PHRASE BOUNDARIES AND PEAK ALIGNMENT: AN ACOUSTIC AND ARTICULATORY STUDY

Doris Mücke, Anne Hermes

I*f*L Phonetik, University of Cologne
{doris.muecke; anne.hermes}@uni-koeln.de

ABSTRACT

The present study investigates the effect of an upcoming phrase boundary on peak alignment in rising pitch accents in a variety of German (Vienna). We measured the synchronisation of F0 peaks with acoustic segments and articulatory movements. As expected, the closer the tone bearing unit is to the phrase boundary, the earlier the F0 peak is aligned. Alignment is more stable in relation to articulatory movements than it is in relation to syllable or segment boundaries. More precisely, the F0 peak is aligned with the oral closing gesture.

With increasing time pressure from a phrase boundary two compensatory effects were found: the closing gesture is slowed down (pre-boundary lengthening) and the H peak is realised earlier (it is shifted away leftwards from the boundary).

Keywords: tonal alignment, articulatory gestures, boundary tone, pre-boundary lengthening, time pressure

1. INTRODUCTION

It has been shown for several related languages that the timing of F0 peaks with segmental boundaries is affected by an upcoming phrase boundary: In Dutch [1] and Mexican Spanish [2], F0 peaks shifted leftwards in phrase final position. While in both languages the end of the rise was pushed leftwards with increasing time pressure, the start of the rise remained more or less constant. As a result the rise contours were steeper and the pitch excursion was shorter.

Upcoming phrase boundaries also affect the timing of segments at the boundary, in that they are lengthened. In the articulatory dimension, final lengthening is analysed as the result of changes in the localised speaking rate [3]. Eg. in English [4], oral closing gestures in VC sequences slowed down in phrase final position. For the present study, we investigate effects of an upcoming phrase boundary on the tonal alignment of F0 peaks with articulatory gestures.

2. METHOD

Speakers: We selected two native speakers from Vienna, both female, one student (JS) and one make-up artist (JR), both in their mid twenties.
Speech Material: We developed a set of speech materials designed to elicit eight target words in phrase medial and final position (table 1). Target words contain bilabial or alveolar nasals.

Table 1: Structure of target words.

	monosyllabic	disyllabic
long vowel	CV:C [ma:m]	CV:CV [ˈma:mi]
short vowel	CVC [mam]	CVCV [ˈmami]

We hypothesised that phrase finality of the test words would exert considerable time pressure for the realisation of the tones, since the LH accent and the L boundary tone (L-%) have to be realised on the same target word. Since nuclear contrastive focus accents have been shown to typically involve rising accents [5], we designed mini-dialogues with target words in contrastive focus in (1) phrase medial and (2) phrase final position:

(1) Q: Hat sie die Nahni oder die Mahmi bestohlen?
- *Has she robbed Nahni or Mahmi?*
A: Sie hat die <u>Nah</u>ni bestohlen.
lit.: - *She has Nahni robbed.*

(2) Q: War es die Nahni oder die Mahmi?
- *Was it Nahni or Mahmi?*
A: Es war die <u>Nah</u>ni.
lit.:- *It was the Nahni.*

Recordings: The speech material was read from a computer monitor in pseudo-randomised order. We recorded 160 tokens per speaker (8 target words x 2 phrase positions x 10 repetitions x 2 speakers). Recordings were made with an Articulograph, AG100, and a time-synchronised DAT recorder (sensors placed on lower lip, tongue tip and tongue body). Kinematic data were recorded at 400Hz (downsampled to 200Hz, smoothed with 40Hz low-pass filter; acoustic data digitized at 44.1kHz). *Labelling procedure:* Acoustic and articulatory data were handlabelled in the EMU Speech Database System. Segments were identified in the acoustic waveform:

(a) monosyllabic: | C1 | V1 | C2 |
landmarks: C1onset | V1onset | C2onset | C2offset

(b) disyllabic: | C1 | V1 | C2 | V2 |
landmarks: C1onset | V1onset | C2onset | V2onset | V2offset

For the rise contour we identified local turning points for the start of the rise, the L-valley (L), and the end of the rise, the high peak (H). Since the rise was followed by an immediate fall, we labelled the turning point for the end of the fall, the low valley L2. Pitch perturbations were ignored. In the kinematic data, vertical targets (maximum/ minimum) of the consonantal opening and closing gestures were located at zero crossings in the respective velocity trace and peak velocity targets were labelled at zero crossings in the respective acceleration trace [7].

3. RESULTS AND DISCUSSION

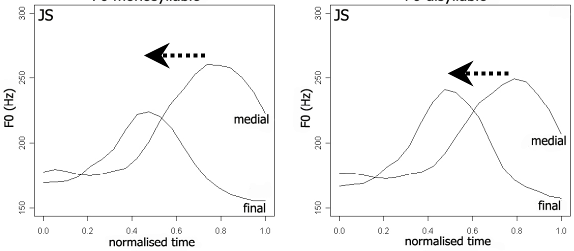
All tokens (320) went into the statistical analysis.

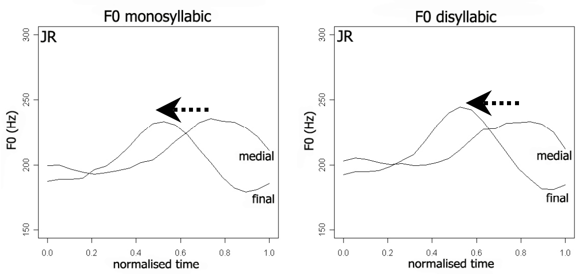
3.1. F0 contour, rise time and rise excursion

Figure 1 shows the averaged F0 contours for mono- and disyllabic target words in phrase final and phrase medial position (e.g. Es war die Mamm vs. Sie hat die Mamm bestohlen). H occurs earlier in phrase final position than in phrase medial ones. The H is shifted leftwards with time pressure.

We then calculated the rise duration and excursion of the rising LH accents for each speaker separately (table 2). For disyllabic target words, the shape of the rise appears to be relatively invariant for both speakers. However, for monosyllabic target words, we found speaker dependent strategies in the realisation of the rise.

Figure 1: Averaged F0 contours for target words in phrasefinal and medial position, all data, speaker separately.





Although the rise *duration* remains constant for speaker JS, (t-test, p>0.05, n.s.), there was a considerable shortening of the rise for speaker JR, (decrease of 17%, c.202ms for target words in phrase medial position and c.167ms for phrase final position, p<0.001). In contrast, the rise *excursion* remains constant for speaker JR, (p>0.05, n.s.), while it shows a decrease of 29% for speaker JS (c.7.15st for phrase medial position and c.5.05st for final position, p<0.001).

Table 2: Mean rise durations (in ms) and excursion sizes (in semitones). Standard deviations in parenthesis.

	monosyllabic				disyllabic			
	phrase medial		phrase final		phrase medial		phrase final	
	long	short	long	short	long	short	long	short
	Rise Duration (ms)							
JS	196 (23)	194 (18)	200 (24)	184 (24)	217 (22)	204 (18)	210 (21)	187 (20)
JR	209 (30)	195 (23)	168 (24)	166 (16)	196 (17)	193 (20)	190 (17)	166 (18)
	Rise Excursion (semitones)							
JS	6.5 (0.5)	7.8 (0.52)	4.4 (0.8)	5.7 (0.78)	5.5 (0.79)	6.4 (0.63)	6.1 (0.8)	6.3 (0.8)
JR	3.7 (0.88	3.9 (0.82)	3.4 (1.1)	3.9 (1.7)	3.2 (0.9)	3.3 (0.83)	4.0 (1.2)	4.3 (1.1)

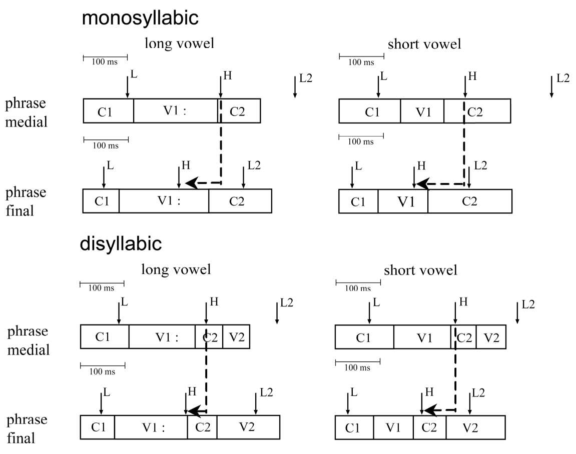
To sum up, rise duration and excursion remained more or less constant for disyllabic words. For monosyllabic words we found speaker dependent strategies to resolve the LH adjustments with time pressure: While speaker JS decreased the F0 scaling and kept the rise duration intact, speaker JR shortened the rise duration and retained the scaling (also reported in [8] for stress clash in Greek). Across all conditions (mono- and disyllabic words), H was shifted leftwards with time pressure.

3.2. Segmental Anchors

Figure 2 shows schematic alignment patterns for L, H and L2 relative to the segmental sequences. The figure shows four different alignment patterns in phrase medial and phrase final position for monosyllabic and disyllabic target words in phonologically long and short vowel conditions, such as /ma:m/ vs. /mam/ (top) and /ma:mi/ vs.

/mami/ (bottom). The figure is to scale, and calculated on the basis of mean durations.

Figure 2: Schematic alignment patterns in phrase medial and phrase final position, one speaker (JR).



As was expected, we found that time pressure in the phrase final condition entails a leftward push of the tonal targets relative to segments. H coincides with the second target consonant, C2, in phrase medial position. In phrase final position H coincides with the accented vowel V1 in monosyllabic condition and either with C2 or the boundary between C2 and V1 (in disyllabic condition).

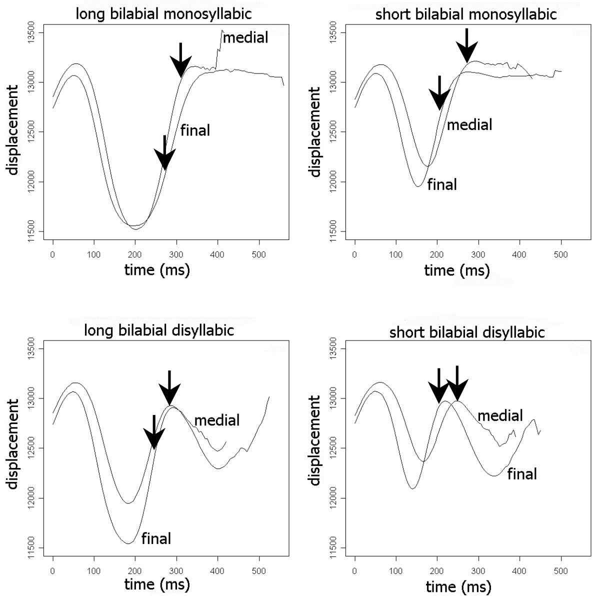
In addition, we examined the alignment of H relative to the end of the target word in phrase final position for each speaker separately. The alignment is affected by vowel length in monosyllabic words (t-test, $p<0.001$ (JS); $p<0.05$ (JR)) as well as in disyllabic words (t-test, $p<0.001$ (JS); $p<0.01$ (JR)).

To sum up, H is neither aligned with the end of the target word nor with a particular segment, e.g. C2 or V1, in phrase final position.

3.3. Kinematic anchors

We examined the coordination of the H peak with landmarks of dynamically defined articulatory gestures of the lower lip and tongue tip movements [6]. Figure 3 shows averaged trajectories of the lower lip for mono- and disyllabic target words in phrase medial and final position. The contours are plotted for the vertical movements of the lower lip (high displacement indicates closed lips). The arrows mark the alignment of the accentual H peak relative to the lower lip movement.

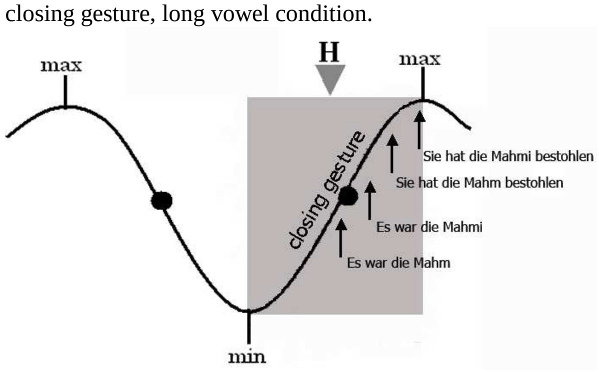
Figure 3: Averaged contour of lower lip movement and temporal occurrence of H, two speakers.



Throughout all conditions, H coincides with the C2 closing gesture. More precisely, H occurs *at* the target in phrase medial position (e.g. in disyllabic target words the latencies for H relative to the articulatory target are around –2ms for long vowel condition and around 19ms in short vowels). In phrase final position H occurs *before* the target (e.g. in disyllabic target words around –48ms for long vowels and around –18ms for short vowels).

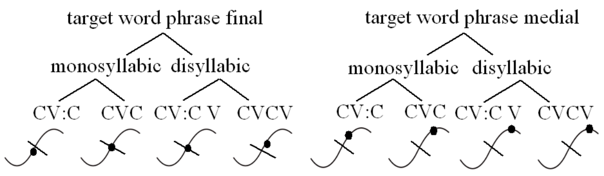
Figure 4 shows stylised alignment patterns for test words in the long vowel condition (<u>Mahm</u> and <u>Mah</u>mi) when the test word is in phrase medial and final position. With increasing time pressure, H was shifted from the target of the closing gesture up to and just beyond the peak velocity.

Figure 4: Stylised alignment patterns for H relative to the closing gesture, long vowel condition.



The differences in the alignment patterns appear to be gradient for target words in phrase final position. Phrase finally, H is aligned (i) earlier in mono- than in disyllabic target words and (ii) earlier in long than in short vowel condition (see figure 5).

Figure 5: Gradient alignment patterns for H relative to the closing gesture, long vowel condition, peak velocity is marked by the slash for orientation.



3.4. Pre-boundary lengthening

Table 3 shows mean duration and displacement for the lower lip and tongue tip kinematics corresponding to the C2 closing gesture. The duration of the closing gesture was affected by the phrase boundary in all test conditions, $p<0.05$. The effect was stronger for monosyllabic target words (e.g. increase of 44%, c.62ms for monosyllabic words in the long vowel condition and 25%, c.25ms for disyllabic ones). The effect was weaker for target words containing a phonologically short vowel (e.g. increase of 2%, c.3ms in short vowel condition, monosyllabic words). The displacement of the closing gesture was affected in disyllabic words, $p<0.001$, but not in monosyllabic words, $p>0.05$, n.s.

Table 3: Mean durations and displacement of the closing gesture, two speakers. Standard errors in parenthesis.

		duration in ms		displacement in mm	
		medial	final	medial	final
mono-syllabic	long	140 (8)	202 (62)	16 (1.7)	15 (1.9)
	short	124 (15)	127 (28)	11 (1.5)	12 (1.4)
di-syllabic	long	100 (11)	125 (15)	12 (3.4)	17 (1.5)
	short	81 (7)	95 (16)	9 (1.3)	12 (0.9)

In monosyllabic target words the upcoming phrase boundary results in a change in duration but not in displacement. In contrast to [4], we found the closing gestures to be longer and more displaced in disyllabic words. Beside these differences in articulatory strategies, we found that all C2 closing gestures in phrase final position were longer (i) in mono- than in disyllabic target words and (ii) longer in the long than in the short vowel condition.

In phonologically long vowel (as opposed to short ones) the closing gesture is lengthened to a greater extend than in the short vowels. Lengthening is not an option for the phonologically short vowel, since it would involve a risk of confusion with its long counterpart.

4. CONCLUSION

In Vienna German, the F0 peak in rising LH accents was affected by an upcoming phrase boundary in phrase final words: as in Dutch [1] and Mexican Spanish [2], the H peak was shifted leftwards to compensate for tonal crowding.

Furthermore, we found more stable alignment patterns for H with articulatory movements than with acoustic segments and syllables. H is aligned with the closing gesture in the second target consonant C2.

Fine phonetic adjustments in the alignment of tonal targets with articulatory movements appear to compensate for time pressure. The H peak is shifted leftwards, while the closing gesture for C2 is slowed down. Both of these adjustments lead to a fuller realisation of the intonation contour without losing the alignment of the closing gesture with the H peak.

5. REFERENCES

[1] Caspers, J., van Heuven, V. 1993. Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50.
[2] Prieto, P., van Santen, J. & Hirschberg, J. 1995. Tonal Alignment Patterns in Spanish, *JPhon*, 23, 429-451.
[3] Byrd, D., Kaun, A., Narayanan, S. & Saltzman, E. 2000. Phrasal signatures in articulation. *Papers in LabPhon V*. Cambridge Univ. Press, 70 – 87.
[4] Edwards, J., Beckman, M.& Fletcher, J. 1991.The articulatory kinematics of final lengthening. *JASA* 89(1), 369-82.
[5] Baumann, S., Grice, M. & Steindamm, S. 2006. Prosodic Marking of Focus Domains - Categorical or Gradient? *Proceedings Speech Prosody 2006*, Dresden, 301-304.
[6] Browman, C.P., Goldstein, L. 1988. Some Notes on Syllable Structure in Articulatory Phonology. *Phonetica* 45, 140-155.
[7] Mücke, D., Grice, M., Becker, J. & Hermes, A. (submitted). Sources of variation in tonal alignment: evidence from acoustic and kinematic data.
[8] Arvaniti, A., Ladd, R.D & Mennen, I. 2000. What is a starred tone? Evidence from Greek. *Papers in LabPhon V*. Cambridge Univ. Press, 119–131.