TONES AS GESTURES: THE CASE OF ITALIAN AND GERMAN

Henrik Niemann[a], *Doris Mücke*[a], *Hosung Nam*[b], *Louis Goldstein*[b,c] *& Martine Grice*[a]

[a]I*f*L Phonetik, University of Cologne, Germany;
[b]Haskins Laboratories & [c]University of Southern California, USA
henrik.niemann@uni-koeln.de; doris.muecke@uni-koeln.de; martine.grice@uni-koeln.de;
nam@haskins.yale.edu; louisgol@usc.edu

ABSTRACT

In this paper we investigate tonal alignment in peak accents in Italian and German. We show that timing differences across the two languages are systematic and result from crucial differences in phonological structure. The F0 rise in Italian is represented as a tonal high gesture, the onset of which is synchronised with the vocalic gesture. The German rise is the result of a sequence of tonal gestures, low and high, which compete for alignment with the vocalic gesture, resulting in a delay in the rise. A comparison of accented and deaccented syllables in German shows that the presence of a non-lexical tone does not affect the timing of consonantal and vocalic gestures.

Keywords: tonal alignment, competitive coupling, tonal gesture, pitch accent

1. INTRODUCTION

Tonal alignment with segments in the acoustic record – segmental anchoring - has been the subject of a great deal of research over the last 13 years. Arvaniti, Ladd & Mennen [1] showed for rising pitch accents in Greek that turning points in the F0 contour are consistently aligned with landmarks in the segmental string, providing evidence for an autosegmental-metrical levels-based analysis of intonation. Subsequent studies have revealed differences in alignment across and within languages, and have fired debate as to whether such differences should be accounted for in terms of phonological structure or simply as phonetic detail.

In peak accents it is often the F0 maximum, corresponding to a H tone that is seen to be primary, encoded through the * notation (H*), and treated as a strong node in a branching structure [8]. However, studies on segmental anchoring often find that the F0 minimum (the *start* of the F0 rise) is more stable than the F0 maximum.

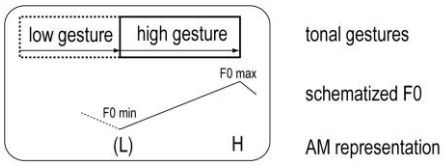
Mücke, et al. [6] explored the possibility of modeling tonal alignment by treating peak accents as *tonal gestures*, analogous to vocalic and consonantal gestures in the Articulatory Phonology framework ([2, 5, 7]). They compared the realizations of one speaker each of Catalan and Viennese German. In their analysis, both types of pitch accent had two tonal gestures, a low followed by a high. Rises in Catalan were relatively early, and involved a direct coupling between the high gesture and the vocalic gesture – they both started simultaneously. The late rise in Viennese German was the result of competition between the low and high tonal gestures for alignment with the vowel, delaying the onset of the second tonal gesture, the high gesture.

In this study, we explore the coordination of tonal and oral constriction gestures in a further variety of German and Italian, a language reported to have an earlier alignment than German. A further goal is to explore whether competition between two tonal gestures affects the coordination between vocalic and consonantal gestures.

1.1. The tonal gesture

A tonal gesture is defined as an articulatory action to achieve a desired tonal goal in F0 space [4, 6] Figure 1 provides a schematized F0 contour for a rising pitch accent represented both as a tonal gesture and as autosegmental-metrical tones. In a rising intonational pitch accent, a high gesture involves a tonal movement to an H target in F0. The onset of a tonal gesture is taken to be the point in time at which F0 begins to move in the direction of that gesture's pitch target.

Figure 1: Levels of representation for tonal gestures.



2. METHOD

2.1. Speakers and recordings

We recorded two native speakers of Italian (MG, Bologna, MS, Bari) and three native speakers of Standard German (CR, MK, KS, Wesel, Lower Franconian), all female, and aged 24-46 years.

The kinematic data were recorded with a 2D Electromagnetic Articulograph (Carstens AG100) at 500 Hz, downsampled to 200 Hz and smoothed with a 40

Hz low-pass filter. The acoustic data were recorded with a time-synchronized DAT-recorder digitized at 44.1 kHz.

Lip and tongue movements were tracked by sensors placed on the vermillion border of the upper and lower lip and sensors on the tongue tip, tongue blade and tongue body. Subjects were instructed to read the speech material displayed on a computer screen in a comfortable and natural way.

We recorded 240 German tokens (3 speaker x 2 accent type x 4 target words x 10 repetitions) and 80 Italian tokens (2 speakers x 4 target words x 10 repetitions).

2.2. Speech material and analyses

The speech material consisted of trochaic target words with a CVCV structure. Consonants were either labial or alveolar nasals or liquids, and vowels alternated in height, see table 1.

Table 1: Italian and German speech material.

	/i/ condition	/a/ condition
Italian	Nina Lina	nani lami
German	Nima Lina	Nami Lani

All target words were embedded in carrier sentences ensuring alternation of low and high vowels throughout the whole utterance. The carrier sentences were designed to elicit a pitch peak on the target words, see (1) for Italian and (2) for German. In addition, the target words were recorded in a deaccented condition in German, see (3).

(1) *Italian acc.:* 'Per favore dimmi la___di nuovo.'
(Please say ___ again.)

(2) *German acc.*: 'Er geht mit der ___ viel lieber.'
(He goes with ___ preferentially.)

(3) *German deacc.*: 'Er **geht** mit der ___ viel lieber.'
(He **goes** with ___ preferentially.)

Acoustic and articulatory data were labelled manually using the EMU speech database system. In the F0 trace we identified local minima and maxima for the rise up to the accentual peak.

For the articulatory landmarks, we identified minima and maxima in the vertical position by zero-crossings in their respective velocity traces. For consonantal gestures we used the trajectories of the upper and lower lip (Lip Aperture index, [3]) and the tongue tip. To capture the vocalic gestures we used the tongue body trajectories.

The following variables were computed:

L-C1 Lag: Acoustic measure. The start of the F0 rise relative to the beginning of the initial C segment (accented syllable).

CV Lag: Articulatory measure. The start of the consonantal gesture relative to start of the vocalic gesture. This variable reflects the intrasyllabic CV coordination.

TV Lag: Articulatory measure. The start of the tone gesture (=the start of the F0 rise) relative to the start of the vocalic gesture.

3. RESULTS

3.1. Italian alignment data

Figure 2 shows medians and quartiles for the *acoustic* L-C1 lags for all target words. The zero-line marks the beginning of the accented syllable, negative values indicate that the start of the F0 rise occurs *before* the start of the accented syllable.

Figure 2: Acoustic data: L-C1 Lag, Italian.

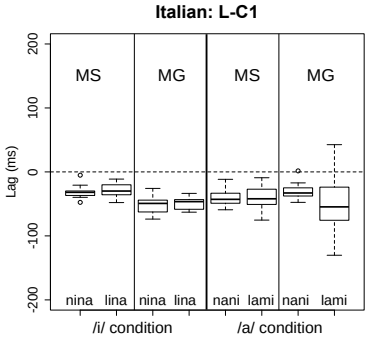


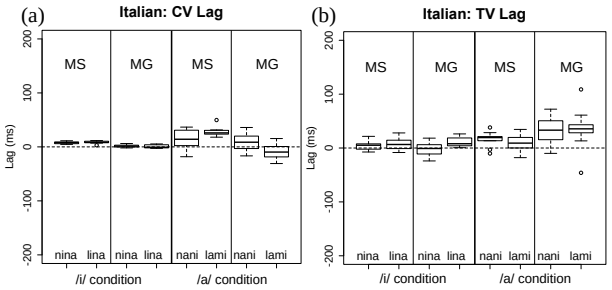
Table 2 provides the means and standard deviations of the acoustic and articulatory lags for all target words. The tone precedes the start of the accented syllable in all investigated conditions (on average -40 ms in the /i/ condition and -41 ms in the /a/ condition). In a series of one-way ANOVAs conducted for each speaker separately the effect of VOWEL HEIGHT revealed no significance (Bonferroni correction $\alpha' = 0.0083$ for multiple comparisons).

Table 2: Mean lags (in ms) and standard deviations for the acoustic and articulatory measures

		Acoustic	Articulation	
	Speaker	L-C1	TV	CV
/nina/	MG	-52 (14)	-1 (12)	2 (3)
	MS	-31 (11)	5 (8)	8 (2)
/lina/	MG	-49 (9)	11 (9)	0 (3)
	MS	-29 (11)	8 (11)	9 (2)
/nani/	MG	-30 (15)	32 (24)	10 (17)
	MS	-41 (14)	17 (14)	13 (20)
/lami/	MG	-50 (52)	35 (39)	-10 (14)
	MS	-41 (20)	10 (15)	28 (9)

Figure 3 displays boxplots for the *articulatory* measures (CV and TV lags). The gestures occur in the order C-V-T, but all lags were very small: In the /i/ condition the tone gesture was on average 6 ms after the V gesture, which was in turn 5 ms after the C gesture. The factor VOWEL HEIGHT revealed significance on the CV measure in speaker MS ([$F(1, 40) = 11.697$, $p \leq 0.001$]), but not in speaker MG.

Figure 3: Articulatory data: (a) CV Lag, (b) TV Lag.



To sum up, in the Italian data the F0 minimum considerably leads the *acoustic* anchor. By contrast, in articulatory terms it is practically synchronous with both the consonantal and the vocalic gestures in all conditions. We found an effect of VOWEL HEIGHT in the acoustic data and to some extent in the articulatory data.

3.2. German alignment data

Figure 4 displays the acoustic L-C1 lags for all target words. The start of the rise occurs considerably *after* the accented syllable, on average 54 ms in the /i/ condition and 108 ms in the /a/ condition.

Figure 4: Acoustic data: L-C1 Lag.

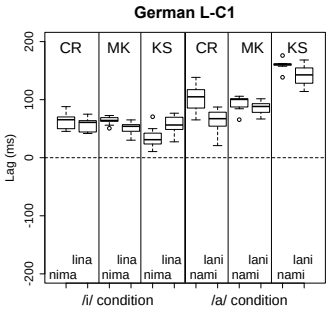


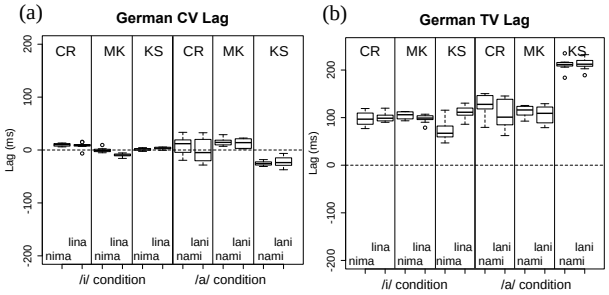
Table 3: Mean lags (in ms) and standard deviations for the acoustic and articulatory measures.

		Acoustic	Articulation	
	Speaker	L-C1	TV	CV
/nima/	CR	63 (14)	98 (14)	10 (2)
	MK	64 (6)	105 (7)	0 (4)
	KS	34 (19)	73 (22)	1 (2)
/lina/	CR	57 (11)	101 (10)	8 (6)
	MK	51 (10)	98 (8)	-10 (3)
	KS	56 (16)	110 (15)	3 (2)
/nami/	CR	104 (23)	126 (22)	9 (17)
	MK	94 (12)	114 (11)	15 (7)
	KS	160 (11)	211 (15)	-25 (5)
/lani/	CR	63 (21)	105 (30)	-1 (21)
	MK	85 (11)	108 (17)	12 (8)
	KS	142 (17)	213 (12)	-23 (11)

Table 3 provides means for the acoustic and articulatory measures. In contrast to the Italian data, the rise occurs on average of 54 ms later in the /a/ condition compared to the /i/ condition. One-way ANOVAs (Bonferroni correction α'=0.0055) revealed a main effect of VOWEL HEIGHT on L-C1 (CR [F(1, 40) = 10.321, p≤ 0.002]; MK [F(1, 40) = 79.547, p<0.001]; KS [F(1, 40) = 262.68, p<0.001]).

In the articulatory domain (figure 5), the F0 minimum was very late in relation to the onset of the vocalic gesture (98 ms in the /i/ condition and 147 ms in the /a/ condition), while consonantal and vocalic gestures were tightly synchronised (2 ms apart in the /i/ condition and -2 ms in the /a/ condition). When comparing the means in table 3 for the articulatory TV and CV measures, the tonal high gesture starts later in the /a/ condition than in the /i/ condition.

Figure 5: Articulatory data: (a) CV Lag, (b) TV Lag.



We found a main effect of VOWEL HEIGHT on the TV measure for speaker KS ([F(1, 40) = 283.31, p<0.001]) and for two out of three speakers on the CV measure (MK: [F(1, 40) = 75.904, p<0.001], KS [F(1, 40) = 155.70, p<0.001]).

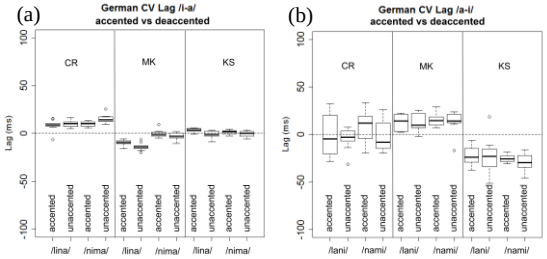
To sum up, in the German data the F0 minimum occurs considerably after the start of the accented syllable on the acoustic surface. In the kinematic data, this minimum is also substantially delayed, while C and V gestures are still synchronised in relation to each other. There was more variability in the timing of all gestures in the /a/ condition than in the /i/ condition, especially for speaker KS. Effects of VOWEL HEIGHT were strong but not systematic. Phonetic variation is stronger in the /a/ condition than in the /i/ condition.

3.3. Accented vs. deaccented word

In German, we compared the CV coordination in target words with a pitch accent (accented) and without a pitch accent (deaccented). There is a tight synchronisation of the CV gestures with respect to each other, in the accented condition (/i/: 2 ms, /a/: -2 ms) as well as in the deaccented condition (/i/: 1 ms, /a/: -6 ms), see figure 6.

An overall repeated measure ANOVA (two-way, 2x2) based on cell means reveals no effect of PITCH ACCENT [F(1, 24) = 0.168, p>0.05] and VOWEL HEIGHT [F(1, 24) = 0.952, p>0.05] on the CV coordination. The presence of the pitch accent does not influence the intrasyllabic timing.

Figure 6: German CV coordination, accented vs. deaccented, (a) /i/ condition and (b) /a/ condition.

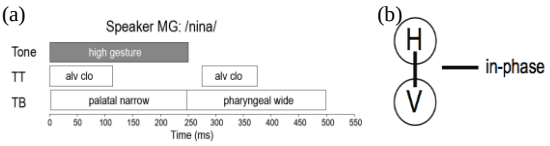


4. DISCUSSION

Despite the fact that the Italian speakers have different regional accents, their tonal alignment in this particular pitch accent is strikingly similar - and consistently different from the three German speakers. We can account for the differences between the two languages (even though they are relatively subtle) in terms of phonologically driven *coupling relations* between tonal and vocalic gestures.

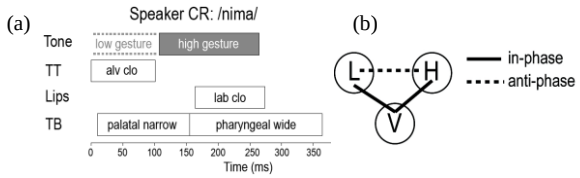
In figure 7(a) the accentual rise in Italian is modeled as one tonal gesture (high gesture), which starts at the same time as the vocalic gesture (palatal narrow) for the accented syllable. The tonal gesture is thus coupled in-phase (7b) with the vowel, leading to a relatively early rise of the pitch accent in the acoustics.

Figure 7: (a) Gestural score and (b) coupling graph for the Italian speaker MG, target word /nina/.



The accent in German, on the other hand, is modeled with two tonal gestures (low gesture, high gesture), which compete with each other for synchronization with the vowel (palatal narrow), see figure 8(a); they are both in phase with the vowel, but at the same time sequenced in relation to each other (8(b)), leading to an earlier low and a later high gesture. On the acoustic surface, this corresponds to a later F0 minimum (corresponding to the onset of the tonal high gesture).

Figure 8: (a) Gestural score and (b) coupling graph for the German speaker CR, target word /nima/.



The alignment patterns in Italian appear to be similar to those in Catalan [6]. However, our analysis of Italian is different in that it does not involve a low tonal gesture preceding the high gesture. The F0 minimum is taken to be simply the onset of the high gesture rather than a target in its own right. This analysis was influenced largely by the fact that the F0 minimum was not low in the speakers' range in Italian, whereas it was in German (rise excursion in German 6 semitones and in Italian 2 semitones).

Another outcome of this study is that, as far as German is concerned, we found no effect of accentuation – the presence or absence of tonal gestures – on CV coordination. This contrasts with Gao's [4] modeling of lexical tones in Mandarin Chinese, in which tonal gestures did affect CV coordination. Our results confirm that intonational tones (which are post-lexical in nature) cannot modify the intrasyllabic coupling relations that define that syllable, as suggested in [6].

5. REFERENCES

[1] Arvaniti, A., Ladd, D.R., Mennen, I. 1998. Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics* 26, 3-25.

[2] Browman, C.P., Goldstein, L. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée* 5, 25-34.

[3] Byrd, D. 2000. Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* 57(1), 3-16.

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

[5] Goldstein, L., Chitoran, I., Selkirk, E. 2007. Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber. *Proc. 16th ICPhS* Saarbrücken, 241-244.

[6] Mücke, D., Nam, H., Hermes, A., Goldstein L. (to appear). Coupling of tone and constriction gestures in pitch accents. *Consonant Clusters and Structural Complexity*. Munich, Germany.

[7] Nam, H., Saltzman, E. 2003. A competitive, coupled oscillator model of syllable structure. *Proc. 15th ICPhS*, Barcelona, 2253-2256.

[8] Pierrehumbert, J., Beckman M. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.