Tonal alignment, scaling and slope in Italian question and statement tunes

Mariapaola d'Imperio

TONAL ALIGNMENT, SCALING AND SLOPE

IN ITALIAN QUESTION AND STATEMENT TUNES

Mariapaola D'Imperio

Abstract

Unlike in languages such as English and Standard Italian, the Italian spoken in Naples shares with other Southern varieties the use of a very similar rising-falling (LHL) tune for both yes/no questions and narrow focus statements (D'Imperio, 2001; Grice et al., to appear). However, it has been claimed that the temporal alignment of the accent peak is later in the pitch accent of yes/no questions (L+H) than in that of statements (L+H*) and that this alignment difference is employed by native speakers to perceptually identify the two tunes (D'Imperio and House, 1997). This study acoustically tested the hypothesis that all three tonal targets of the rise-fall are timed and scaled differently in questions and statements. Moreover, slope differences for both rise and fall were also tested by employing logistic regression modeling. Two speakers of Neapolitan Italian produced utterances whose target words differed in question/statement modality, syllable structure and segmental environment. The results show that all three targets within the rise-fall are timed later in questions than in statements. By contrast, no systematic difference was found for the slope of the rise nor for the slope of the fall. The exact contribution of* F_0 *height to signaling the contrast could not be determined, though. In fact, while one speaker marked the difference by producing higher peaks for statements, the other did not produce any difference.*

Keywords : prosody, tonal alignment, Italian, slope, questions.

Résumé

Contrairement à certaines langues comme l'anglais et l'italien standard, l'italien parlé à Naples partage avec d'autres variétés du Sud de l'Italie un patron mélodique ascendant/descendant (LHL) très similaire pour les questions oui/non et pour les affirmations à focalisation étroite (D'Imperio, 2001 ; Grice et al., *à paraître).*

Néanmoins, il a été proposé que les pics d'accent mélodique des questions oui/non (L+H) sont alignés temporellement plus tard que ceux des affirmations (L+H*) et que cette différence d'alignement est employée par les auditeurs de langue maternelle napolitaine pour identifier perceptuellement les deux modalités (D'Imperio and House, 1997). Cette étude teste au niveau acoustique l'hypothèse que les trois cibles tonales de la montée/descente sont alignées temporellement différemment et sont produites sur des échelles différentes pour les questions et les affirmations. De plus, les différences de pente des montées et des descentes ont aussi été analysées grâce à un modèle de régression logistique. Deux locuteurs napolitains ont produit des phrases dont les mots cibles, la structure syllabique et l'environnement segmental différaient pour la modalité question et la modalité affirmation. Les résultats montrent que les trois cibles dans les montées/descentes sont alignées plus tard pour les questions que pour les affirmations. Par contre, aucune différence systématique n'a été mise en évidence pour les pentes des montées ou pour les pentes des descentes. La contribution exacte de la hauteur de la* F_0 *au contraste question/affirmation n'a pu être déterminée. En effet, alors qu'un locuteur marquait le contraste en produisant des pics plus élevés pour les affirmations, l'autre locuteur n'a produit aucune différence significative entre les deux modalités.*

Mots-clés : prosodie, alignement tonal, italien, pente, questions.

D'IMPERIO, Mariapaola, (2002), Tonal alignment, scaling and slope in Italian question and statement tunes, *Travaux Interdisciplinaires du Laboratoire Parole et Langage,* vol. 21, p. 25-43.

1. Introduction

Tonal targets can be defined according to two dimensions, i.e., *alignment* (their temporal location relative to the segments in the string) and *scaling* (their F_0 value). Recent work on alignment has shown that it is affected by a number of factors in a variety of languages (Silverman and Pierrehumbert 1990, Arvaniti *et al.* 1998, Ladd *et al.* 2000). However, it appears that under controlled conditions tones are timed to cooccur quite systematically with specific segments in the string. These regularities appear also to be language-specific and subject to phonological constraints proper to the language under scrutiny. For instance, the findings in Ladd *et al.* (2000) suggest that vowel duration, as a consequence of phonological length, affects peak alignment in Dutch rises. Moreover, Arvaniti *et al.* (1998) found that Greek prenuclear peaks are earlier relative to the onset of the stressed syllable when the onset of the postaccentual syllable is a nasal (as opposed to a stop or a fricative). Hence, alignment has been claimed to be one of the defining properties of "accent identity", while other more holistic features, such as accent shape, duration and slope, would not be controlled by the speaker, hence they would not be *specified*.

Target alignment can be then employed categorically with the purpose of contrasting pitch accent categories. While yes/no questions of most Northern and Central varieties of Italian are characterized by a terminal rise, i.e., a L* nuclear accent followed by a H- rising phrase accent, Southern varieties exhibit a rising (LH) pitch accent on the nuclear accented syllable followed by a later fall. In the Neapolitan variety, this tune has been recently analyzed as a combination of a L*+H accent followed by a HL- phrase accent (D'Imperio 2000, Grice *et al.* to appear). In most cases, this configuration is acoustically realized as a sequence of three tonal targets, LHL, due to "merging" of the H tone sequence in nuclear position. This rise-fall tune is similar to that of narrow focus statements of the same variety, though the alignment of its H peak is later, and the pitch accent is transcribed as a L*+H. Moreover, peak alignment appears to affect the identification of questions and statements (D'Imperio and House, 1997). Hence, it appears that question H peaks are, *ceteris paribus*, later than statement peaks. Since the accent rise is a unitary event, we expect that a later peak will correspond to a later preceding L (henceforth L1). Also, given the analysis of the HL- phrasal fall (which is the same in questions and statements), we also expect that the target for the second L of the LHL sequence (henceforth L2) will be realized later in questions than in statements. This is because the HL- starts at the H target location for the pitch accent rise.

An additional hypothesis tested here is that tonal targets are timed to occur at a specific, absolute location in the utterance (e.g., at a certain temporal distance after vowel or syllable

onset). This predicts that peaks will be invariantly produced at such locations, independent of structurally-dependent vowel duration differences, such as syllable structure (open vs. closed syllable), and segmental environment. Alternatively, if target alignment is proportional to the entire vowel and/or syllable duration, target location is expected to differ as a consequence of these factors.

Moreover, scaling of F_0 values appears to be affected by the question/statement distinction in a variety of languages, with questions presenting usually higher peaks and/or higher global range (Jun and Oh, 1996). This distinction seems also to be relevant for the perceptual identification of questions and statements. For instance, the higher peaks of Hungarian questions have been reported to help listeners to perceptually identify them (Gósy and Terken, 1994). Therefore, we also tested the hypothesis that Neapolitan questions are characterized by higher F_0 peaks. Additionally, from impressionistic observations, it appeared that, apart from the H peak, both L1 and L2 are higher in questions. Hence, in addition to peak F_0 level, L1 and L2 F_0 values were also analyzed. Finally, according to a strict autosegmental-metrical view (Ladd, 1996), slope gradient within a rise or a falling F_0 contour should be irrelevant and entirely dependent on the timing and scaling characteristics of the tonal targets involved. Such a hypothesis was tested here by measuring the slope of both LH rise (from L1 to H) and HL fall (from H to L2).

2. Methods

2.1. Corpus

The corpus consisted of a group of sentences in which modality (question or statement; QS henceforth), structure of the stressed syllable and segmental environment within the target word were varied. The stressed syllable within the target words could either be closed or open (Open/Closed, henceforth). Closed syllables within the target word were always closed by a nasal. In order to manipulate possible segmental effects induced by the onset of the postaccentual syllable, target syllables could either be followed by a nasal or a stop postaccentual onset (Nasal/Stop factor). The above combinations resulted in a corpus of 8 target words, shown in Table 1. Target stressed syllables were always penultimate and the word was embedded as the direct object in a fixed carrier sentence *Vedrai il [...] dopo* "You will see the [...] afterwards", both as a statement and as a question, such that focus scope was always narrow over the object.

OPEN SYLLABLE/NASAL	OPEN SYLLABLE/STOP
nano "ninth"	*mago* "magician"
CLOSED SYLLABLE/NASAL	**CLOSED SYLLABLE/STOP**
mammo "man-mother"	*mango* "mango"

Table 1
Target words for the corpus

In all cases, the test word was a noun, followed by an adverb. The expected stress pattern was a nuclear LH accent on the target word, followed by a fall. Specifically, a L*+H was produced as the nuclear accent of yes/no questions and a L+H* as the nuclear accent of narrow focus statements, as expected. Two Neapolitan speakers, one male (LD) and one female (MD), produced each target sentence. The speakers were both brought up in Naples and spoke Standard Italian with a Neapolitan accent. In order to ensure the statistical validity of the results, given the relatively low number of participants, each of the 8 target utterances were repeated 10 times, for a total of 160 utterances[1].

The recordings were made, for MD, in a double-walled soundproof booth in the Phonetics Lab of The Ohio State University. LD was recorded in a studio at the University "Federico II" of Naples. Both subjects used a head-mounted microphone recording into a Marantz tape recorder. The sentences were written on cards, whose order was randomized. The cards were read one time each. The subjects spoke at a normal pace and as naturally as possible. Misread or disfluent utterances were repeated. The recordings were then digitized at 16 bit, 16 kHz using ESPS Waves+ on a SUN Sparc 10 station. Individual sentences were stored in separate files, and F_0 was later extracted at every ms using the *get_f0* ESPS pitch tracker. This utility uses an auto-correlation method of F_0 estimation and dynamic programming to adjust for local intensity variations and F_0 trends.

2.2. Acoustic measurements

The F_0 tracks were inspected in combination with waveform and spectrogram for each utterance. Both duration and F_0 measurements were performed. Specifically, within the rise-fall

1. Note, though, that the main goal of this study was to obtain a good range of values within a speaker, therefore the number of participants was not a concern.

configuration, the F_0 value in Hz was measured at the first L target (L1), at the H target and at the second L target (L2). In most cases, a peak value corresponding to the H was easy to locate. The H target was in fact measured at the single highest F_0 point within the accented syllable and was labeled "F_0max" (see Figure 2, for instance). In some cases, instead of a well-defined peak, the accented syllable showed a short F_0 plateau. In these cases, the H target was measured as the point halfway through the plateau, whose boundaries were marked as "pl1" and "pl2". Such points were determined as the first and the last high F_0 points within a series of points with very similar value[2]. We did not know whether this location would correspond to a production or a perceptual target, but it ensured consistency of measurement. Plateaus were common both in question and statement utterances of this corpus (29 plateaus in questions and 35 in statements were globally measured).

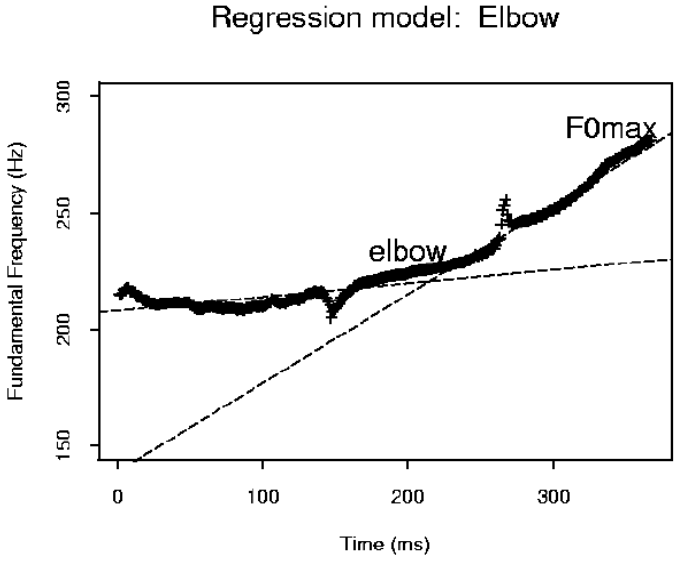

Figure 1
F_0 trace with fitted lines that intersect at the elbow time location

2. Similarity was determined in base of a criterion by which the F_0 value for such points could not differ more than +/- 2 Hz relative the previous point.

The measurement of L1 and L2 proved to be more challenging. L1 was particularly difficult to discern in statements because of the frequent lack of a visible trough at either accented syllable onset or vowel onset. Therefore, L1 and L2 were estimated by means of linear regression models. Specifically, an automatic procedure (D'Imperio, 2000) was employed by which two straight lines were fitted to the F_0 segment going from "b" to F_0 max (see Figure 2). The parameters of the two linear models were estimated by means of conventional linear least-squares methods. To estimate the elbow position, i.e., the intersection of the two fitted lines, two linear regressions were computed for each possible elbow location (from 1 to n, where n is the number of samples within the F_0 segment). The location eventually selected as the "elbow" was the one leading to the smallest total modeling error. The elbow was then automatically inserted in the "label file" of a specific utterance, at the location corresponding to the x-intersection of the fitted lines. The same procedure was also employed in order to locate the end of the fall from the measured peak (L2). This location was labeled "elbow2" (see Figure 2).

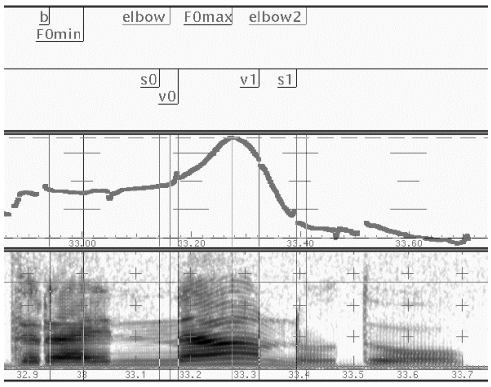

Figure 2

Tone labels, F_0 trace and spectrogram for one of the statement utterances of the corpus, showing duration measurements (s0, v0, etc.) as well as F_0 measurements (Fomin, Fomax, etc.). The target word here is nano.

Duration measurements were made from waveforms in combination with spectrograms following standard criteria of segmentation. We marked locations for the accented syllable onset (s0), the stressed vowel onset (v0), the postaccentual vowel onset (s1) and stressed vowel offset (v1). Such marks are shown in Figure 2. Boundaries between laterals and nasal (e.g., i*l n*ano) were marked at locations where sudden changes in amplitude and formant structure occurred. Vowel onset was marked at the start of high amplitude periodicity. Since one of the target words *mammo* contained a geminate, the problem arose of where to place the syllable boundary. Because of the lack of an uncontroversial criterion, apart from arbitrarily cutting at the geminate halfpoint, it was decided that s1 would be placed at the onset of voicing of the following vowel (within the postaccentual syllable), after the nasal burst. For consistency sake, s1 was always placed at this location also in the other target words. In other words, the stressed syllable could be made, in the simplest case, of a CVC sequence (i.e., *nan*o and *mag*o), a CVCC sequence where the postvocalic CC sequence was a geminate (e.g., *mamm*o) or a CVCC sequence where the postvocalic CC was a cluster (i.e., *mang*o). Figure 2 illustrates how the various labels were placed.

Duration and latency measurements included (i) stressed vowel and syllable duration; (ii) the distance of elbow1 (L1) from both v0 and s0; (iii) the distance of Fomax (H) from v0 and s0, as well as from v1 and (iv) the distance of elbow2 (L2) from v1, s1 and Fomax. The measurements are schematized in Figure 3.

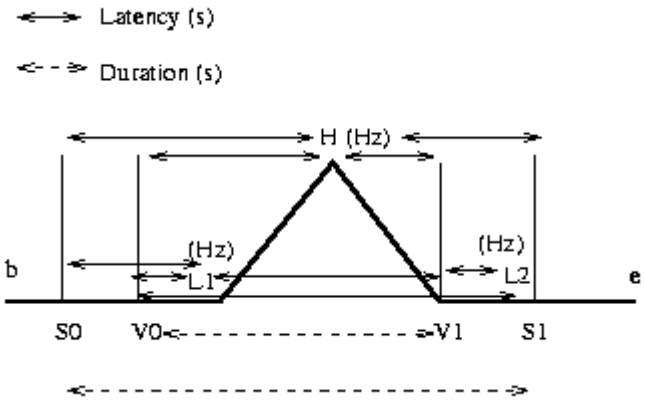

Figure 3
Schematized version of the measurements performed

At a first approximation, it appeared as if questions were characterized by a shallower rise than statements, since their peak is reached later and spans the entire stressed vowel. Such an impressionistic remark was already made in (D'Imperio and House, 1997) and was invoked speculatively to account for the intrinsic difference in the results for their declarative-base and interrogative-base stimuli. In that experiment, all things being equal, the interrogative-base stimuli produced more question responses than the declarative-base stimuli. The reason for such a discrepancy was then attributed to the difference in rise slope, which had not been controlled for, and was steeper for declarative base stimuli. Therefore, rise and fall slope within the LHL configuration were calculated here through various means, in order to find out if systematic differences between questions and statements could be discerned. We observed that the F_o segments for the LH rise and the HL fall present a shape that can be modeled in terms of a logistic curve, which was then employed to obtain slope values. Specifically, to estimate the peak velocity of an F_o segment (either the LH rise or the fall within the LHL configuration), a logistic model was fitted to it:

$$F_0(t) = c + \frac{a}{1 + e^{-b(t-t_0)}}$$

where t stands for time, to for the temporal coordinate of the curve inflection point, and a, b and c are the parameters of the model. The peak velocity was defined by the slope of the model at its inflection point, i.e., the first derivative of the model measured in t_o. The first derivative of the model is:

$$\frac{dF_0(t)}{dt} = \frac{abe^{-b(t-t_0)}}{\left[1 + e^{-b(t-t_0)}\right]^2}$$

The peak velocity was therefore estimated by ab/4, again expressed in Hz/ms. This model resulted in a very good fit to the data, as can be seen in Figure 4. This model is much less sensitive to noise in the data relative to other methods (D'Imperio, 2000). In order to check the modeling results, visual inspection of each fitting was carried out. Note also that a simple straight line fitting would not yield satisfactory results, given the particular shape of the F_o segments under investigation. Consequently, a linear fitting would crucially obtain higher modeling error values than a logistic fitting.

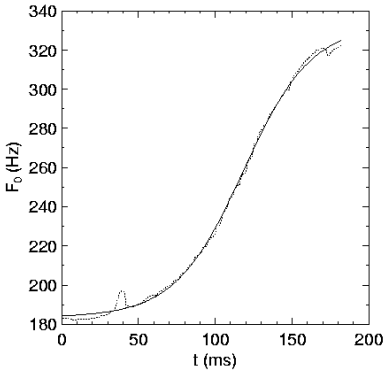

Figure 4
Fundamental frequency curve (dotted line) and logistic curve (solid line) fitted to the LH rise within a question produced by MD. The perturbation corresponds to the nasal-vowel edge

3. Results

3.1. Temporal Alignment

In Figure 5, mean latency from vowel onset and F_0 value for L1, H and L2 are plotted separately for questions and statements, for each speaker. Note that for both speakers all three targets were realized earlier in statements than in questions.

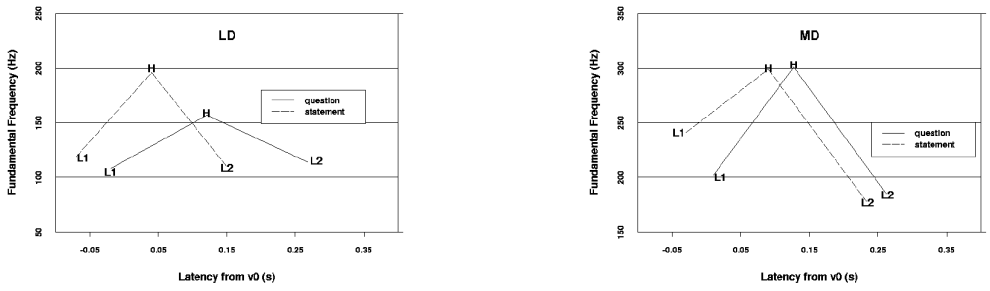

Figure 5
Fo values and latency from vo (vowel onset) for L1, H and L2 for speaker LD (upper) and MD (lower)

For MD, whose rise-fall configuration was globally later than for LD, L1 was located just before the stressed vowel onset in statements and at or soon after the same location for questions. For

LD, L1 was located quite before the stressed vowel onset for statements, while it was located immediately before it in questions. The H peak tended to be timed with the offset of the stressed vowel for MD questions, while it was timed 36 ms before vowel offset in LD. In statements, H was timed to occur around the 2/3 of the vowel for MD, and closer to vowel onset for LD. Question peaks were aligned 37 ms later than statements for MD and 81 ms later for LD. While for MD L2 tended to occur at the postaccentual vowel onset for statements and around the middle of the same vowel for questions, LD differentiated L2 target alignment more drastically between the two modalities. Namely, in statements, L2 was aligned with stressed vowel offset, while in questions it was timed with the onset of the postaccentual vowel.

First, let us look closely at the results for the elbow latency relative to vowel onset. As shown in Figure 5, L1, measured in term of the elbow location, was earlier in statements (-0.07 s for LD and -0.03 s for MD) than in questions (-0.02 s for LD and 0.01 s for MD). In other words, the elbow was circa 50 ms earlier in statements for LD, and circa 40 ms earlier for MD. The statistic results for L1 alignment showed that that the QS factor affected significantly the means when the latency was measured relative to vowel onset [$F(1,144) = 68.8$; $p = 0$]. Also, the Nasal/Stop factor reached significance [$F(1,144) = 7.45$; $p<0.01$] as well as Speaker [$F(1,144) = 55.6$; $p = 0$] while Open/Closed [$F(1,144) = 0.4$; $p = 0.84$] did not. None of the interactions was significant.

Analogous results were found regarding the latency of H (F0max) relative to vowel onset (HtoVons). Specifically, the latency was smaller for statements (0.04 s for LD and 0.09 s for MD) than for questions (0.121 s for LD and 0.127 s for MD). Here, QS was again significant [$F(1,144) = 490$; $p = 0$], while Open/Closed was not [$F(1,144) = 5.9$; $p = 0.02$]. Also, Nasal/Stop was not significant [$F(1,144) = 0.8$; $p = 0.4$], while Speaker was [$F(1,144) = 122$; $p = 0$]. Among the interactions, only QS by Speaker [$F(1,144) = 77.4$; $p = 0$] and QS by Open/Closed by Nasal/Stop [$F(1,144) = 13.5$; $p<0.01$] were significant. In fact, at least for MD, the latency was smaller for closed statements (0.088 s) than for open statements (0.097 s), while for LD there was a small difference in the opposite direction (0.043 s for closed statements vs. 0.039 s for open statements). On the other hand, HtoVons latency was smaller for closed questions (0.117 s for LD and 0.121 s for MD) than for open questions (0.125 s for LD and 0.133 s for MD), and this was true for both speakers.

When measured relative to syllable onset (HtoSons), only QS [$F(1,144) = 295$; $p = 0$] and Speaker [$F(1,144) = 33.5$; $p = 0$] were significant, while neither Open/Closed [$F(1,144) = 0.74$; $p = 0.4$] nor Nasal/Stop [$F(1,144) = 2.6$; $p = 0.1$] were. The only significant interaction was QS by Open/Closed by Nasal/Stop [$F(1,144) = 7.5$; $p<0.01$]. Therefore, it appears that the only strong effect on H

alignment, when measured relative to stressed vowel onset, is QS. Different latency measures were then compared in order to test whether they would be more sensitive to syllable structure and segmental environment. In Figure 6 such measures are compared for H peaks in MD questions. Apart from latency from stressed vowel onset (HtoVons), we find latency from stressed syllable onset (HtoSons), latency relative to stressed vowel offset (HtoVoff) and latency relative to stressed syllable offset (HtoSoff). Note that HtoVons (as well as HtoSons) are pretty constant throughout the conditions, while the other alignment measures do vary in a more dramatic way. Similar results were obtained for statements and for LD.

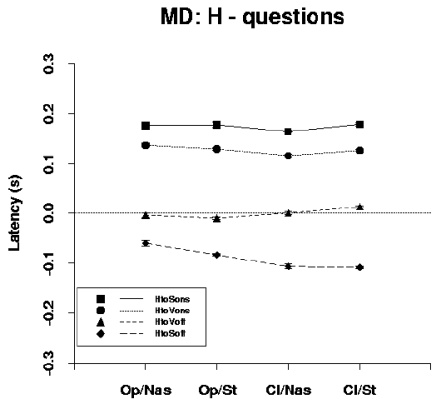

Figure 6

Mean H (Fomax) latency from vowel onset (Hto\-Vons), vowel offset (Hto\-Voff) and syllable offset (HtoSoff) for MD (Op/Nas = open syll., nasal; Op/St = open syll., stop; Cl/Nas = closed syll., nasal; Cl/St = closed syll., stop). The dotted line is the reference point for each latency measurement. Standard error is indicated by vertical bars.

If the peak is timed to occur relative to vowel or syllable onset, we do not expect any effect of syllable structure. On the other hand, we expect those effects to reveal themselves in measures that are relative to the right edge of the stressed syllable, such as HtoVoff and HtoSoff. For instance, the peak should be located closer to vowel offset in closed syllables, whose vowels are short. Alignment should then result to be constant only relative to the left edge of the target syllable.

Such a prediction was supported by the data. When H latency was calculated relative to vowel offset (HtoVoff), apart from the expected effect of QS [F(1,144) = 754.2; $p = 0$], Open/Closed was also highly significant [F(1,144) = 84.03; $p = 0$]. Moreover, there was no Open/Closed by Speaker interaction [F(1,144) = 2.5; $p = 0.12$] (analogously to what was found for HtoVons). In sum, in closed syllables, the peak tended to be within the coda for speaker MD, and towards vowel offset for speaker LD (who generally presented earlier latency values for the entire LHL configuration). Again, Speaker [F(1,144) = 559.7; $p = 0$] affected significantly the results, while Nasal/Stop did not reach significance [F(1,144) = 3.56; $p = 0.06$].

For both speakers, the latency of L2 from vowel onset was greater for questions (0.268 s for LD and 0.262 for MD) than for statements (0.151 s for LD and 0.233 s for MD). The results of the four-way ANOVA revealed a significant effect of QS [F(1,144) = 607; $p = 0$], Nasal/Stop [F(1,144) = 17.04; $p < 0.01$] and Speaker [F(1,144) = 160.7; $p = 0$], but also here no Open/Closed effect was found [F(1,144) = 0.84; $p = 0.36$]. However, Nasal/Stop by Speaker [F(1,144) = 16; $p < 0.01$] was significant, while Open/Closed by Speaker was not [F(1,144) = 4.34; $p = 0.04$]. This interaction was due to the fact that Nasal/Stop was significant only for MD [F(1,72) = 26.07; $p < .0001$] and not for LD [F(1,72) = .01; $p < .9$].

3.2. Fundamental frequency target values

As shown above in Figure 5, L1 F_0 values were lower for questions (108 Hz for LD and 202 Hz for MD) than for statements (119 Hz for LD and 241 Hz for MD). The ANOVA run on L1 F_0 values for both speakers uncovered in fact a highly significant effect of QS [F(1,144) = 154; $p = 0$], and Speaker [F(1,144) = 2893; $p = 0$], while neither Open/Closed [F(1,144) = 0.43; $p = .51$] nor Nasal/Stop [F(1,144) = 5.315; $p = 0.02$] reached significance (at 0.01 level). Given the fact that questions tend to have higher global and local (especially within the peak) F_0 values in various languages of the world, we tested the hypothesis that H values would be higher for questions than for statements. The results showed that the F_0 peak (H) was apparently not affected by QS for speaker MD [F(1,72) = 0.24; $p < 0.63$], but it was for LD [F(1,72) = 105.28; $p = 0$], though in a direction opposed to the expected one. That is, for LD statement peaks were higher than question peaks. In the ANOVA conducted on H F_0 values for both speakers, Open/Closed did not reach significance [F(1,144) = 2.7; $p < 0.1$], nor did Nasal/Stop [F(1,144) = 0.09; $p < 0.77$], while Speaker was obviously significant [F(1,144) = 1966; $p = 0$].

Also, note in Figure 5 that the L2 target presented higher values in questions (114 Hz for LD and 185 Hz for MD) than in statements (109 Hz for LD and 178 Hz for MD). QS was in fact significant [F(1,144) = 31.64; $p < 0.01$] as well as Speaker [F(1,144) = 4185; $p = 0$], while neither

Nasal/Stop [F(1,144) = 1.33; $p = 0.25$] nor Open/Closed [F(1,144) = 0.61; $p = 0.44$] were significant. The interaction of Open/Closed by Speaker was almost significant [F(1,144) = 6.22; $p<0.014$]. Also, three of the three-way interactions reached significance, i.e., QS by Open/Closed by Nasal/Stop [F(1,144) = 10.19; $p<0.01$], QS by Nasal/Stop by Speaker [F(1,144) = 12.77; $p<0.01$] and Open/Closed by Nasal/Stop by Speaker [F(1,144) = 8.08; $p<0.01$]. Three-way ANOVAs were then run on each speaker, which showed that, Nasal/Stop was significant only for LD [F(1,72) = 10.92; $p<0.01$], and not for MD [F(1,72) = 0.09; $p = 0.76$]. Such a discrepancy between speakers might be due to the fact that the presence of a stop at the elbow2 location can render the measurement difficult to perform and inconsistent. On the other hand, the same measurement is usually straightforward within a nasal segment.

3.3. Slope values

Mean slope values and standard error relative to the F_0 rise region from syllable onset to Fomax, obtained through the logistic curve modeling described above, are shown in Figure 7. The effect of the QS manipulation on slope values was not significant [F(1,144) = 2.75; $p = 0.1$]. Analogously, neither the Open/Closed [F(1,144) = 0.02; $p = 0.9$] nor the Nasal/Stop manipulation [F(1,144) = 0.06; $p<0.8$] reached significance. However Speaker was significant [F(1,144) = 33.41; $p = 0$] as well as two of the two-way interactions, i.e., QS by Nasal/Stop [F(1,144) = 15.62; $p<0.01$] and QS by Speaker [F(1,144) = 109.64; $p = 0$]. The last interaction is particularly interesting. In fact, as expected from the higher *Fo* mean value for statement peaks for LD, slope values for such speaker were greater for statements than for questions (see Figure 7), and the difference was significant in a three-way ANOVA [F(1,72) = 130.23; p = 0]. However, the opposite was true for MD. For this speaker, statement rise slopes were shallower than for questions, and the result reached significance in a three-way ANOVA. [F(1,72) = 27.06; $p<0.01$]. Remember that no significant difference for F_0 peaks was reported for MD, though a minor trend for higher question peaks was noticeable.

A logistic curve was also fitted to the F_0 segment in the region between the peak and the following L2. At a first approximation, it appeared in fact that the H to L2 fall was shallower in questions than in statements. However, this was only true for LD [F(1,72) = 63.4; $p = 0$], who had higher peaks for statements than for questions, and not for MD [F(1,72) = 0.49; $p = 0.5$]. Therefore, the difference in slope was entirely predictable from the F_0 results.

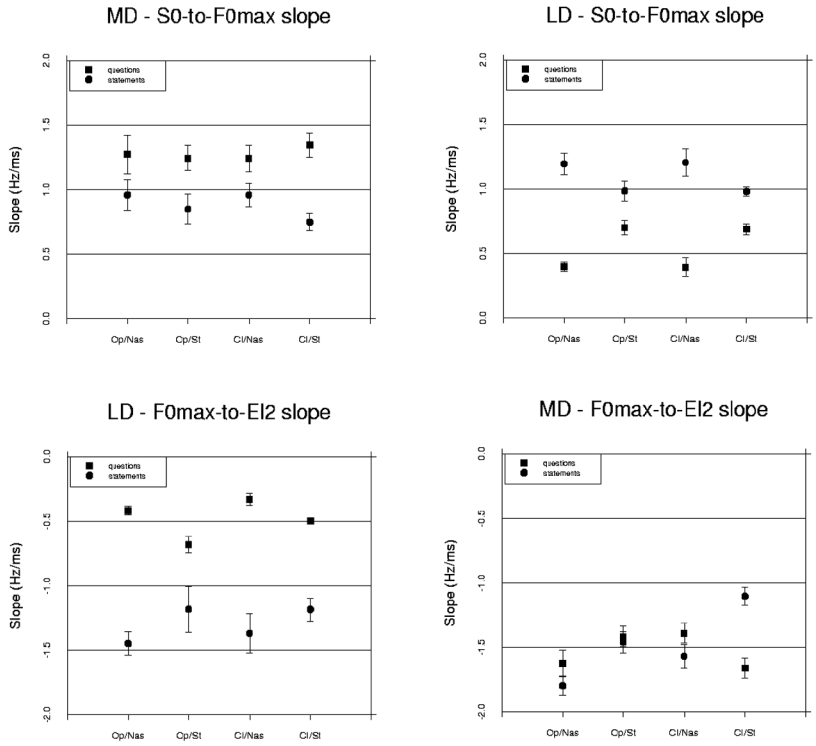

Figure 7
Mean slope values in Hz/ms of the LH (S0-to-Fomax) rise and the HL (Fomax-to-El2) fall in questions and statements for both speakers (Op/Nas = open syllable, nasal; Op/St = open syllable, stop; Cl/Nas = closed syllable, nasal; Cl/St = closed syllable, stop).
Standard error is indicated by vertical bars

4. Discussion

The hypothesis that the pragmatics of the utterance would affect the alignment of the H peak and the preceding L was supported by the data, with questions showing later L1 and H targets. Additionally, we also expected that the timing of L2 would be affected by the coupled displacement of L1 and H in questions and statements, since the HL- fall is here assumed to start at the location for the H target of the LH rise, for both questions and statements. The hypothesis was confirmed by the results, which are novel.

The study also examined the stability of target alignment under variability in both syllable structure and segmental composition of the postaccentual onset. Specifically, it was

hypothesized that if targets are aligned relative to syllable or vowel onset, i.e., the left edge of a phonological or phonetic domain/unit (such as the stressed syllable or the stressed vowel), duration variation induced by either structural or segmental factors would not affect the results. The results appear to support this hypothesis for all three targets.

This means that, despite the expected duration difference between stressed vowels in open and closed syllables (D'Imperio, 2000), peak alignment relative to vowel onset is stable and therefore not proportional to overall vowel duration. The findings parallel those of Dutch prenuclear L (Ladd *et al.*, 2000), in that no relative displacement of L1 alignment was found as a result of differences in phonological length of the stressed vowel when measured relative to syllable (or vowel) onset (though vowel duration had an effect on H peak alignment in that study). The Nasal/Stop factor had only an effect on L1 alignment relative to vowel onset. The origin of such a long-distance effect is still unclear at this point.

When measured relative to the right edge of the stressed vowel/syllable, alignment was more variable but just as far as the syllable structure factor is concerned. Namely, the Open/Closed factor affected the alignment of L1 and L2 only when alignment was measured relative to the right edge of the syllable (both accented vowel offset and following vowel onset). This was not the case when alignment was measured from the left edge of the syllable. Generally, the targets were timed closer to both vowel offset and postaccentual vowel onset when the target syllable was closed. It is well known that closed syllables are characterized by shorter stressed vowels in Italian (D'Imperio and Rosenthall, 1999). If targets are to occur at a certain distance from the left edge of a specific unit, the position relative to the right edge would expectedly vary in order to meet the alignment condition.

In fact, a syllable structure effect was found for H latency only when it was measured relative to vowel offset, i.e., relative to the right edge of a likely domain unit. The findings are reminiscent of those of (Ladd *et al.*, 2000), where a phonological length effect was found on H alignment, which was only measured relative to stressed vowel offset. If peaks were timed to occur relative to vowel offset, however, a very complex situation would arise in Neapolitan Italian. Namely, the alignment of H peaks in statement pitch accents that are associated to a closed syllable can have alignment specifications that could render them almost undistinguishable from question pitch accents associated to an open syllable.

A surprising segmental effect was found, on the other hand, affecting target alignment even when measured relative to vowel and syllable onsets. Such a result was rather speaker-dependent, though. Specifically, the alignment of both L1 and L2, relative to the left edge of the stressed syllable, was later if the postaccentual syllable had a stop consonant in onset position.

However, this was consistently true (for both L1 and L2) only for MD. LD showed such an effect only for L1 alignment (and quite markedly for questions). A speaker-dependent segmental effect, though on the alignment of H, was also found for Greek by Arvaniti *et al.* (1998).

Here, remember that for MD, L2 was generally aligned around the onset of the postaccentual vowel onset in statement utterances. Hence, when syllables are closed and followed by a stop, L2 alignment for MD is disrupted by a localized F_0 rise, which is a consequence of a microprosodic effect in the immediate surroundings of the voiced stop. In such conditions, the automatic procedure tended to place L2 later than in other utterances, at a point where F_0 starts to fall again after the small rise. But this was the case only for MD closed syllables followed by a stop. As indicated above, L2 (*i.e.*, elbow2) generally cooccurred with a much earlier location, i.e., vowel offset, for LD, as it can be seen from Figure 8. This might be why the Nasal/Stop effect was consistently found only for MD.

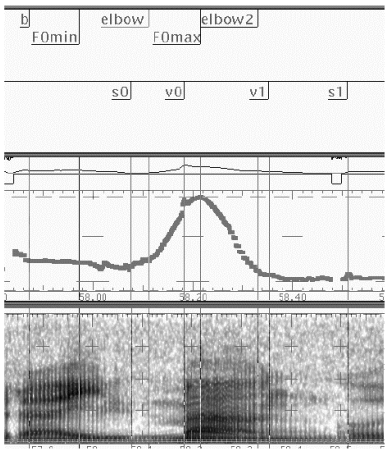

Figure 8
Labels, Fo trace and spectrogram for the statement Vedrai il mango dopo*, produced by LD*

Another interesting result was the syllable structure effect when L2 (elbow2) latency was measured relative to the right edge of the accented vowel, i.e., vowel offset. In this case, the

target was later in closed than in open syllables. How can we account for such an effect? At this point, one can speculate that the HL- falling gesture has a rather fixed duration. Hence, by the time the *F0* fall is completed, the speaker is closer to the syllable offset and might already be into the post-accentual vowel (as for both MD and LD questions) when the target syllable is open, hence shorter. Though the segmental effect on L2 alignment can somehow be accounted for, it is harder to account for such an effect on L1. When latency was measured relative to vowel and syllable onset, both speakers tended to place L1 at later positions when the syllable was followed by a postaccentual stop onset. Such a long-distance segmental effect was entirely unexpected. Unlike the results reported in Arvaniti *et al.* (1998), no segmental effect on H alignment was found in the present study, independent of the reference point for the alignment measure. The most consistently timed target appears therefore to be the H peak. When measured relative to syllable onset, only the question/statement contrast appears to affect its alignment, in that question peaks are aligned relatively later than statement peaks.

The F_0 evidence regarding L2 is quite puzzling, though. If we assume that the HL- fall is the result of the same phrasal gesture in statements as well as in questions (as already proposed for the HL fall of questions in D'Imperio (2001) then we would expect L2 to attain the same F_0 level in both modalities. But this was not the outcome of the F_0 analysis. Both L1 and L2 were significantly different in *F0* terms between questions and statements, for both speakers. Unlike L1 and L2, H peaks did not show a consistent QS effect on F_0. Though the effect was significant when results of both speakers were pooled, it was found that only LD produced a significant difference for the individual results. Moreover, the difference went in the opposite direction to that expected from other languages, for which questions are generally characterized by higher peaks and expanded F_0 range (cf. Jun and Oh, 1996 for Korean). This speaker actually showed the reverse of the expected effect, that is his statements presented higher peaks than questions.

However, when listening to LD utterances, it is clear that he put more emphasis on the statements than on the questions, though he was not instructed to do so. This might be due to the fact that narrow focus statements are harder to produce than narrow focus questions: there is in fact no difference between the pitch accent of narrow focus and broad focus questions in Neapolitan, while the opposite is true for narrow focus and broad focus statements. Hence, narrow focus statements are "marked" and intrinsically emphatic in this variety, since they are signaled by a different pitch accent than the one employed to signal broad focus (while questions do not present such a contrast). It is possible, therefore, that LD tried hard to signal the narrow focus contrast on statements, putting more effort than in the question production.

Heightened effort in the narrow focus statement productions of LD might have translated in greater emphasis, and, consequently, heightened F_0 values.

I believe that the gradient raising of the peak F_0 value is therefore optional, showing strong inter- and intra-speaker variability, in a way that is similar to the use of "emphatic stress". Hence, it is plausible that the difference, which was only significant for LD can be merely attributed to individual differences in signaling narrow focus. In other words, there might be just no systematic difference in peak height between two modalities, as speaker MD shows.

Regarding the F_0 difference for L1 and L2, while this was quite conspicuous for L1 (especially for MD, i.e. 39 Hz), the L2 difference was only equal to 5 Hz for LD and to 7 Hz for MD. Though small, such a difference appeared to be accompanied by a significant difference in slope gradient between the two falls. Nevertheless, individual analyses revealed that the slope difference was consistently maintained only by LD. This is quite suspicious, given two points. First, as noticed above, LD produced always higher H peaks for statements than for questions. Hence, the steeper slope gradient for his statements can be plausibly analyzed as an epiphenomenon of the different emphasis degree attributed to utterances in the two modalities. Second, it appeared as if LD employed a strategy to displace L2 as far as possible towards the onset of the postaccentual vowel. This could only render the HL fall of questions shallower than the fall of statements. A different control of the fall in questions and statements is therefore still uncertain at this point.

5. Conclusion

To summarize, the hypothesis that the alignment of L1, H and L2 (expressed in terms of the elbow, F0max and elbow2 measures) would be affected by the question vs. statement contrast was confirmed. This replicates the findings of D'Imperio (1996, 1997a). The hypothesis that syllable structure would affect target alignment relative to the left edge of the syllable was not confirmed, while a difference was found when alignment was measured relative to the right edge of the syllable. Additionally, the hypothesis that segmental environment would affect the alignment of the H peak when measured relative to syllable onset was not confirmed. The hypothesis was instead supported for L1 and partially for L2 alignment. Assuming that the strong hypothesis of invariant alignment of such targets relative to syllable or vowel onset holds, it will be interesting to see how such factors interact in the perception of the question/statement contrast. Finally, the hypothesis that slope would differ in questions and statements was not supported for the LH rise, but it was partially supported for the HL fall.

6. References

ARVANITI, A., LADD, D.R. and I. MENNEN (1998), "Stability of tonal alignment: The case of Greek prenuclear accents", *Journal of Phonetics*, 26:3-25.

D'IMPERIO, M. (1996), "Caratteristiche di timing degli accenti nucleari in parlato italiano letto", In *Atti del XXIV Convegno Nazionale dell'Associazione Italiana di Acustica*, Trento, Italy, June 12-14, 1996, p. 55-60.

D'IMPERIO, M. (1997), "Breadth of focus, modality and prominence perception in Neapolitan Italian", In K. Ainsworth-Darnell and M. D'Imperio (Eds.) *The Ohio State University Working Papers in Linguistics - Papers from the Linguistics Laboratory*, vol. 50, p. 19-39, OSU.

D'IMPERIO, M. (2000), *The Role of Perception in Defining Tonal Targets and their Alignment*, Ph.D. Thesis, The Ohio State University.

D'IMPERIO, M. (2001) "Focus and tonal structure in Neapolitan Italian", *Speech Communication*, 33(4):339-356.

D'IMPERIO, M. and D. HOUSE (1997), "Perception of questions and statements in Neapolitan Italian". In G. Kokkinakis, N. Fakotakis, and E. Dermatas (Eds.), *Proceedings of Eurospeech '97*, Volume 1, Rhodes, Greece, p. 251-254.

D'IMPERIO, M. and S. ROSENTHALL (1999), "Phonetics and phonology of main stress in Italian", *Phonology*, 16(1): 1-28.

GRICE, M., D'IMPERIO, M., SAVINO, M. and C. Avesani (to appear), "Towards a strategy for labelling varieties of Italian", In S.-A. Jun (Ed.), *Prosodic Models and Transcription: Towards Prosodic Typology*, Oxford: Oxford University Press.

GÓSY, M. and J. TERKEN (1994), "Question marking in Hungarian: timing and height of pitch peaks", *Journal of Phonetics*, 22:269-281.

JUN, S.-A. and M. OH (1996), "A prosodic analysis of three types of wh-phrases in Korean", *Language and Speech*, 39:37-61.

LADD, D.R. (1996), *Intonational Phonology*, Cambridge: Cambridge University Press.

LADD, D.R., MENNEN, I. and A. SCHEPMAN (2000), "Phonological conditioning of peak alignment in rising pitch accents in Dutch", *JASA*, 107(5):2685-2695.

SILVERMAN, K. and J.B. PIERREHUMBERT (1990), "The timing of prenuclear high accents in English". In Kingston, J. and M.E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, Cambridge: Cambridge University Press, p. 71-106.