakers placed peaks earlier in target words involving a labial closure, and not an alveolar closure (on average 15 ms earlier in the labial condition). The results were confirmed by the individual analysis ($p<0.01$). All speakers employ the same strategy by aligning peaks earlier in the labial condition (later peaks in labial condition: speaker Gi, $\Delta = 9$ ms; speaker Gu, $\Delta = 23$ ms; speaker JS, $\Delta = 8$ ms; speaker JR, $\Delta = 19$ ms).

In order to compare these nuclear anchors to the prenuclear ones from the first experiment, the effect of Accent Status was analysed in the intonational hierarchy on the alignment of H relative to the beginning of the unstressed vowel, V2. In a repeated measures ANOVA (all data) with Speaker as the random factor, the within-subject factor Accent Status reached significance [$F(1,\ 609) = 147.58$, $p<0.01$]. All speakers align peaks (H-V2onset) 90 ms earlier in nuclear accents compared to prenuclear accents. This finding is confirmed by the individual analysis (one-way ANOVAs conducted for each speaker and measure separately). The independent variable Accent Status was highly significant ($p<0.001$) for each speaker with earlier peaks in nuclear accents for H peaks relative to the acoustic anchor (speaker Gi, $\Delta = 78$ ms; speaker Gu, $\Delta = 102$ ms; speaker JS, $\Delta = 82$ ms speaker JR, $\Delta = 108$ ms).

3.2.2. *Articulatory anchors (nuclear accents)*

In nuclear accents, the nearest anchor was the maximum closure during the production of the intervocalic target consonant, C2 (maximum lower lip or tongue tip closure). Fig. 9 provides alignment values for H relative to lower lip closure in labial consonants (left figure) and tongue tip closure in alveolar consonants (right figure). The H peaks were closely phased with the consonant target in the Vienna variety (on average 8 ms after the lower lip target and 4 ms before the tongue tip target). In contrast, alignment latencies were rather large in the Düsseldorf variety (on average H occurred 19 ms before the labial closure and 36 ms before the tongue tip closure).

In the dialect comparison, later H peaks were found in the Vienna variety than in the Düsseldorf one (on average 33 ms later in the Vienna group; lower lip and tongue tip data pooled). Again, that difference failed to reach significance in the repeated measures ANOVA [$F(1,\ 292) = 5.4068$, $p = 0.145$ ns].

The factor Syllable Structure reached significance in the repeated measures ANOVA [$F(1, 292) = 51.94$, $p<0.01$]. H peaks were systematically later in closed syllables compared to open syllables (on average 22 ms later, all speakers and both constrictor types together). That result is confirmed by the individual analysis of each speaker ($p<0.001$): All speakers placed peaks later in closed syllables for H relative to the articulatory anchor (speaker Gi, $\Delta = 13$ ms; speaker Gu, $\Delta = 23$ ms; speaker JS, $\Delta = 26$ ms; speaker JR, $\Delta = 24$ ms).

The Place of Articulation factor marginally failed to reach significance in the repeated measures ANOVA [$F(1,\ 292) = 7.0539$, $p = 0.07$ ns]. Furthermore, there was no interaction between Place of Articulation and Syllable Structure. However, all speakers placed peaks on average 16 ms earlier in target words involving an alveolar closure (which was rather surprising, since in the prenuclear data tendencies were found for later and not earlier peaks in the alveolar condition). In the individual analysis,

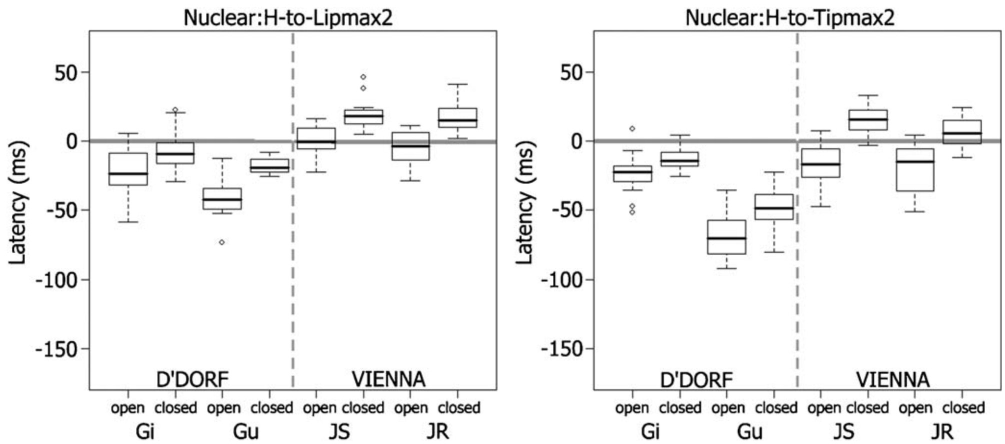


Fig. 9. Latencies for the H peak relative to the maximum of the lower lip (left figure) and tongue tip (right figure) during the production of the second target consonant, C2. Articulatory anchors, nuclear data.

three out of four speakers show the same tendencies: For those speakers the main factor Place of Articulation ($p<0.001$) reached significance with later peaks in labial condition (speaker Gu, $\Delta = 29$ ms; speaker JS, $\Delta = 11$ ms; speaker JR, $\Delta = 13$ ms). No such effect was found ($p = 0.19$) for the Düsseldorf speaker Gi (on average $\Delta = 4$ ms later peaks in labial condition).

To test the effect of Accent Status on H alignment relative to gestural anchors, the anchor corresponding to the production of the unstressed vowel, V2 (the transvocalic minimum: lower lip opening and the tongue tip lowering during V2) was selected. An overall ANOVA (repeated measures ANOVA as in Section 3.2.1, all data) revealed a strong effect of Accent Status on the articulatory H alignment [$F(1, 609) = 198.59$, $p<0.001$]. Peaks were earlier aligned in nuclear accents, on average 121 ms earlier across all speakers. The individual analysis (one-way ANOVAs for each speaker and measure separately, as in Section 3.2.1) revealed the same overall tendencies: all speakers align H peaks considerably earlier relative to the articulatory anchors in nuclear accents (speaker Gi, $\Delta = 114$ ms; speaker Gu, $\Delta = 159$ ms; speaker JS, $\Delta = 110$ ms; speaker JR, $\Delta = 115$ ms).

3.2.3. *Acoustic and articulatory anchors compared (nuclear accents)*

Table 2 provides means and standard deviations for the pooled articulatory and acoustic alignment values.

First the means for the pooled acoustic and articulatory data sets were compared by calculating ANOVAs. In the Vienna variety, the articulatory alignment latencies were *smaller* than acoustic ones ($p<0.001$): While H occurred on average 2 ms after the articulatory C2 target (lower lip and tongue tip data together), H occurred 41 ms after the acoustic boundary of C2 (C2 onset). However, the opposite was found in the Düsseldorf group, where articulatory latencies were *larger* than the acoustic ones with $p<0.001$: While H occurred on average 31 ms before the articulatory C2 target, H occurred 11 ms after the acoustic C2 boundary.[4]

The standard deviations were then compared for the pooled acoustic and articulatory data sets using F-tests. Only in 3 out of 16 conditions the F-values reached significance (with $p>0.5$ ns). As with the prenuclear data, there was no systematic difference in dispersion between the pooled articulatory and acoustic anchors. Both anchor types show a comparable stability.

3.3. *Discussion (nuclear accents)*

3.3.1. *Acoustic anchors (nuclear accents)*

Fig. 10 summarises the findings for the acoustic data (based on statistical means). In nuclear accents H was aligned with the intervocalic consonant, C2. The dashed arrows indicate the anchor shift from prenuclear to nuclear accents, since H peaks occurred with V2 in prenuclear accents (reported above, see Fig. 5), but with C2 in the nuclear accents.

Concerning differences due to dialectal background our results on contrastive nuclear accents are in line with the findings reported by Braun (2007) for prenuclear contrastive accents: Even though there were tendencies for peak placement to be consistently later in Vienna than in the Düsseldorf group, on average 28 ms later in Vienna, there is no clear cut distinction between the dialectal groups.

However, as in Dutch (Ladd et al., 2000), Spanish (Prieto & Torreira, 2007), Neapolitan Italian (D'Imperio et al., 2007a) and (for falling H*L accents) in German (Jilka & Möbius, 2007), strong effects of syllable structure were found. H was consistently later in closed syllables than in open syllables with respect to the nearby landmark, C2, in

[4] It is not clear what is behind this difference. The alignment of *H*, which is early in relation to the articulatory anchor for *C*2, is possibly affected by the proximity of the preceding *L* tone. Differences in competition for alignment between two pitch accent tones, according to pitch accent structure along the lines of Grice (1995), is an issue which will be addressed in further work (Mücke, Nam, Prieto, & Goldstein, accepted).

Table 2
Mean alignment latencies (ms) with standard deviations in parenthesis, *p*-values (comparison of the means and comparison of the standard deviations) for pooled acoustic (H-C2onset) and articulatory data (H-Lipmax2 and H-Tipmax2).

Nuclear	Bilabial				Alveolar			
	H-segment (H-C2ons)	H-gesture (H-Lipmax2)	*p*-Value (mean)	*p*-Value (sd)	H-segment (H-C2ons)	H-gesture (H-Tipmax2)	*p*-Value (mean)	*p*-Value (sd)
Düsseldorf								
Gi open	17 (13)	−22 (15)	***	ns	10 (13)	−23 (13)	***	ns
Gi closed	32 (10)	−7 (13)	***	ns	21 (9)	−13 (8)	***	ns
Gu open	0 (20)	−42 (19)	***	0.0016	−17 (16)	−69 (16)	***	ns
Gu closed	22 (7)	−18 (6)	***	0.009	8 (9)	−47 (14)	***	ns
Grand mean	19 (17)	−21 (19)			3 (19)	−41 (26)		
Vienna								
JS open	36 (8)	0 (10)	***	ns	18 (12)	−17 (14)	***	ns
JS closed	56 (9)	19 (10)	***	ns	58 (8)	15 (10)	***	ns
JR open	35 (11)	−4 (12)	***	ns	10 (20)	−20 (17)	***	0.03
JR closed	61 (10)	17 (10)	***	ns	48 (10)	7 (11)	***	ns
Grand mean	47 (15)	8 (15)			34 (24)	−4 (20)		

Nuclear data.

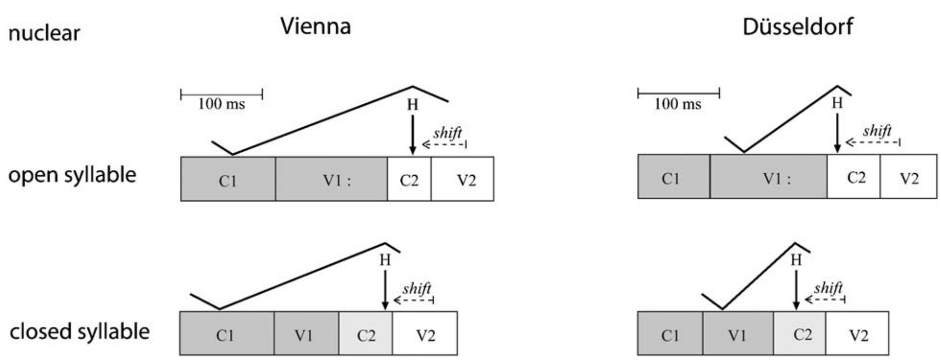


Fig. 10. Nuclear accents: acoustic alignment patterns in open and closed syllables in the varieties Vienna (speaker JR) and Düsseldorf (speaker Gu). Figure to scale.

both varieties (on average 25 ms later in closed syllables, Düsseldorf and Vienna together). In open syllables, H was placed early in C2 in the Düsseldorf variety and around the middle of C2 in Vienna. In closed syllables, H was placed around the middle of C2 in the Düsseldorf variety and late in C2 in Vienna.

Furthermore, later peaks were found in target words containing alveolar closures than labial closures for H relative to the onset of the following vowel (while it was the other way round in prenuclear accents).

Segmental anchor shift: Our acoustic results support the claims made by Silverman and Pierrehumbert (1990) and by Schepman et al. (2006) for other West Germanic languages: for both varieties the f0 peak (H target) is aligned earlier in nuclear position than in prenuclear position. In prenuclear position, H co-occurred with the nucleus of the posttonic syllable and in nuclear position with the intervocalic consonant. In the acoustic dimension, a segmental anchor shift was found. H was aligned with the unstressed vowel, V2, in prenuclear rises and with the intervocalic consonant, C2, in nuclear rises. H was shifted from one segment to the preceding one in the segmental string.

(a) Occurrence of H in prenuclear accents C1 V1 C2|V2.
(b) Occurrence of H in nuclear accents C1 V1|C2 V2. (4)

See Section 3.3.3 for a suggested explanation.

3.3.2. *Articulatory anchors (nuclear accents)*

The nearest landmark for H was the articulatory target of the intervocalic consonant C2 (the consonantal closure). Fig. 11 provides average trajectories for the lower lip movement in the nuclear CV(:)CV target words (opening and closing movements). The arrows mark the timing of H. The H peaks co-occurred with the maximum target of the second closing gesture.

It is clear from the figure that the peaks were, on average, later in the Vienna group than in the Düsseldorf group (33 ms later in Vienna). The differences in dialectal background measured for the articulatory anchor types

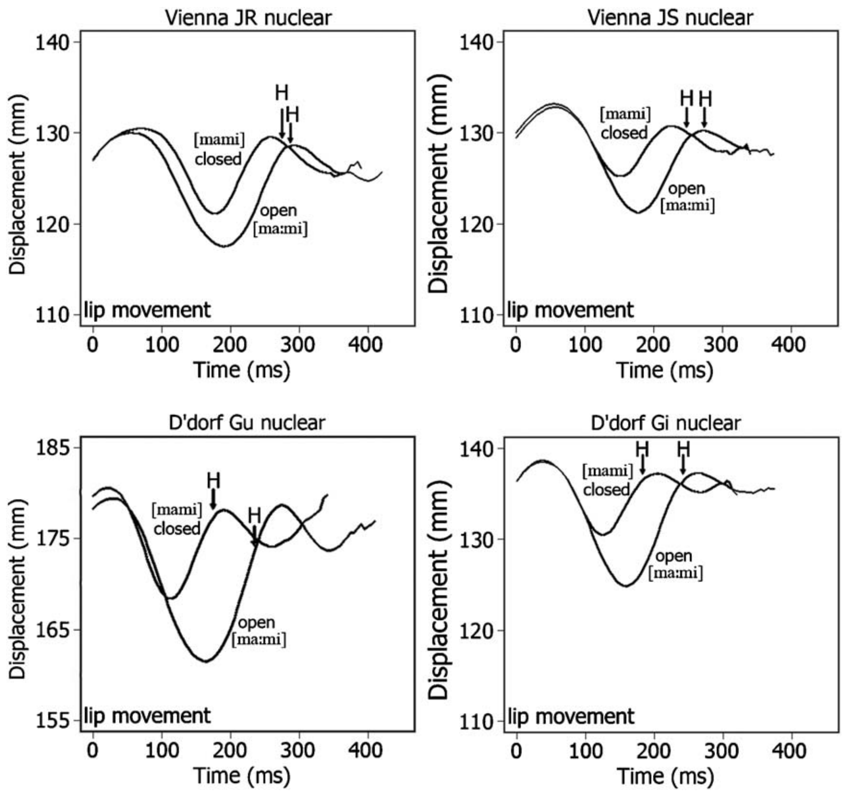


Fig. 11. Averaged contours of vertical lower lip movements for open and closed syllables in the (nuclear) target words Mahmi and Mammi, separately for Vienna speakers JR and JS (above), and Düsseldorf speakers Gu and Gi (below). Nuclear accents. The arrows mark the timing of the f0 peaks (H) relative to the articulatory trajectories.

(33 ms delay) were comparable to those measured for the acoustic anchor types (28 ms delay). They were too large to be explained by different consonant–vowel articulations across the varieties. (Recall that the differences in articulatory alignment between the varieties were rather small in prenuclear accents with only 10 ms delay.) Despite the observed differences in the means, the factor dialectal background failed to reach significance. This can be explained by looking at the individual speakers. Speaker Gi's (Düsseldorf) values tend to overlap with those of the two Vienna speakers. We assume that speakers differ in the extent to which they express pragmatic context (in this case contrasts). Alternatively, contrastive contexts might induce different degrees of delay in the timing of H. In both cases the delay for contrast would be competing with dialectal information (Braun 2007, see also Section 2.3.1).

Unlike our results for the prenuclear accents, there was a strong effect of syllable structure. H peaks were systematically later in closed syllables in the realisations of all speakers, on average 21 ms later in closed syllables than open syllables (Vienna and Düsseldorf together). In open syllables, H occurred 39 ms before the consonant target in Düsseldorf, and 10 ms before it in Vienna. In closed syllables, H occurred 21 ms before the target in Düsseldorf and 15 ms after it in Vienna.

As in Section 2.3.2, the effect of place of articulation was weaker in the articulatory record than in the acoustic one.

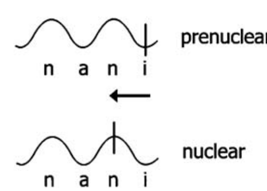


Fig. 12. Schematic articulatory anchor shift between prenuclear and nuclear accents in the synchronisation of H tonal targets and gestural targets (for both the Vienna and the Düsseldorf varieties).

Moreover, peaks were later in target words containing labial consonants compared to alveolar consonants.

Articulatory anchor shift: In prenuclear accents the high f0 target co-occurred with the transvocalic target of the lower lip and tongue tip during production of the unstressed vowel, V2. In nuclear accents, articulatory anchors for the high target corresponded to the consonantal closure during the production of the intervocalic consonant, C2. We argue that the differences between prenuclear and nuclear accents can be best described as an anchor shift in the articulatory dimension, schematised in Fig. 12.

3.3.3. Excursus: the role of L2 for the anchor shift

In line with Silverman and Pierrehumbert (1990) for English and Schepman et al. (2006) for Dutch, we might ask whether the timing difference between nuclear and prenuclear accents is due to the tonal context. They argue that the H tone undergoes a leftwards push, resulting from an additional L following the LH accent in nuclear position.

In fact, in the nuclear accents of both varieties investigated here, the rise in f0 was followed by an immediate fall to an 'elbow' (see Fig. 7), which was labelled in our data as L2. In both varieties, L2 was outside the test word. In the Vienna data, L2 occurred on average 40 ms after the end of the test word in open syllables and 50 ms in closed syllables. In the Düsseldorf data, L2 occurred on average 13 ms after the end of the test word in open syllables and 19 ms in closed syllables.

The fact that L2 was found to be systematically later in the Viennese than the Düsseldorf data, in the same way as H, makes it difficult to argue that it might be microprosodic effects which are responsible for the position of L2. Moreover, since the nuclear syllable is followed in all cases by four unstressed syllables before the phrase boundary, there is unlikely to be time pressure from a straight edge tone. This makes the extra tone appear to be associated, along with the previous H tone, to the accented syllable.

Table 3 shows durations for the fall for each speaker separately. The similarity of the duration of the fall, i.e. the time lag between H and L2, across speakers and varieties supports the interpretation that L2 is associated with the accented syllable. L2 could then be interpreted as a *phrase accent*, in the sense of Grice et al. (2000), with a secondary association to the nuclear syllable, as in Dutch (Schepman et al., 2006). See Grice et al. (2000), Lickley, Schepman, and Ladd (2005) for more comprehensive discussions of phrase accents.

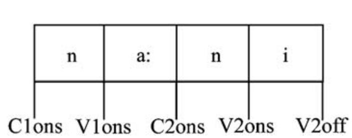
3.3.4. *Acoustic and articulatory anchors compared (nuclear accents)*

Smaller alignment latencies were found for articulatory anchors compared to acoustic ones only for the speakers of the Vienna group. In the Düsseldorf group, latencies were smaller for acoustic rather than articulatory anchors (recall that in prenuclear accents all articulatory alignment values were smaller than the acoustic values).

Furthermore, in a comparison between the standard deviations of the pooled acoustic and articulatory C2 anchors, no systematic difference in dispersion was found (Scheme 1).

Table 3
Mean durations for the fall (L2-H) in nuclear accents.

Speaker	Open syllable Fall duration (ms)	Closed syllable Fall duration (ms)
JS Vienna	126	135
JR Vienna	159	139
Grand mean	143	137
Gi D'dorf	147	142
Gu D'dorf	170	165
Grand mean	154	149



Scheme 1.

4. Conclusion

Our results enable us to compare segmental and articulatory landmarks in terms of alignment latencies and variation within and across conditions. For prenuclear accents latencies were smaller between the H peak and the *articulatory* anchor than between the H peak and *segmental* anchors. However, the picture is not as clear for the nuclear accents. The latencies were smaller for articulatory anchors in the Vienna group but not in the Düsseldorf group, where latencies were smaller for the segmental anchors. Thus, as in D'Imperio et al.'s (2007b) study, articulatory anchors tend to be more closely aligned with the f0 targets than acoustic ones, indicating that looking at articulatory anchors might provide a better insight into how the *tune* is coordinated with the *text*. However, just like D'Imperio et al. (2007b), we did not find that articulatory anchors were less variable in their timing than acoustic ones when comparing different instances within one condition. Each of the three investigated effects on peak alignment will now be treated in turn.

Syllable structure: In nuclear accents effects of syllable structure were ascertained in the alignment of H with the consonantal landmark, both in the articulatory and the acoustic dimensions: peaks were later in closed syllables, as in Spanish and Dutch, providing a degree of support for the view that peaks in rising accents align with the syllable edge. However, although these effects could be found for all speakers in the nuclear accents, only one out of four speakers had later peaks in closed syllables on the prenuclear accents. The picture is thus not as clear as it might be. More research is needed to resolve the issue as to the role of syllable structure in the timing of f0 peaks.

Dialectal background: In both experiments peaks tended to be later in Vienna than in Düsseldorf, even though the dialectal differences were much smaller in the articulatory dimension than in the acoustics. There was, however, a degree of variation within each dialectal group. In fact, in some conditions the Düsseldorf speakers had values approaching those of the Viennese speakers.[5] There is thus no evidence for two distinct accent types, contrary to earlier studies on dialectal variation in German, where

[5]Since the accents investigated here are all contrastive, and since it has been shown for German that contrast can lead to tendentially (although not categorically) later peaks (Braun, 2007), we cannot tell for certain whether the speakers produce a varying amount of contrast or whether their alignment patterns differ regardless of pragmatic context.

Southern and Northern varieties have often been analysed as having two accent categories (Sievers, 1903; Gibbon, 1998). It is clear, then, that our results accord with those of Atterer and Ladd (2004) and Braun (2007), providing further evidence for gradience, supporting a unified account of the rising accents across the different varieties.

Accent status: The situation when comparing prenuclear and nuclear accents is quite different from the situation when comparing the two varieties. Here the differences in alignment are (a) greater and (b) statistically significant, indicating a discrete distinction. In fact, an *anchor shift* for the f0 peak could be observed—from an articulatory movement corresponding to the vowel in prenuclear accents to one corresponding to the intervocalic consonant in nuclear accents. This leftwards shift for nuclear accents appears to be related to a fall in pitch occurring immediately after the rising pitch accent. The fall ended in a relatively consistently aligned elbow which could be taken to be an L phrase accent, along the lines of Schepman et al. (2006) for Dutch and Barnes, Shattuck-Hufnagel, Brugos, and Veilleux (2006, see also Grice et al., 2000) for English. Alternatively it could be analysed as a trailing tone of a tritonal pitch accent. Whatever the source of this tone may be, the difference appears to be discrete and representable in symbolic terms.

Thus, both gradient and discrete differences could be found in alignment with both segmental and articulatory anchors, depending on the source of the variation. More work is clearly needed to pin down how exactly these differences should be represented in phonetics and phonology.

Acknowledgements

The research reported here was funded by the Deutsche Forschungsgemeinschaft (DFG) under the project "The Structure of Tonal Representations: Evidence from Tune-Text Synchronisation (STRETTS)", awarded to Martine Grice and Jonathan Harrington.

Appendix. Speech materials

Prenuclear contrastive accents (one example for each target word)

Mit der NANni nahm sie den SECHS-Uhr-Bus.
Mit der NAHni nahm sie den EIN-Uhr-Bus.
Mit der NENni nahm sie den DREI-Uhr-Bus.
Mit der NAEHni nahm sie den SIEBEN-Uhr-Bus.
Lit.: With the **A** took she the **B** o'clock-bus.
(With **A** she took the **B** o' clock-bus.)

Mit der MAMmi macht sie um NEUN Uhr Schluss.
Mit der MAHmi macht sie um ZWEI Uhr Schluss.
Mit der MEMmi macht sie um ZEHN Uhr Schluss.
Mit der MAEHmi macht sie um VIER Uhr Schluss.
Lit.: With the **C** made she at **D** stop.
(With **C** she finished at **D**.)

Nuclear contrastive accents (one example for each target word)

Q: Hat sie die MAMmi oder die MAHmi bestohlen?
A: Sie hat die MAHmi bestohlen
Lit.: Q: Has she the **E** or the **F** robbed?
A: She has the **F** robbed.
(Q: Has she robbed **E** or **F**?
A: She has robbed **F**.)

Q: Hat Sie die MAMmi oder die MEMmi bestohlen?
A: Sie hat die MAMmi bestohlen.

Q: Hat sie die MEMmi oder die MAEHmi bestohlen?
A: Sie hat die MEMmi bestohlen.

Q: Hat sie die MAEHmi oder die MAHmi bestohlen?
A: Sie hat die MAEHmi bestohlen.

Q: Hat sie die NANni oder die NAHni bestohlen?
A: Sie hat die NAHni bestohlen.

Q: Hat sie die NANni oder NENni bestohlen?
A: Sie hat die NANni bestohlen.

Q: Hat sie die NENni oder die NAEHni bestohlen?
A: Sie hat die NENni bestohlen.

Q: Hat sie die NAEHni oder die NAHni bestohlen?
A: Sie hat die NAEHni bestohlen.

References

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics*, *32*(2), 177–197.
Arvaniti, A., & Ladd, D. R. (1995). Tonal alignment and the representation of accentual targets. In: *Proceedings of the 13th International Congress of Phonetic Sciences*, Vol. 4, Stockholm (pp. 220–223).
Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*, *26*, 3–25.
Barnes, J., Shattuck-Hufnagel, S., Brugos, A., & Veilleux, N. (2006). The domain of realization of the L-phrase tone in American English. In R. Hoffman, & H. Mixdorff (Eds.), *Proceedings of Speech Prosody 2006*. Dresden: TUDpress.
Baumann, S., Grice, M., & Steindamm, S. (2006). Prosodic marking of focus domains – categorical or gradient? In R. Hoffman, & H. Mixdorff (Eds.), *Proceedings of the third conference on Speech Prosody 2006* (pp. 301–304). Dresden: TUDpress.
Becker, J. (2005). Artikulation und Akzentuierung: Pränukleare Kontrastakzente im Deutschen. Master Thesis, unpublished.
Braun, B., & Ladd, D. R. (2003). Prosodic correlates of contrastive and non-contrastive themes in German. In: *Proceedings of the eighth European conference on speech communication and technology 2003*, Geneva (pp. 789–792).
Braun, B. (2007). Effects of dialect and context in the realisation of German prenuclear accents. In: *Proceedings of the 16th international congress of phonetic sciences*, Saarbrücken (pp. 961–964).
Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*, 140–155.

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In T. Kingston, & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech* (pp. 341–376). Cambridge: Cambridge University Press.

Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. In M. Broe, & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the lexicon* (pp. 70–88). Cambridge: Cambridge University Press.

Caspers, J., & van Heuven, V. J. (1993). Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica*, *50*, 161–171.

Cassidy, S., & Harrington, J. (2001). Multi-level annotation in the EMU speech database management system. *Speech Communication*, *33*, 611–677.

D'Imperio, M. (2000). *The role of perception in defining tonal targets and their alignment*. Ph.D. Thesis, Department of Linguistics, The Ohio State University.

D'Imperio, M., Nguyen, N., & Munhall, K. G. (2003). An articulatory hypothesis for the alignment of tonal targets in Italian. In: *Proceedings of the 15th international congress of phonetic sciences 2003*, Barcelona (pp. 253–256).

D'Imperio, M., Petrone, C., & Nguyen, N. (2007a). Effects of tonal alignment on lexical identification in Italian. In C. Gussenhoven, & T. Riad (Eds.), *Tones and tunes*, Vol. 2 (pp. 79–106). Berlin: Mouton de Gruyter.

D'Imperio, M., Espesser, R., Lœvenbruck, H., Menezes, C., Nguyen, N., & Welby, P. (2007b). Are tones aligned with articulatory events? Evidence from Italian and French. In J. Cole, & J. I. Hualde (Eds.), *Laboratory Phonology 9: Phonology & phonetics* (pp. 577–608). Berlin: Mouton de Gruyter.

Gibbon, D. (1998). Intonation in German. In D. Hirst, & A. Di Cristo (Eds.), *Intonation systems. A survey of 20 languages* (pp. 78–95). Cambridge: Cambridge University Press.

Grice, M. (1995). Leading tones and downstep in English. *Phonology*, *12*(2), 183–233.

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

Gussenhoven, C. (2002). Intonation and Interpretation: Phonetics and phonology. In: *Proceedings of the first conference on speech prosody 2002*, Aix-en-Provence (pp. 47–57).

Igarashi, Y. (2004). "Segmental anchoring" of F0 under changes in speech rate: Evidence from Russian. In: *Proceedings of the second conference on Speech Prosody 2004*, Nara (pp. 25–28).

Jilka, M., & Möbius, B. (2007). The influence of vowel quality features on peak alignment. In: *Proceedings of Interspeech 2007* (pp. 2621–2624).

Ladd, D. R. (2008). *Intonational phonology* (second ed.). Cambridge: Cambridge University Press.

Ladd, D. R., Faulkner, D., Faulkner, A., & Schepman, A. (1999). Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of Acoustic Society of America*, *106*(3), 1543–1554.

Ladd, D. R., Mennen, I., & Schepman, A. (2000). Phonological conditioning of peak alignment of rising pitch accents in Dutch. *Journal of Acoustic Society of America*, *107*(5), 2685–2696.

Lickley, R. J., Schepman, A., & Ladd, D. R. (2005). Alignment of "phrase accent" lows in Dutch falling rising questions: Theoretical and methodological implications. *Language and Speech*, *48*(2), 157–183.

Möbius, B., & Jilka, M. (2007). Effects of syllable structure and nuclear pitch accents on peak alignment: A corpus-based analysis. In: *Proceedings of the 16th international congress of phonetic sciences 2007*, Saarbrücken (pp. 1173–1176).

Mücke, D., Grice, M., Becker, J., Hermes, A., & Baumann, S. (2006). Articulatory and acoustic correlates of prenuclear and nuclear accents. In R. Hoffman, & H. Mixdorff (Eds.), *Proceedings of the third on Conference Speech Prosody 2006* (pp. 297–300). Dresden, Dresden: TUDpress.

Mücke, D., & Grice, M. (2006). Articulatory anchors for tonal targets in rising pitch accents. Poster presentation at *10th conference on laboratory phonology*, Paris, June 29–July 1.

Mücke, D., Grice, M., Hermes, A., & Becker, J. (2008). Prenuclear rises in northern and southern German. In: *Proceedings of the fourth conference on Speech Prosody 2008*, Campinas (pp. 245–248).

Mücke, D., Nam, H., Prieto, P., & Goldstein, L. (accepted). Coupling of tone and constriction gestures in Catalan and German. In: *Phonetics and phonology in Iberia 2009*, Las Palmas de Gran Canaria.

Prieto, P., van Santen, J., & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, *23*(4), 429–451.

Prieto, P., Mücke, D., Becker, J., & Grice, M. (2007). Coordination patterns between pitch movements and oral gestures in Catalan. In: *Proceedings of the 16th international congress of phonetic sciences 2007*, Saarbrücken (pp. 989–992).

Prieto, P., & Torreira, F. (2007). The segmental anchoring hypothesis revisited. Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics*, *35*(4), 473–500.

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

Schiller, N. O., Meyer, A. S., & Levelt, W. J. M. (1997). The syllabic structure of spoken words: Evidence from the syllabification of intervocalic consonants. *Language and Speech*, *40*(2), 103–140.

Silverman, K. (1987). *The structure and processing of fundamental frequency contours*. Ph.D. Thesis, University of Cambridge.

Silverman, K., & Pierrehumbert, J. (1990). The timing of prenuclear high accents in English. In J. Kingston, & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech* (pp. 72–106). Cambridge: Cambridge University Press.

Sievers, E. (1903). Über ein neues Hilfsmittel philologischer Kritik. Reprinted in: *Rhythmisch-Melodische Studien: Vorträge und Aufsätze* (pp. 78–111). Heidelberg: Karl Winter (1912).

Welby, P., & Lœvenbruck, H. (2006). Anchored down in Anchorage: Syllable structure, rate, and segmental anchoring in French. *Rivista di Linguistica*, *18*(1), 73–124.

Wichmann, A., House, J., & Rietveld, T. (2000). Discourse constraints on F0 peak timing in English. In A. Botinis (Ed.), *Intonation: Analysis, modelling and technology*, Vol. 15 (pp. 163–182). Dordrecht: Kluwer.

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, *55*, 179–203.

Xu, Y. (2002). Articulatory constraints and tonal alignment. In: *Proceedings of the first conference on Speech Prosody 2002*, Aix-en-Provence (pp. 91–100).

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*, 159–197.