At the interface between phonetics and pragmatics: Non-local f_0 effects on the perception of Cosenza Italian tunes

Caterina Petrone

Zentrum für Allgmeine Sprachwissenshaft

petrone@zas.gwz-berlin.de

Abstract

Analysing the tune meaning is often equivalent to analysing the meaning contribution of the "nucleus" (the intonation region starting from the last pitch accented syllable until the end of the intonation phrase). In this paper we report results from an identification task in Cosenza Italian, suggesting that listeners are able to identify the contrast between questions and statements by exploiting phonetic differences in the implementation of the rise-fall prenuclear contour. Such results mirror previous studies on Neapolitan and German and raise questions on the contribution of prenuclear f_0 details to tune meaning.

Index Terms: Intonation, perception, phonetic variability, meaning compositionality, Cosenza Italian.

1. Introduction

Many theories of intonation claim that the intonation contour can be decomposed in smaller intonational "events" (tonal movements or level tones), and that such events (or their combination) contribute to the meaning of the tune in which they are integrated. Traditionally, though, a primary importance has been assigned to the "nucleus" (the intonation region starting from the utterance's last pitch accented syllable until the end of the intonation phrase) over the "prenuclear" region (the stretch of contour preceding the nuclear accent) For example, in the British tradition, the nucleus is the only mandatory element in an intonation contour, and the number of intonational choices is higher for nuclear tones than for the "head" and the "prehead". As a consequence, the analysis of tune meaning is focussed mainly on analysing the contribution to meaning of the nucleus [1].

The special status of the nucleus has been rejected by [3], according to which nuclear and prenuclear accents are selected from the same inventory, the nuclear accent being merely the last pitch accent within the phrase. Tones in a tune freely combine, since tune meaning is the sum of the independent contribution of the meaning of its individual tones. Despite such assumptions, works on tune meaning couched within the AM model continued to consider the nucleus, less or more implicitly, as the semantic "heart" of tunes. Therefore, such works are aimed at investigating the meanings associated to the nuclear accents and to the following edge tones. For example, in the description of American English intonation by [4], the combination of the nuclear H* accent with the L- phrase accent and the L% is regarded as typical of broad focus statements (S), while the L* H- H % tune is typical of yes/no questions (Q).

Recently, work on Neapolitan Italian and German suggested that listeners are able to perceive the Q/S contrast well before hearing the nucleus. In Neapolitan, yes/no questions and statements are distinguished only intonationally, a strong cue being the choice of the nuclear accent [5]. Furthermore, in questions, the f_0 fall after a prenuclear rise is shallower, so that the f_0 contour assumes a convex shape. In statements, the f_0 rapidly falls from the prenuclear peak to the end of the accented prosodic word [6]. Such a steep fall is a strong cue for statement interpretation, even in absence of the nuclear accent information ([7]). In German, questions and statements can be distinguished by lexical, syntactic and intonational means. Moreover, [8] reported that German listeners use differences in the alignment (early vs. late) of the prenuclear peak and in the shape (concave vs. convex) of the immediately following f_0 fall to identify "intonational" questions (questions marked only intonationally) and statements.

This paper continues this line of research by focussing on Cosenza Italian. In this variety, as in Neapolitan, the choice of the nuclear accent is important in cuing the Q/S distinction. Specifically, broad focus statements are characterized by a nuclear H+L* accent [1], while yes/no questions are characterized by a nuclear L+H* accent [9]. Moreover, the phonetic implementation of the prenuclear rise-fall contour differs across the two modalities in many respects, though in a partial different way than in Neapolitan and German. In fact, as showed in Fig. 1, a steep f_0 fall follows the prenuclear accents in both modalities. However, the prenuclear rise (analysed by [9] as H*) is aligned earlier in statements. Specifically, in statements, the f_0 rise starts at the beginning of the paroxyton word *Marinella* and ends at the accented vowel offset. In questions, the f_0 rise starts in the syllable preceding the accented one, and ends right after the accented vowel offset (i.e., within the following coda consonant). Also, while the melodic value of the f_0 rise onset is similar across the two modalities (Q: 218 Hz; S: 222 Hz), the prenuclear peak is higher in questions (Q: 400 Hz; S: 335 Hz). The end of the fall is located in the middle of the syllable /vuo/, where a low turning point is visible in both modalities. However, such a point is lower in questions (Q: 188 Hz; S: 208 Hz). As a consequence, in questions, both the rise and the fall slopes are steeper. Finally, the rise shape appears to be convex in questions and concave in statements.

Our hypothesis is that phonetic details in the implementation of the prenuclear rise-fall contour are exploited to perceptually enhance the Q/S distinction. If this is true, one or more of such f_0 dimensions (tonal alignment, f_0 scaling, slope and shape) will be employed by Cosenza listeners for the purposes of the intonational contrast. If, on the contrary, such a phonetic variability does not matter, listeners will be not able to distinguish questions and statements in absence of the nuclear information.

[1] In Cosenza Italian, as in other Italian varieties, the peak of this accent is acoustically downstepped, even though perceptually salient (Fig. 1).

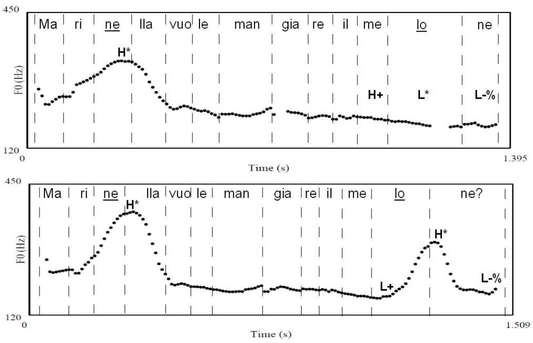

Figure 1: f_0 *contour and phonological analysis of a broad focus statement and a yes/no question for the sentence "Marinella vuole mangiare il melone". Accented syllables are underlined. Because of the difficulty in cutting geminates, the coda consonant in the accented syllable "-nel-" is segmented with the following syllable.*

2. Methods

2.1. Stimuli preparation

A natural yes/no question was selected from a small corpus of sentences read by a female speaker from Cosenza. The sentence was composed of: (1) a utterance-initial proper name (*Marinella*) in which the syllable *-nel-* was associated to a prenuclear pitch accent; (2) an unaccented verbal phrase (*vuole mangiare*, "wants to eat") and (3) a paroxyton noun (*il melone*, "the melon") carrying a nuclear L+H* accent. The utterance was first cut at the end of *vuole*, to prevent listeners from exploiting nuclear information in identifying Q vs. S. Subsequently, a linear stylization of the pitch contour was carried out, in which five points were interpolated: one point at the utterance beginning; two points at the beginning (L1) and at the end (H) of the prenuclear f_0 rise; one point at the end of the f_0 fall (L2); one point at the end of the fragment. As for tonal scaling, a baseline condition was created, in which the f_0 was kept at intermediate level for the L1-H-L2 configuration (i.e., between the average values found in Q and S for this speaker). These f_0 values will be referred to as "medium" ("m") as opposed to "high" ("h") and "low" ("l") values. For example, the "mmm" combination will indicate that L1, H and L2 have all medium values, while the "hhh" combination will indicate that L1,H and L2 have all high values. Our manipulations as well as the hypotheses underlying them are listed below (Hyp. **(a)**-**(g)**). The hypotheses are mainly based on acoustic observations.

- **(a)** Rise-fall Timing. The L1-H-L2 configuration in "mmm" stimuli was shifted right- and leftwards relative to the accented vowel offset in six steps of 15 ms (T1-T6, Fig. 2). This continuum will be referred to as "primary" ("p") as opposed to other continua in which tonal targets were independently manipulated (see below). Hyp: the Q score will increase as the L1-H-L2 contour is shifted rightwards, since, in questions, the prenuclear accent is aligned later.

To avoid a proliferation of the stimuli, the other f_0 parameters were modified only for two Timing Steps, T3 and T5. At these timing values, the prenuclear peak alignment was similar to that found in representative examples of statements and questions, respectively (Fig.1). Moreover, the choice of timing steps in the middle (rather than at the extrema) of the alignment continuum ensured that additional manipulations did not compromise stimuli naturalness:

- **(b)** L1 Timing (4 levels). L1 was shifted 30 ms (L1-veryEarly) or 15 ms before (L1-early) and 15 ms after (L1-late) the corresponding values found in the primary continuum (L1-p). Hyp: Earlier L1 locations will cue more S, while later L1 locations will cue more Q.
- **(c)** H timing (3 levels). The H peak was shifted 15 ms before (H-early) or after (H-low) the corresponding values found in the primary continuum (H1-p). Hyp: Earlier H alignment will cue to more S, while late alignment within the coda consonant will cue more Q.
- **(d)** Rise-fall scaling (3 levels). The height of the L1-H-L2 contour was lowered (lll) or raised (hhh) by 20 Hz relative to the mmm stimuli. Hyp: Higher L1- H-L2 peaks will increase the Q score, since higher pitch is cross-linguistically more typical of questions.
- **(e)** L2 Height (3 levels). L2 was shifted 20 Hz lower (mml) or higher (mmh) than the corresponding values in the mmm stimuli. Hyp: Lower L2 values will cue more Q.
- **(f)** Rise-fall Steepness (3 levels). Shallow and steep rise-fall contours were created by shifting the H target 20 Hz lower (mlm) or higher (mhm) than the corresponding values in the mmm stimuli. Hyp: Shallow rise-fall slope (mlm) will cue more S, while steep rise-fall slope (mhm) will cue more Q.
- **(g)** Rise Shape (3 levels). The linear L1-H interpolation in the mmm stimuli was manipulated by shifting the f_0 value in the middle of the transition 20 Hz up and down, to create concave (cv) and convex (cx) stimuli. Hyp: Convex f_0 rises will cue more Q.

A total of 19 stimuli was obtained from the combination of all these experimental variables. Two control stimuli were added (created from the Q and the S bases in Fig. 1), in which the mean values typical for the prenuclear rise-fall configuration were combined with the typical nuclear configuration of statements (H+L*) and questions (L+H*). All the stimuli, created through PRAAT [10], were resynthesized by means of PSOLA. More details for the stimuli manipulations can be found in [5], [7], [8], by which our methods are inspired.

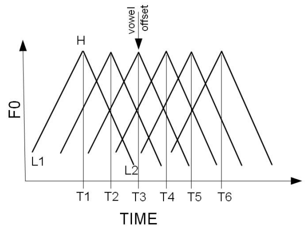

Figure 2: *Manipulation of the* f_0 *rise-fall timing.*

2.2. Procedure

The stimulus set, preceded by 10 practice fragments, was played through a laptop 5 times in the same randomized block by means of PERCEVAL [11]. At the end of the session, participants listened to the control sentences containing the nuclear accent. This presentation order was intended to avoid “learning” effects, which could have biased listeners’ responses for tune fragments.

Fourteen Cosenza listeners, aged 19-39, participated in the experiment. They were told that they were going to listen to some fragments of sentences, containing same words but different melodies, as if the speaker was suddenly interrupted while formulating them. After listening to each fragment, listeners had to decide whether the speaker was going to formulate a Q or as a S by pressing the left (for Q) or the right (for S) arrow as soon as they were able to identify it. These labels were also visualized on the computer screen, always in the same order. Reaction Times (RT) were simultaneously recorded [2].

3. Results

For the sake of brevity, only major results are reported. As for statistics, generalized linear models with mixed effects were applied on the Q/S identification score. These models all included the listeners as the random term, while they differed as for the number and type of fixed factors. The effect of the rise-fall timing manipulation was tested by a model with Timing as the only fixed factor (point **(a)**, par. 2.1). For factors listed in **(b-g)**, the main effects of each of these factors, as well as the effect of Time Steps (T3/T5) and their interactions were calculated. Another series of models were run without the interaction terms, when they were not significant.

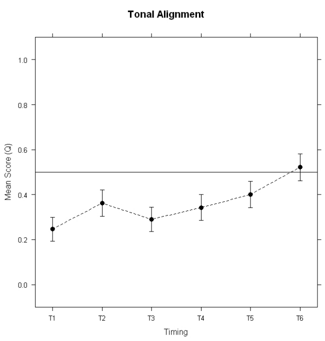

Figure 3: *Mean Q score (y axis) across Timing (x-axis) pooled for all subjects. The straight line indicates the chance level (0.5).*

Fig. 3 shows that the intonation identification is function of Timing. Specifically, at T1, the Q identification score is at 0.25 (i.e., 25%). The identification score increased towards the chance level as the rise-fall was shifted rightwards, so that, at T6, the Q score is at 0.52 (i.e., 52%). The statistical analysis showed a strong effect of the Rise-fall Timing (z = 3.4, p = 0.0005). One might wonder whether this effect is due to the shift of the whole rise-fall contour or to the shift of a specific target within it (L1, H, L2). Separate analyses were run to test the effect of L1 and H alignment (i.e., the only targets whose temporal alignment was independently manipulated; points **(b-c)**, par. 2.1). The Q score significantly increased from L1-p to L1-late stimuli (z = 2.1, p = .03), even though it was never above the chance level. Moreover, neither the contrast between L1-p and L1-early stimuli nor that between L1-p and L1-veryEarly stimuli was significant. The H alignment manipulation did not yield to significant results.

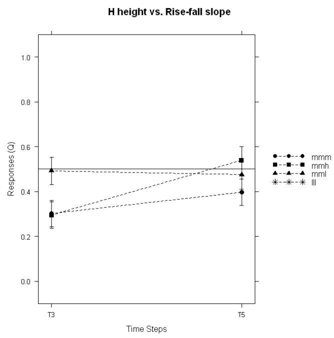

Figure 4: *Mean Q score (y axis) across Timing Steps (x-axis) for “mmm” (dots), “hhh” (squares) and “mhm” (triangles) stimuli. The straight line indicates the chance level (0.5).*

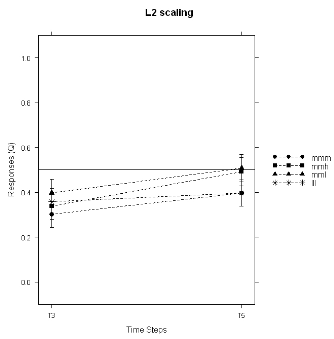

Figure 5: *Mean Q score (y axis) across Timing Steps (x-axis) for the L2 scaling manipulation. The straight line indicates the chance level (0.5).*

Tonal timing is not the only factor at work. Fig. 4 plots results for the mmm, mhm and hhh stimuli. It is important to take into account that hhh and mhm stimuli were created to test the effect of scaling and slope on the Q/S identification (points **(d)** and **(f)**, par. 2.1). In mhm stimuli, the steeper rise-fall slope was obtained by raising the peak height and by keeping L1 and L2 at intermediate melodic values; in hhh stimuli, the three targets had all high values. The graph shows that mhm stimuli scored around the chance level at T3 and T5, with the contrast between mmm and hhh being significant at T3 (z = 2.3, p = .01), but not at T5 ($p > .05$). Moreover, mhm stimuli scored similarly as mmm stimuli at T3, while they scored 0.53 at T5. However, the Q increase at T5 was only marginally significant (p = .07).

For scaling manipulations described at point **(d)**, we assumed the effect of intonation modality to go in the same direction for L1, H and L2. However, in natural stimuli, it appears that L1 has similar values in questions and statements, H is higher in questions while L2 is lower in questions (thus resulting in highly asymmetric contours). Therefore, the effect of the independent manipulation of L2 scaling (point **(e)**) was tested. Graph 5 shows that lowering the entire rise-fall contour (lll), raising (mmh) or lowering (mml) L2 led to an increase of

[2]RT differences across the gated stimuli were never significant, probably because of the difficulty of the experiment. RTs are here employed for data selection, i.e., responses whose RT was shorter than 200 ms after the stimulus onset and longer than 1200 after the offset were discarded from the statistical analysis of the data (corresponding to 7.4% of the dataset).

Q responses relative to the mmm condition at T3. Indeed, the contrast between mml and mmm (z = 2.4, p = .01) stimuli was significant, while that one between mmh and mmm (z = 1.84, p = .06) and between lll and mmm (z = 1.82, p = .06) were marginally significant.

4. Discussion

This experiment suggests that Cosenza listeners are able to distinguish yes/no Q and broad focus S even when the nuclear accent information is not available. However, the impact of our manipulations varies with intonation modality. The early alignment of the rise-fall prenuclear contour is a very strong cue for interpreting the fragments as statements. In fact, this pattern is typical of prenuclear accents in S modality. While the effect of the independent manipulation of H was not significant, L1 alignment seems to play a role in Q/S identification. Since L2 alignment was not independently manipulated, we cannot establish yet whether the rise-fall alignment effect is due to a shift in L1 or L2 alignment, to an interplay between them, or to a more global shift in the whole rise-fall configuration.

The Q score gradually increased from T1 to T6. This is expected, since our manipulations were intended to reproduce phonetic differences *within* the same phonological category (H*). However, when the rise-fall contour was shifted rightwards, the Q score was around the chance level, i.e., listeners could not "safely" categorize such stimuli as questions. Additional manipulations in slope steepness/rise-fall scaling increased the Q score but they never yielded to results significantly above 50%. The effects of L2 scaling are unclear, since either raising or lowering it led to results around the chance level. Results for questions are surprising because the base stimulus for our manipulations was a question. In Italian, many prosodic and segmental cues other than f_0 can be used to enhance the Q/S contrast [5]. Therefore, listeners' responses should have been biased towards the Q interpretation.

Different explanations can be invoked to account for such a discrepancy between Q and S identification. The choice of the timing region (T1-T6) for the construction of the continuum was based on observations in the production of a small corpus. However, it is possible that in Cosenza Italian the rise-fall contour should be pulled much later in order to perceive a question. Another explanation is based on the properties of the prenuclear fall. We know that the contrast between low and high pitch is often associated to the contrast between statements and questions, and that its phonetic instantiation can be different across languages [12]. For example, high pitch in questions might be signaled by the nuclear peak delay in Neapolitan, and by the utterance-final rise in German. Moreover, in Neapolitan and German, the prenuclear fall has a more convex shape and a shallower slope (and thus, a higher pitch) in questions than in statements. In Cosenza Italian, on the contrary, both modalities are characterized by a steep fall (even steeper in questions), which was also present in our gated stimuli. As a consequence, it is possible that, while its presence is a strong cue to S interpretation, more ambiguous results are obtained when perceiving a question. An objection which can be raised is that, in Cosenza Italian, the high pitch in questions is not signaled by a slope difference in the fall, but by a wider pitch span, since questions are characterized by higher H and lower L2. In future research, we thus aim at testing the specific impact of the prenuclear fall on intonation identification in this Italian variety. Moreover, cross-linguistically, low pitch is a stronger cue for statement interpretation than high pitch for question interpretation [12]. This is also confirmed in our previous studies focussing on the role of the prenuclear f_0 contour in Neapolitan and German. In Neapolitan, both questions and statements are well identified before hearing the nucleus, though the presence of a steep fall is a stronger cue for statements than the shallow slope for questions. In German, the early alignment of the prenuclear rise-fall contour is, by itself, a strong cue for statements. However, more f_0 cues should be present to perceive questions: the late prenuclear peak alignment and the concave shape of the prenuclear fall. Similarly to German listeners, Cosenza listeners exploited the early rise-fall alignment for statements identification. However, the combination of more f_0 cues only resulted in a slight increase of the Q score. A comparative study is thus necessary to test the meaning contribution of the low vs. high pitch contrast in the prenuclear contour as well as the perceptual saliency of the high pitch in languages/varieties in which the Q/S contrast is differently realized, such as Neapolitan, Cosenza Italian and German.

To sum up, in Cosenza Italian, the systematic variability of the prenuclear f_0 rise-fall contour is exploited for intonation identification. This also means that f_0 cues to meaning might spread over different portions of the tune, thus helping to perceptually enhance intonational contrasts. Our findings do not support theories which do not take into account the contribution of the prenuclear f_0 region to tune meaning and which disregard the role of phonetic detail in phonological representation.

5. Acknowledgements

Thanks to Dr. Cinzia Citraro and to Francesco Imperio for data collection at University of Calabria (Italy), and to Oreste Salerno for reading an early draft of the paper. This work was funded by a grant given by the German Research Foundation to the SPP1234 program.

6. References

[1] Halliday, M.A.K. (1967), Intonation and Grammar in British English. The Hague: Mouton.

[3] Pierrehumbert, J. B. (1980). The Phonetics and Phonology of English Intonation. PhD Thesis, MIT.

[4] Pierrehumbert, J. & J. Hirschberg (1990), The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (eds.), Intentions in Communication, pp. 271-311. Cambridge, Massachusetts: MIT Press.

[5] D'Imperio, M. (2000). The Role of Perception in Defining Tonal Targets and their Alignment. Ph.D. Thesis, OSU.

[6] Petrone, C. & M. D'Imperio (2008). Tonal structure and constituency in Neapolitan Italian: Evidence for the accentual phrase in statements and questions. Proc. of Speech Prosody, 301-304.

[7] Petrone, C. and D'Imperio, M. (to appear). From tones to tunes: The contribution of the prenuclear region in the perception of Italian questions and statements.

[8] Petrone, C. & O. Niebuhr (submitted). The intonation of intonational questions.

[9] Sorianello P. (1995) Contorni intonativi e confini sintattici. Un'indagine acustica sull'italiano parlato a Cosenza. PhD Thesis, University of Florence.

[10] Boersma, P. & D. Weenink (2009). Praat: doing phonetics by computer (Version 5.1.20). From: http://www.praat.org/

[11] André, C., A. Ghio, C. Cavé & B. Teston (2007). PERCEVAL (Version 3.0.4). From: http://www.lpl.univ-aix.fr/

[12] Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. In Proc. of the 1st Conference on Speech Prosody, pp. 44-57.