TONES AS GESTURES: THE CASE OF ITALIAN AND GERMAN

Henrik Niemann\textsuperscript{a}, Doris Mücke\textsuperscript{a}, Hosung Nam\textsuperscript{b}, Louis Goldstein\textsuperscript{b,c} & Martine Grice\textsuperscript{a}

\textsuperscript{a}IfL Phonetik, University of Cologne, Germany; \textsuperscript{b}Haskins Laboratories & \textsuperscript{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 synchronized 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 CVCC 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            | nani             |
|                 | Lina            | lami             |
| German          | Nima            | Nami             |
|                 | Lina            | Lami             |

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 ___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.

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

<table>
<thead>
<tr>
<th></th>
<th>Acoustic</th>
<th>Articulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Speaker</td>
<td>L-C1</td>
</tr>
<tr>
<td>/nina/</td>
<td>MG</td>
<td>-52 (14)</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>-31 (11)</td>
</tr>
<tr>
<td>/lami/</td>
<td>MG</td>
<td>-49 (9)</td>
</tr>
<tr>
<td>/nani/</td>
<td>MS</td>
<td>-29 (11)</td>
</tr>
<tr>
<td></td>
<td>MG</td>
<td>-30 (15)</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>-41 (14)</td>
</tr>
<tr>
<td></td>
<td>MG</td>
<td>-50 (52)</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>-41 (20)</td>
</tr>
</tbody>
</table>

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 ≤ 0.001]), but not in speaker MG.
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.

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.

3.3. 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.
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.

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).

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