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

Mariapaola D’Imperio

LORIA – INRIA, France
dimperio@loria.fr

Abstract

Unlike in languages such as English and Standard Italian, Neapolitan Italian yes/no questions and narrow focus statements share a rising-falling (LHL) tune [1, 2]. However, the alignment of the target H peak has been claimed to be later in questions. 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.

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 number of factors in a variety of languages [3, 4, 5]. 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 contraints proper to the language under scrutiny. For instance, the findings in [5] suggest that vowel duration, as a consequence of phonological length, affects peak alignment in Dutch rises. 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 is also 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, 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 [6, 2]. 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 [7]. 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- 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)

<table>
<thead>
<tr>
<th>Open syllable/nasal</th>
<th>Open syllable/stop</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;ninth&quot;</td>
<td>&quot;magician&quot;</td>
</tr>
<tr>
<td>&quot;mammo&quot;</td>
<td>&quot;mango&quot;</td>
</tr>
</tbody>
</table>

Table 1: Target words for the corpus. ("mammo" is a man that takes on the tasks of a mother.)

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.

Scaling of \( F_0 \) values has been reported to be affected by the question/statement distinction in a variety of languages, with questions presenting usually higher values (cf. [8]). Therefore, we also tested the hypothesis that Neapolitan questions are characterized by higher fundamental frequency 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 [9], 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 will be 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. Two Neapolitan speakers, one male (LD) and one female (MD), produced each target sentence 10 times, in randomized order, for a total of 160 utterances. The recordings were then digitized at 16 bit, 16 kHz using ESPS Waves* on a SUN Sparc 10 station and F0 was later extracted at every ms using the getf0 ESPS pitch tracker.

2.2. Acoustic measurements

Both duration and F0 measurements were performed. Specifically, within the rise-fall configuration, the F0 value in Hz was measured at the first L target (L1), at the H target and at the second L target (L2). The H target was measured at the single highest F0 point within the accented syllable and was labeled “F0max” (see Figure 1). 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 (see [6] for details). The measures employed for L1 and L2 were labelled “elbow” and “elbow2” respectively. Labels are shown in Figure 1.

We marked locations for the accented syllable onset (s0), the stressed vowel onset (v0), the postaccentual vowel onset (s1) and stressed vowel offset (v1). 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 F0max (H) from v0 and s0, as well as from v1 and (iv) the distance of elbow2 (L2) from v1, s1 and F0max.

We also observed that the F0 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 F0 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, \(t_0\) for the temporal coordinate of the curve inflection point, and \(a, b, c\) 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 \(t_0\). The first derivative of the model is:

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

The peak velocity was therefore estimated by \(\frac{dF_0}{dt}\), again expressed in Hz/ms. This model resulted in a very good fit to the

[...]

3. Results

3.1. Temporal Alignment

In Figure 3, mean latency from vowel onset and F0 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.

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. Question peaks were aligned 37 ms later than 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.

The statistic results for L1 alignment showed 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 (H0vons). 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 \(F(1, 144) = 400; p = 0\), while Open/Closed was not \(F(1, 144) = 5.9; p = 0.02\). Nasal/Stop was not significant \(F(1, 144) = 0.8; p = 0.4\), while Speaker was \(F(1, 144) = 12.2; p = 0\).

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 4 such measures are compared for H peaks in MD questions. Apart from latency from stressed vowel onset (H0vons), we find latency from stressed
syllable onset (HtoSons), latency relative to stressed vowel off-
set (HtoVoffs) and latency relative to stressed syllable offset
(HtoSoff). Note that HtoVons (as well as HtoSons) are pretty
contant throughout the conditions, while the other alignment
measures do vary in a more dramatic way. Similar results were
obtained for questions and for LD.

For both speakers, the latency of L2 from vowel onset was
greater for questions (0.268 s for LD and 0.262 s for MD) than
for statements (0.151 s for LD and 0.233 s for MD). The re-
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\}.

Figure 3: \(F_0\) values and latency from \(v_0\) (vowel onset) for L1,
H and L2 for speaker LD (upper) and MD (lower).

![Figure 3: \(F_0\) values and latency from \(v_0\) (vowel onset) for L1, H and L2 for speaker LD (upper) and MD (lower).](image)

Figure 4: Mean \(H\) (\(F_0\)max) latency from vowel onset (Hto-
Vons), vowel offset (HtoVoffs) and syllable offset (HtoSoffs) 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.

Figure 5: Mean slope values in Hz/ms of the LH (S0-to-F0max)
rise and the HL (F0max-to-E12) 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.

3.2. Fundamental frequency target values

As it shown in Figure 3, 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 lan-
guages 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; \(p = 0\)\}.

Note in Figure 3 that the L2 target presented higher val-
ues 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; \(p < 0.01\)\}, while nei-
ther Nasal/Stop \{F(1, 144) = 1.33; \(p = 0.25\)\} nor Open/Closed
\{F(1, 144) = 0.61; \(p = 0.44\)\} were significant. Speaker was
also significant. \{F(1, 144) = 4185; \(p = 0\)\}.

3.3. Slope values

Mean slope values and standard error relative to the \(F_0\) rise re-
gion from syllable onset to \(F_0\)max, obtained through the logis-
tic curve modeling described above, are shown in Figure 5. The
effect of the QS manipulation on slope values was not signif-
ificant \{F(1, 144) = 2.75; \(p = 0.11\)\}. 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
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, peak alignment relative to vowel onset is stable and therefore not proportional to overall vowel duration. The findings parallel those of Dutch prenuclear L [5], in which 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 uncertain at this point.

When measured relative to the right edge of the stressed vowel/syllable, alignment was more variable. 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 be expected to vary in order to meet the alignment condition.

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 F0. 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 F0 range. Note, though, that narrow focus statement productions of LD might have translated in greater emphasis, and, consequently, heightened F0 values. Regarding the F0 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.

4. Discussion

The hypothesis that the pragmatics of the utterance would affect the alignment of the H peak and the preceding Low 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, peak alignment relative to vowel onset is stable and therefore not proportional to overall vowel duration. The findings parallel those of Dutch prenuclear L [5], in which 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. 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 be expected to vary in order to meet the alignment condition.

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 F0. 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 F0 range. Note, though, that narrow focus statement productions of LD might have translated in greater emphasis, and, consequently, heightened F0 values. Regarding the F0 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 would be affected by the question vs. statement contrast was confirmed. 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


