Segmental and positional effects in tonal alignment: An articulatory approach

Henrik Niemann, Martine Grice & Doris Mücke

IfL Phonetik, University of Cologne, Germany

{henrik.niemann|martine.grice|doris.muecke}@uni-koeln.de

Abstract

In this production study, we investigate the effects of phrasal position and segmental makeup on the alignment of tonal events corresponding to rising nuclear pitch accents. Conditions included target words in initial, noninitial and final position, monosyllabic and disyllabic target words, and open and closed target syllables. Although cross linguistically the beginning of the rise (corresponding to the L tone) is relatively stable in its alignment with acoustic and articulatory landmarks, the endpoint, or target, of the rise (H) is notoriously variable. This highly variant peak alignment pattern on the acoustic surface corresponds to a stable coordination between the tone and the articulatory vowel gesture.

***Keywords**: tonal alignment, articulatory phonology, coordination, contextual effects*

1. Introduction

In recent years there have been many studies investigating the temporal arrangement of F0 events with respect to the words with which they are coproduced. The first were concerned with landmarks in the acoustic signal, followed by studies exploring landmarks in the articulatory gestures used to produce the words (see Ladd et al. 2008 for an overview). Most of these studies have been concerned with the question of stability – or invariance – of F0 targets or events relative to these landmarks under different prosodic and segmental conditions (see Wichman et al. 2000 for an overview). The main focus of attention has been on pitch accents with a rise up to a high peak (L+H*), and consequently on F0 events corresponding to the low (L) and high (H) tones.

It has been found across many studies in a number of languages that the L is consistently aligned just before or after the acoustic onset of the accented syllable (Arvaniti, Ladd & Mennen 1997, Ladd et al. 1999), and at the beginning of the articulatory gesture for the onset consonant (Mücke et al. 2012, Niemann, Mücke, Nam, Goldstein & Grice 2011). However, the alignment of the F0 peak for H shows a high degree of variation. Its timing appears to be sensitive to context, both the proximity of an upcoming phrase boundary and syllable structure. For example, it is aligned earlier in phrase-final position compared to other positions within the intonation phrase (Silverman & Pierrehumbert 1990, Caspers & van Heuven 1993) and is aligned later in closed syllables than in open ones (Mücke et al. 2009, Prieto & Torreira 2007). There is evidence that the effect of syllable structure on peak alignment disappears when investigating the peak alignment with articulatory gestures (Mücke, Grice & Hermes 2008). However, most of these studies were restricted to transvocalic gestures, that is they investigated consonantal gestures (opening of the lips during the vowel), and only indirectly captured the vocalic gesture (the tongue body movement directly related to vowel production). In this study, we systematically investigate the timing of rising nuclear pitch accents with landmarks in the acoustics and in relation to both consonantal and vocalic gestures.

2. Method

2.1. Speakers and recordings

We recorded four native speakers of German (S1-S4), all female and aged between 26 and 32 years. All speakers grew up north of the Benrather isogloss.

Kinematic data were obtained with a 3D Electromagnetic Articulograph (Carsten AG 500) at 200 Hz and smoothed with a 3-step floating mean. The acoustic data were recorded with the built-in, time-synchronized microphone and digitized at 16 kHz. Articulatory movements were tracked by sensors on the upper and lower lips as well on the tongue tip, blade and body. We recorded 288 tokens (4 spkrs x 3 target words x 3 phrasal positions x 8 repetitions).

2.2. Speech material and measurements

The speech material consisted of monosyllabic and trochaic disyllabic target words, the latter having either an open or a closed syllable (CV: /ma:/, CV:.CV /ma:mi/ and CVC.CV /mamzi/). All target words were placed in carrier sentences designed to elicit nuclear rising pitch accents in (a) phrase-initial, (b) phrase-noninitial and (c) phrase-final position. Subjects were instructed to read this carrier sentence displayed on a computer screen in a comfortable and natural way. For each syllable structure, we maintained the sequence of stressed and unstressed syllables in the carrier sentence. The following carrier sentences with the target word /ma:/ exemplify our corpus. Brackets indicate prosodic boundaries. Target words are bold and underlined.

(a) [Dann dachte sie:]$_{ip}$[**Ma** mineralisierte das Wasser.]$_{ip}$
(Then she thought: Ma mineralised the water.)
(b) [Die **Ma** mineralisierte das Wasser.]$_{IP}$
(The Ma mineralised the water.)
(c) [Sie sah dann die **Ma**.]$_{IP}$
(Then she saw the Ma.)

Acoustic and articulatory data were labelled manually using the EMU speech database system (Harrington 2010). In the F0 trace, we identified local turning points for L (local minimum) and H (local maximum) in the vicinity of the rising nuclear LH pitch accent. Consonantal gestures were labelled via the Lip Aperture Index (Byrd 2000). We labelled local minima and maxima in the interlip distance trace corresponding to the opening and closing gesture of the bilabial consonants. In addition, we determined the peak velocities of the consonantal gestures taking zero-crossings in their respective acceleration traces. The target of the vocalic gesture was identified by means of a local minimum in the tongue body trace taking a zero-crossing in the velocity derivation.

The following variables (lags) were computed in the acoustic and articulatory domains:

(I) Measures relative to acoustic landmarks (fig. 1):
L-C1ons: Beginning of the nuclear rise relative to acoustic onset of the accented syllable
H-EndSyll End of the rise relative to the acoustic offset of the accented syllable. This landmark refers either to the end of the accented vowel (as in the case of CV: and CV:.CV) or to the end of the coda consonant (in the case of CVC.CV)

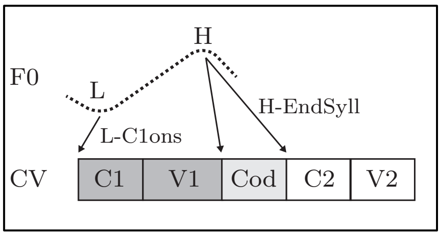

Figure 1: Acoustic landmarks for F0 alignment patterns

(II) Measures relative to articulatory landmarks (fig. 2):
L-minC1: Beginning of the rise relative to the articulatory maximum of the onset's closing gesture (landmark 1)
L-relC1pvel: Beginning of the rise relative to the peak velocity of the onset consonant's release (landmark 2)
H-maxC1: End of the rise relative to the maximal lip opening between C1 and C2 (transvocalic opening, landmark 3)
H-targV: End of the rise relative to the articulatory target of the vocalic gesture (landmark 4)

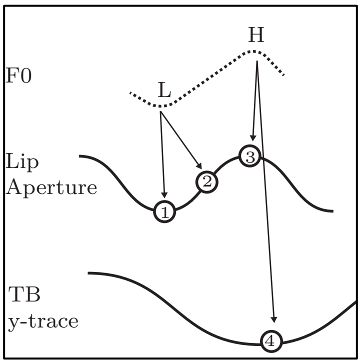

Figure 2: Articulatory landmarks for F0 alignment patterns

3. Results

We conducted an overall rmANOVA (repeated measures ANOVA) based on cell means with POSITION and SEGMENTAL MAKEUP as independent variables and SPEAKER as random factor on the different acoustic and kinematic alignment measures. In case of significant effects, subsets were created and tested pairwise in Bonferroni-corrected post-hoc tests.

3.1. Alignment of L

Figure 3 displays the acoustic alignment of L relative to the beginning of the accented syllable.

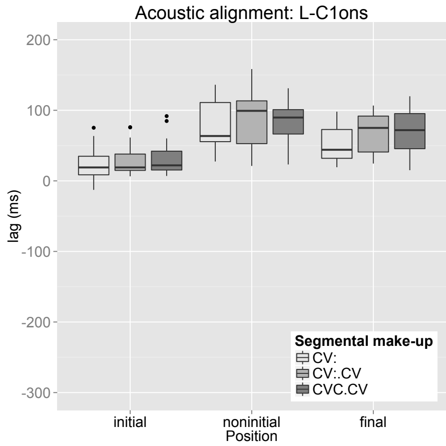

Figure 3: Alignment lags between L and the acoustic onset of C1

L occurs on average of 59 ms after the beginning of the accented syllable. The rmANOVA revealed a small but significant effect of POSITION on the alignment of L relative to the onset of the accented syllable. [$F(2,6) = 7.92$, $p<0.05$ *]. However, subsequent Bonferroni-corrected post-hoc tests did not show any significant differences. The factor SEGMENTAL MAKEUP did not reach significance indicating a stable alignment across our target words.

A similar picture arises in the articulatory domain. Figure 4 displays the articulatory alignment patterns in terms of the L relative the maximum of the onset's closing gesture (L-minC1). L occurs slightly after the consonant's maximum closure. The rmANOVA showed only a small but significant effect of SEGMENTAL MAKEUP on the alignment of L relative to this articulatory landmark [$F(2,6)=6.26$, $p<0.05$ *]. On average, L occurs 27 ms after the maximum constriction of the initial C and 40 ms before the following peak velocity (L-relC1pvel). In terms of articulation, this time window correlates with the release of the consonantal gesture.

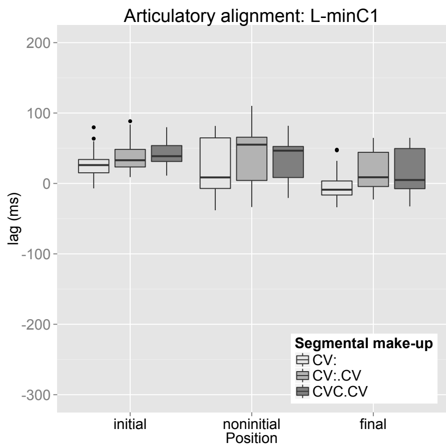

Figure 4: Alignment lags between L and the articulatory maximum of the onset's closing gesture

Table 1 provides means and standard deviations for the alignment of L relative to acoustic and articulatory landmarks.

		Acoustics	Articulation	
position	segmental makeup	L-C1	L-minC1	L-relC1pvel
initial	CV:	24 (23)	26 (21)	-54 (20)
	CV:.CV	29 (21)	39 (22)	-38 (23)
	CVC.CV	32 (22)	43 (18)	-29 (23)
noninitial	CV:	79 (33)	24 (39)	-42 (38)
	CV:.CV	91 (40)	40 (41)	-23 (37)
	CVC.CV	84 (28)	34 (30)	-22 (29)
final	CV:	50 (23)	-2 (29)	-70 (24)
	CV:.CV	68 (29)	18 (29)	-48 (30)
	CVC.CV	70 (30)	15 (31)	-44 (34)

Table 1: Means and standard deviations in parenthesis for the alignment lags of L in ms

3.2. Alignment of H

In contrast to the alignment of L, H shows a considerably amount of variation when investigated relative to segmental landmarks. Figure 5 depicts the alignment of H relative to the end of the accented syllabe which at the same time refers to the end of the vowel in CV: and CV:.CV and to the end of the coda consonant in the case of CVC.CV.

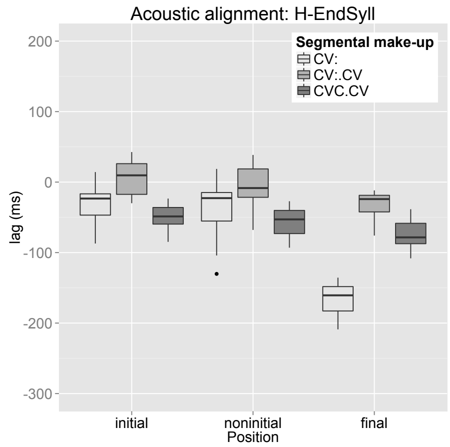

Figure 5: Alignment lags between H and the end of the accented syllable

The rmANOVA revealed a strong effect of both factors and an interaction between them [POSITION: F(2,6)=56.88, p<0.001*, SEGMENTAL MAKEUP: F(2,6)=29.66, p<0.001*, POSITION x SEGMENTAL MAKEUP: F(4,12)=41.53, p<0.001*]. Posthoc tests showed that both the target words and phrasal positions differ from each other significantly. In CV:, the peak was aligned on average 78 ms before the end of the accented syllabe, in CV:CV, it was aligned 7 ms before it and in CVC.CV it was aligned 49 ms before the end of the accented syllable.

Figure 6 depicts the alignment of H relative to the maximum opening of the onset consonant gesture.

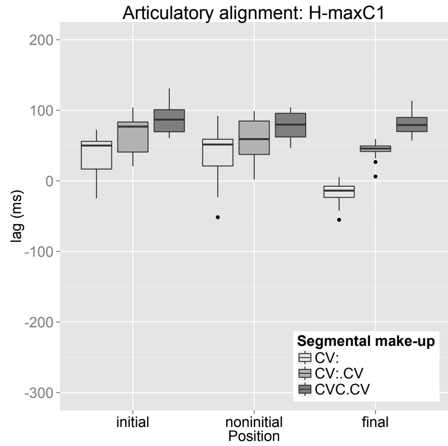

Figure 6: Alignment lags between H and the max opening of the onset consonant (transvocalic gesture)

Again, the rmANOVA revealed an effect of both factors including an interaction [POSITION: F(2,6)=8.81, p<0.01*, SEGMENTAL MAKEUP: F(2,6)=29.34, p<0.001*, POSITION x SEGMENTAL MAKEUP: F(4,12)=22.87, p<0.001*]. Taking these results together, the f0 peak is neither aligned with the end of the accented syllable nor with the transvocalic opening of the consonant.

However, the picture changes when investigating the alignment of the f0 peak relative to the articulatory target of the vocalic gesture in the tongue body trajectory (see figure 7). Here, the effects of segmental makeup disappear. Note that we had to exclude the data from CV: in phrase-final condition as we could not reliably detect the vocalic target in the tongue body trace.

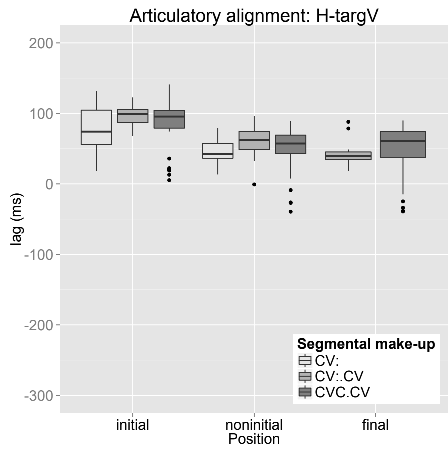

Figure 7: Alignment lags between H and the target of the vocalic gesture

The rmANOVA revealed **no effect** of the segmental make-up but it did reveal an effect of the phrasal position [POSITION: F(2,6)=23.96, p<0.001]: On average, the peak is aligned 86 ms after the vocalic target in phrase-initial, 52 ms in phrase-noninitial and 44 ms in phrase-final condition. To sum up the, we found a stable alignment of H relative to the vocalic gestures. Table 2 summarises our results by providing means and standard deviations for the acoustic and articulatory alignment lags for H.

		Acoustics	Articulation	
position	segmental makeup	H-EndSyll	H-maxC1	H-targV
initial	CV:	-34 (26)	37 (31)	78 (33)
	CV:.CV	6 (23)	65 (26)	98 (14)
	CVC.CV	-48 (15)	87 (18)	84 (36)
noninitial	CV:	-34 (35)	39 (36)	47 (17)
	CV:.CV	-3 (26)	60 (28)	61 (20)
	CVC.CV	-58 (21)	78 (18)	45 (35)
final	CV:	-167 (24)	-16 (15)	
	CV:.CV	-31 (17)	44 (12)	41 (17)
	CVC.CV	-75 (19)	81 (15)	46 (42)

Table 2: Means and standard deviations in parenthesis for the alignment lags of H in ms

4. Discussion and conclusion

In this study, we investigated segmental and positional effects on the alignment of rising nuclear pitch accents in German. We confirmed that the beginning of the nuclear rise (L) exhibits a stable alignment, both to the acoustic onset of the accented syllable and the release of the consonantal gesture. As expected, the end of the rise, the H peak, was highly variable relative to acoustic and transvocalic consonantal landmarks. Both in the acoustic analysis and the analysis of the consonantal gestures, the H peak was consistently aligned later, the greater the amount of segmental material available. However, we found a stable alignment of the f0 peak relative to the target of the vocalic gesture. Thus, we propose that the target of the actual vocalic gesture (rather than the transvocalic gesture) serves as a trigger for the f0 peak.

Our results contribute to the growing body of evidence of a direct relation between laryngeal and supralaryngeal gestures (Gao 2009, Katsika 2012). Utilizing a model of coupled oscillators within the framework of Articulatory Phonology (Browman & Goldstein 1992, Nam & Saltzman 2003) and conceiving pitch accents as tone gestures, it has already been shown that different alignment patterns between languages can be related to underlying coupling structures (Mücke et al. 2012, Niemann et al. 2011). Our results show that not only peak alignment differences found cross-linguistically but also contextually conditioned differences can be explained in terms of gestural coordination.

5. References

Arvaniti, A., Ladd, D.R. & I. Mennen (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics 26*, pp. 3-25.

Browman, C.P. & Goldstein, L. (1992). Articulatory Phonology. An Overview. *Phonetica* 49, 155-180.

Byrd, D. Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica, 57 (1), pp*. 3-16.

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

Gao, M. (2009). Gestural coordination among vowel, consonant and tone gestures in Mandarin Chinese. *Chinese Journal of Phonetics*. Beijing: Commercial Press.

Harrington, J. (2010). *The phonetic Analysis of Speech Corpora*. Wiley-Blackwell.

Katsika, A. (2012). Coordination of prosodic gestures at boundaries in Greek. Ph.D. Thesis. Yale University: New Haven, CT.

Ladd, D. Robert (²2008). Intonational Phonology. Cambridge, MA: Cambridge University Press.

Ladd, D.R., Faulkner, D., Faulkner H. & A. Schepman (1999). Constant „segmental anchoring“ of F0 movements under changes in speech rate. *The Journal of the Acoustical Society of America* 106 (3): pp. 1543-1554.

Mücke, D. & A. Hermes (2007). Phrase Boundaries and Peak Alignment: An Acoustic and Articulatory Study. *Proc. 16th ICPhS, Saarbrücken*, Germany.

Mücke, D., Grice, M., Becker, J., Hermes, A. (2009). Sources of variation in tonal alignment: evidence from acoustic and kinematic data. *Journal of Phonetics 37 (3),* pp. 321-338.

Mücke, D., H. Nam, A. Hermes & L. Goldstein (2012). Coupling of tone and constriction gestures in pitch accents. In: P. Hoole, M. Pouplier, L. Bombien, Ch. Mooshammer & B. Kühnert (Eds.). *Consonant Clusters and Structural Complexity*. Mouton de Gruyter, pp. 205-230.

Mücke, D., M. Grice & A. Hermes (2008). The vowel triggers the tone: Evidence from German. Special Session: Coordination of speech gestures: Models and data. 8th Phonetic Conference of China 2008, Beijing, China.

Nam, H. & Saltzman, E. (2003). A competitive, coupled oscillator of syllable structure. *Proc. 12th ICPhS*, Barcelona, Spain, pp. 2253-2256.

Niemann, H., D. Mücke, H. Nam, L. Goldstein & M. Grice (2011). Tones as Gestures: The Case of Italian and German. Proceedings of the 17th International Congress of Phonetic Sciences, Hongkong, China, pp. 1486-1489.

Prieto, P. & F. Torreira (2007). The Segmental Anchoring Hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics 35*, pp. 437-500.

Silverman, Kim E.A. & Janet Pierrehumbert (1990). The timing of prenuclear high accents in English. In: Kingston, John & Mary Beckman (Eds). *Papers in Laboratory Phonology I: Between the grammar and Physics of Speech*, p. 72-106.

Wichmann, A., J. House & T. Rietveld (2000). Discourse Constraints on F0 Peak Timing in English. In: Botinis, A. (Ed). *Intonation. Analysis, Modelling and Technology*. Dordrecht et al.: Kluwer Academic Publishers, pp. 163-184.