Articulatory and acoustic correlates of prenuclear and nuclear accents

Doris Mücke, Martine Grice, Johannes Becker, Anne Hermes & Stefan Baumann

I*f*L Phonetik
University of Cologne, Germany
{doris.muecke; martine.grice; anne.hermes; stefan.baumann}@uni-koeln.de,
jbecker@smail.uni-koeln.de

Abstract

In this paper we investigate acoustic and articulatory anchors for F0 targets corresponding to prenuclear and nuclear accent peaks in German, both across two different articulation rates (normal and fast) and across two different syllable structures (CV: and CVC). For the articulatory measurements we used Electromagnetic Midsagittal Articulography (EMMA).

Whereas in Dutch the H peak of a rising prenuclear accent has been shown to occur at the edge of the accented syllable [1], in German the peak occurs during the syllable following the accented one, in the vowel. Like in English and Dutch, nuclear peaks in German are aligned earlier than prenuclear ones: H peaks were found to occur at some point during the consonant following the vowel of the accented syllable, although no consistent acoustic anchor could be identified.

We found that F0 turning points aligned more systematically with minima and maxima in the kinematic signals than with acoustically defined events. Furthermore, we interpret the difference in alignment between prenuclear and nuclear accents as a shift from a gesture corresponding to a vowel to a gesture corresponding to a consonant. Within each accent type the kinematic alignment was stable across the different conditions.

1. Introduction

We report on two production experiments on German, examining the temporal synchronisation of tonal targets (realised as minima and maxima in the F0 contour) with landmarks in the acoustic and kinematic signals.

Acoustic studies have revealed that F0 peaks in prenuclear rises in Dutch are found to align with the right edge of the accented syllable [1], i.e. at the end of the vowel in open syllables and the coda consonant in closed syllables. In experiment A we compare our acoustic results to those for Dutch using similar speech material.

Further acoustic studies have shown for English [2] as well as for Dutch [1, 3] that pitch accent status in the intonational hierarchy (nuclear, prenuclear) affects the alignment of F0 peaks. Furthermore, the alignment of F0 events with acoustic landmarks in Dutch nuclear accents was sensitive to vowel length, and, indirectly, syllable structure: F0 peaks were aligned *midway* through the accented vowel if phonologically long and at the *end* of the accented vowel if phonologically short [3]. In experiment B we compare these results with those for German prenuclear and nuclear accents.

It has been claimed elsewhere [4, 5], that some of the variation in aligning F0 peaks with segments in the acoustic signal might be the result of differing degrees of vowel-consonant articulation in the supra-laryngeal system. This implies that there would be less variation if, instead, F0 peaks were aligned with articulatory anchors. Both experiments explore how far this is the case.

2. Experiment A: Prenuclear rises

2.1. Method

Our first experiment explored stability of peak alignment in prenuclear rises in German with landmarks in the articulatory and acoustic dimension. Stability was tested at two articulation rates (normal/fast) and across two different phonological vowel lengths (long/short). If German is like Dutch [1], F0 peaks should be aligned consistently with landmarks of the right edge of the accented syllable. Neither phonological vowel length nor articulation rate should affect the measured alignment pattern.

Target words include the segmental string in (1) (C=nasal [n, m], V1 and V3 = non-high vowel [a, ɛ] and V2 =high vowel [i]).

(1) **C1 V1** C2 V2 C3 V3 [naːni naː] **“Nah**ni nah(m)”

Target words were produced in sentence pairs, where they were contrastive topics as in (2) (see [6] for contrastive topic intonation in German).

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

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

Two speakers were selected, one female (GI) and one male (GU), both students in their mid-twenties who have spent their first 20 years in the low Franconian speech area near to Düsseldorf. Recordings were made with the Carstens articulograph AG100 and a time-synchronized DAT recorder. Sensors were placed on lower lip, tongue tip and tongue body (4cm behind the tip). The kinematic data were recorded at 400hz, downsampled to 200hz and smoothed with a 40Hz low-pass filter. The acoustic data were digitized at 44.1kHz. The speakers read the sentence pairs from a monitor, first at a self-selected (normal) rate (10 repetitions), then at a fast rate (10 repetitions). Lists were in pseudo-randomized order (2 speakers x 8 target words x 20 repetitions = 320 tokens in total).

Acoustic and articulatory data were labelled in the EMU speech database system by hand. For the initial rise contour, we identified the local turning points for a low valley and a high peak in the F0 track, following the conventions described in [1, 6]. Like in [6], the valley for the local minimum was often rather broad and in some cases more difficult to identify, while the peak was usually clearly identifiable. Segment boundaries of consonants and vowels of the accented and postaccented syllable were annotated in the acoustic waveform (e.g. C1onset, V1onset). In the kinematic data, vertical targets were located at zero-crossings in the respective velocity trace. For consonantal gestures, we labelled the maximum for the primary constriction (lower lip and tongue tip) in C1 and C2. The transvocalic gestures were located at the tongue tip or lower lip minimum between two constrictions for C1 and C2 [7]. The vocalic gestures were labelled at maximum targets for the V1-to-V2 ([a] or [ɛ] to [i]) tongue body raising movement.

Utterances were removed from analysis if no clear turning points were locatable either in the kinematic curve or in the pitch track. All tokens minus four for speaker GU (156) and all tokens minus twenty four for speaker GI (136) were analysed.

2.2. Results and discussion

First, we investigated the phonetic alignment of the H target with the segmental string and found that German is unlike Dutch (see 2.1). Figure 1 shows the schematic alignment for the initial rise relative to the accented and unaccented syllable for the German data. The figure is to scale, and based on means for two speakers at normal articulation rate.

The H peak occurs late in V2, c.31ms after V2onset for (a) and c.45ms for (b). We conducted two-way ANOVAs for each speaker with vowel length and articulation rate as independent variables, and the latency of H relative to the end of the accented syllable as dependent variable. There was a large effect of rate on peak alignment with the syllable end for speaker GU [$F(1, 153) = 81.3815$, $p<0.0001$], as well as a small effect of vowel length [$F(1, 153) = 6.7543$, $p<0.05$]. For speaker GI, we found a highly significant effect of both factors, rate [$F(1, 133) = 224.19$, $p<0.0001$] and vowel length [$F(1, 133) = 178.60$, $p<0.0001$]. The results show that the syllable based hypothesis for peak alignment is not valid for German prenuclear rises.

We then tested the alignment of the H peak with other nearby segmental landmarks. Since H appeared to occur late in V2, we calculated latencies for H relative to the acoustic onset of V2.

Further latencies were calculated for the H peak relative to articulatory landmarks: the maximum of the tongue body raising in V2 (vocalic gesture) and the corresponding articulatory minimum of the primary constrictors, lower lip and tongue tip (transvocalic gesture). All latencies for speaker GU were plotted in figure 2. The boxes mark quartiles and medians in ms for the occurrence of the H peak relative to the investigated landmarks (marked at the zero line). In the acoustic dimension, all H latencies relative to the V2onset were rather large (2a). In the articulatory dimension, all H latencies were relatively short (2b-d). The temporal occurrence of H appeared to be synchronized with the articulatory maximum of the vocalic gesture and the minimum of the transvocalic gesture (for the acoustic V2 segment).

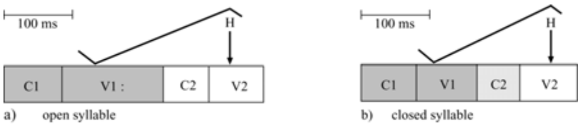

Figure 1: *Schematic prenuclear peak alignment with (a) phonologically long and (b) short vowels*

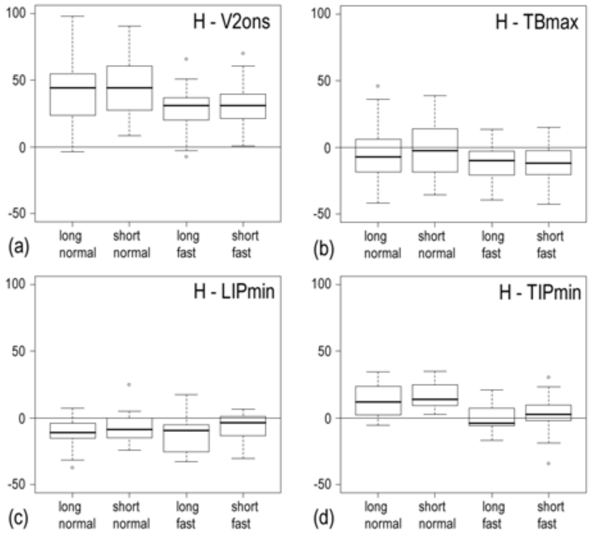

Figure 2: *H latencies for speaker GU, prenuclear peaks*

Table 1 shows means and standard deviations (in parenthesis) for two speakers of H alignment relative to the acoustic and articulatory anchor types under investigation. The following anchor types are differentiated: acoustic vowel onset of V2, target of the vocalic gesture TBmax and target of the transvocalic gesture LIP/TIPmin during V2. The mean values for two speakers show large latencies for H relative to V2onset. In the phonologically short vowel condition, H occurred c.45ms after V2onset at normal rate and c.29ms at fast rate. But H was closely phased with the articulatory targets measured in the production of the respective vowel: for the same short vowel condition mentioned above, H occurred c.1ms before the constrictor's minimum (LIP/TIPmin) at normal rate and c.3ms before it at fast rate.

2 speakers (ms)		long	short
H-V2ons	normal	31 (25)	45 (26)
	fast	24 (14)	29 (13)
H-TBmax	normal	-13 (21)	-6 (19)
	fast	-14 (13)	-9 (13)
H-LIP/TIPmin	normal	-10 (20)	-1 (18)
	fast	- 9 (14)	-3 (11)

Table 1: *Means and standard deviations (in parenthesis) of two speakers for three anchors*

We conducted a series of two-way ANOVAs (vowel length x articulation rate) with latencies of the H peak relative to the anchors (H-V2ons, H-TBmax, H-TIPmin, H-LIPmin) as dependent variables for each speaker. Significant values were set on standard, $p=0.05$.

For speaker GU we found the peak alignment to be affected significantly by articulation rate for H-V2ons [$F(1, 153) = 20.4857$, $p<0.0001$], H-TBmax [$F(1, 153) = 8.5637$, $p<0.0001$] and H-TIPmin [$F(1, 77) = 25.7340$, $p<0.0001$]), but we found no significant effect of vowel length. For the alignment of H-LIPmin, we found no significant effects either of vowel length [$F(1, 77) = 3.2922$, $p=0.07$] or of rate [$F(1, 73) =0.0161$, $p=0.89$], which means that the occurrence of the H peak relative to the transvocalic minimum of the lower lip was robust across all tested conditions.

A different picture arose for speaker GI. In contrast to the first speaker, we found significant effects of vowel length on peak alignment, but not of rate. This was true for all articulatory anchors (H-TBmax [F(1, 133) =17.4595, p<0.0001], H-TIPmin [F(1, 58) =11.5643, p<0.001] and H-LIPmin [F(1, 72) =16.076, p<0.001]), as well as for the acoustic anchor (H-V2ons [F(1, 133) =21.7552, p<0.0001]). Even though there was an effect of articulation rate on the acoustic anchor H–V2ons [F(1, 133) =11.5643, p=0.01876], the rate factor had no significant effect on peak alignment for any of the articulatory anchors.

Results show H was closely phased with the vowel production of V2 in prenuclear rises in German, even though the alignment patterns were affected by vowel length and articulation rate for the two speakers in a different way. The tightest synchronization for the alignment of H was with the transvocalic minimum, but there was also a close alignment with the vocalic maximum. Both correspond to the maximum of vocal tract openness and therefore a maximum of sonority during V2.

3. Experiment B: Nuclear rises

3.1. Method

In the second experiment we investigated the effect of pitch accent status. To do this we compared the alignment of prenuclear and nuclear accent peaks. In line with [2, 3], we expect H to be aligned earlier in nuclear than in prenuclear accents.

We recorded the same speaker GI from Experiment A within the same recording session and used the same recording and annotation procedure. For the speech material, we constructed sentences with the same target words as in Experiment A, but placed them this time in such a way that they would be produced with nuclear accents. We used contrastive contexts and short dialogs as in (3):

(3) Q: Hat sie die MAMmi oder die MAHmi bestohlen?
Did she rob MAMmi or MAHmi?

A: Sie hat die MAHmi bestohlen.
She robbed MAHmi.

Like in Experiment A, phonological vowel length (long/short) and articulation rate (normal/fast) were varied, as well as constriction type (lower lip/tongue tip), making a total of 160 stimuli.

3.2. Results and discussion

All nuclear tokens minus two (158) were analysed, along with all the prenuclear tokens for speaker GI (136) from experiment A. First, we measured the "Silverman/ Pierrehumbert"-peak delay [2], which is peak position relative to V1 onset in the acoustic signal. Figure 3 show prenuclear and nuclear peak delay plotted as a function of vowel duration (two rates, two vowel lengths, one speaker). These results support the hypothesis that H is earlier in nuclear accents than in prenuclear ones [2, 3]. Figure 3 also shows that, across conditions, peak delay increases with an increase in vowel duration.

We also investigated the temporal alignment of H with other landmarks in the segmental string. Figure 4 shows the schematic alignment for the nuclear rise with the accented and unaccented syllables (analogous to figure 2). H occurred in the C2 segment. The segmental anchor for the H peak shifted from the V2 segment in prenuclear position to the preceding C2 segment in nuclear position. As in the Dutch study [3], German nuclear peak alignment was affected by phonological vowel length. H occurred early in C2 for (a), c.13ms after C2onset, but late in C2 for (b), c.27ms after C2onset.

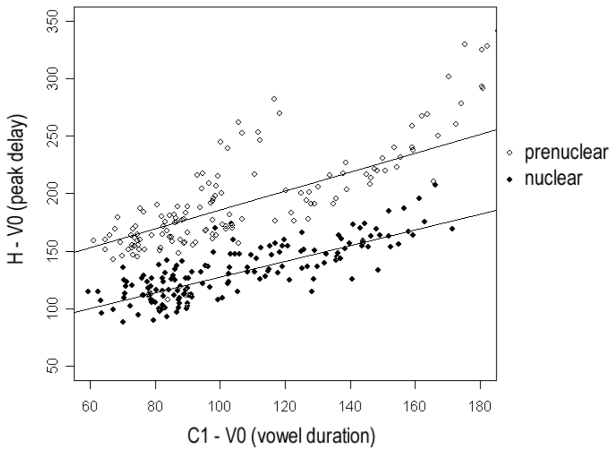

Figure 3: *Peak delay as a function of vowel duration*

Since H was located (acoustically) in the C2 segment, we calculated latencies for H relative to the onset of C2 and V2 (which is simultaneously the offset of C2). For the articulatory dimension, we investigated peak alignment with the maximum closure in the production of C2. The constrictor maximum corresponds to the target of the consonantal gestures for C2. We calculated latencies for H relative to the maximum of lower lip and tongue tip constriction. The boxplots given in Figure 5 show large latencies for H relative to the segmental boundaries for C2. The latencies for H relative to the articulatory events corresponding to the C2 production are smaller than the acoustic ones. However, H alignment with articulatory gestures in nuclear accent position does not appear to be as stable as in the prenuclear condition. This may be due to the fact that nuclear accents allow for more variation in prominence, since they generally carry a higher semantic and pragmatic load.

We conducted two-way ANOVAs (vowel length x articulation rate) with the latencies for nuclear F0 peaks relative to the different segmental and articulatory anchors (H-V2ons, H-C1ons, H-TIPmax, H-LIPmax) as dependent variables. Even though we show that the mean latencies were shorter for H relative to articulatory anchors than for H relative to acoustic anchors, all latencies were significantly affected by vowel length and articulation rate with p<0.01.

Figure 4: *Schematic nuclear peak alignment with (a) phonologically long and (b) short vowels*

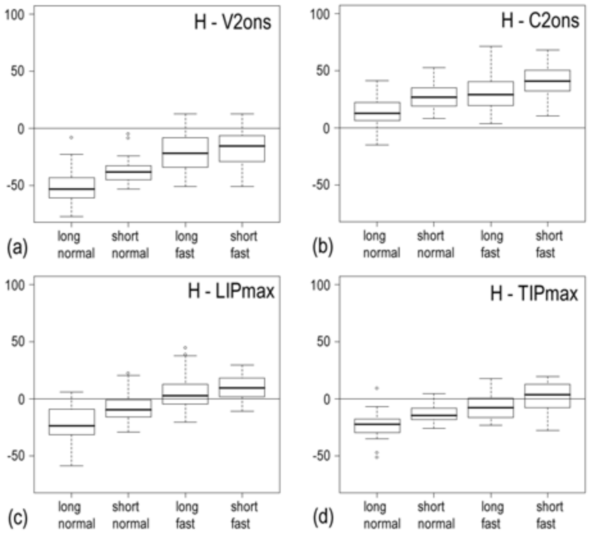

Figure 5: *H latencies for speaker GI, nuclear peaks*

To summarise, we confirm previous findings that nuclear accent peaks are earlier than prenuclear ones. The articulatory explanation we offer is that there is an anchor shift from a gesture corresponding to a vowel to one corresponding to a consonant. To be more precise, the peak in prenuclear accents was aligned with the transvocalic minimum and vocalic maximum, both corresponding to the vowel production in V2 (e.g. minimum for the tongue tip movement in figure 6a). In nuclear accents, the peak was aligned with the consonantal maximum in the C2 production (e.g. maximum for the lower lip movement in figure 6b).

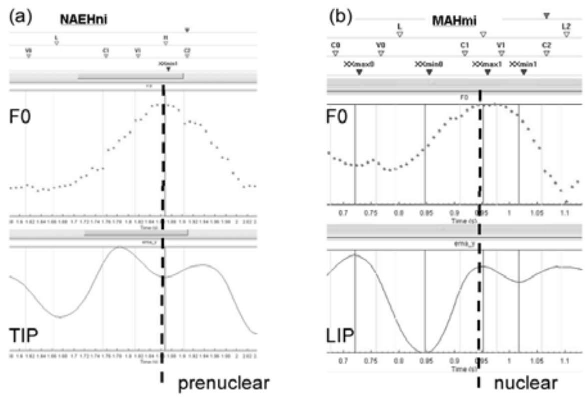

Figure 6: *Alignment of F0 peak with kinematic landmarks for (a) prenuclear and (b) nuclear accents*

4. Conclusion

German, in contrast to Dutch, does not align the prenuclear accent peak with the end of the accented syllable. Rather, the peak occurs within the postnuclear vowel. In articulatory terms, the F0 peak is aligned with targets corresponding to the postaccented vowel (transvocalic minimum of lower lip or tip and vocalic maximum of tongue body, figure 7a-b). In nuclear accents, by contrast, peaks are timed with the articulatory maximum of the consonantal gesture (figure 7c), although alignment was not as clear as in prenuclear position.

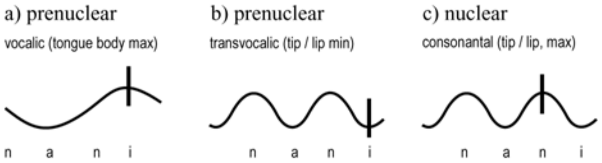

Figure 7: Targets for articulatory gestures serving as anchors for tonal alignment

Even though there are inter-speaker differences (alignment of speaker GU was mainly affected by articulation rate, whereas alignment of speaker GI was predominantly affected by phonological vowel length), there was a closer synchronisation of F0 peaks with articulatory gestures than with acoustic landmarks.

These results shed light on the findings of a study on lip movements for Neapolitan Italian questions and statements [4], which revealed that F0 peaks were systematically aligned with a lip aperture maximum for L+H* rises in statements and with a lip aperture minimum for the L*+H rises in questions. The H peak can be interpreted as being aligned with the target of a transvocalic gesture in statements and with the target of the consonantal gesture in questions, suggesting that there is also an anchor shift, similar to the one reported on above for prenuclear and nuclear accents in German. The difference is that the anchor shift in Neapolitan represents a paradigmatic choice on the part of the speaker, whereas the difference in the German data is a syntagmatic one, since nuclearity is, inter alia, positionally defined.

Thus, what might appear to be arbitrary alignment points in the acoustic signal turn out to be highly restricted in the articulatory dimension, and are closer to alignment patterns which have been shown to be both distinctive and categorical than inspection of acoustic data is able to reveal.

5. References

[1] Ladd, D.R.; I. Mennen, A. Schepman, 2000. Phonological conditioning of peak alignment of rising pitch accents in Dutch**.** *JASA* 107, 2685-2696.

[2] Silverman, K.; J.Pierrehumbert (1990). The timing of prenuclear high accents in English. In Kingston, J. & Beckman, M. (eds.), *Papers in Laboratory Phonology I: between the grammar and physics of speech.* Cambridge: CUP

[3] Schepman, A.; R.Lickley; D. R. Ladd, (in press). Effects of vowel length and "right context" on the alignment of Dutch nuclear accents**.** To appear in *Journal of Phonetics.*

[4] D'Imperio, M.; N. Nguyen; K.G. Munhall, 2003. An articulatory hypothesis for the alignment of tonal targets in Italian. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 253-256.

[5] Mücke, D.; M. Grice (2004). Gestural anchors for F0 targets in German speech: an EMA-study, International Conference on Tone and Intonation, Santorini, September 2004.

[6] Braun, B.; D. R. Ladd (2003): Prosodic Correlates of Contrastive and Non-Contrastive Themes in German. *Proceedings of the 8th European Conference on Speech Communication and Technology*, September 2003, Geneva, 789-792.

[7] Browman, C.P. & L. Goldstein 1988. Some Notes on Syllable Structure in Articulatory Phonology. *Phonetica* 45, 140-155.