Phonetic Detail in Intonation Contour Dynamics

Francesco Cangemi

PHONETIC DETAIL IN INTONATION CONTOUR DYNAMICS

Francesco Cangemi
Laboratoire Parole et Langage – Université de Provence
francesco.cangemi@lpl-aix.fr

ABSTRACT

The Autosegmental-Metrical theory of intonation investigates the relationship between f0 contours and post-lexical meaning. Phonetic data are represented in the phonology as a sequence of discrete, local events. The properties of the *transitions* between one event and the next are considered to be phonologically irrelevant (§1).

We present data on Neapolitan Italian which show a significant correlation between the shape of these transitions and the pragmatic context in which a sentence is uttered. This correlation is stronger than the one displayed by traditional autosegmental-metrical indices (§2 and §3).

In the conclusions, we discuss the usefulness of our findings as a step towards the fine-tuning of the autosegmental-metrical theory (§4).

1. INTRODUCTION

Intonational phonology aims at describing how phonetic suprasegmental features convey post-lexical meaning in a linguistically structured way (see Ladd, 2008 for an introduction). For instance, vocal fold vibration rate data (*f0*) are used to describe acoustic differences of the same utterance in different pragmatic contexts, as in the opposition between assertive and questioning modality.

In the frame of the autosegmental-metrical theory of intonation (*AM*), phonetic (continuous) f0 data are translated into a phonological (discrete) inventory of tunes, composed by combining only two tones, high (*H*) and low (*L*). Intonation contours consist of a string of tonal events, linked to the prosodic structure of the sentence. Some tonal events, mainly the *pitch accents* (i.e. those associated to prominent syllables), can phonetically appear as a rise (or a fall) in the f0 curve. In these cases, they are analyzed as the succession of two tones (L H for rises, H L for falls)[1].

In AM, the f0 path between the two tones which compose a rising pitch accent is not regarded as phonologically relevant. Speech synthesis systems based on this framework (e.g. Pierrehumbert, 1981, Anderson *et alii*, 1984, Black & Hunt, 1996) use a simple monotonic interpolation between the two tones. Nonetheless, data from Neapolitan Italian (*NI*, D'Imperio *et alii* 2008) show that, in different pragmatic contexts, the intonation contour of the same segmental string also differs systematically in terms of the f0 path between the two tones. The curve seems to follow a concave or convex[2] path, depending on the pragmatic context in which the sentences are uttered (see Figure 1).

[1] See D'Imperio (1999) for Neapolitan Italian. Similar treatments have been proposed also for spanish (Hualde, 2000; Face, 2001) and english (Ladd & Schepman, 2003).
[2] Note that the attributes of concave and convex refer to the half-plane above the curve.

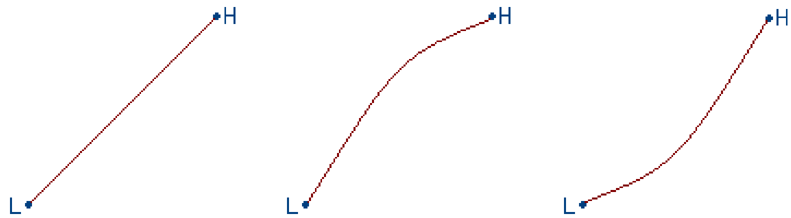

Figure 1. Sketch of linear, concave and convex interpolation.

Figures 2 and 3 display the spectrogram and the f0 contour for the sentence *Milena lo vuole amaro* 'Milena drinks her coffee unsweetened'. In the first case, the sentence is uttered as a statement, while in the second it is uttered as a question. The most striking difference between the two f0 contours is visible in the movement associated to the last stressed syllable of the sentence ("aMAro", highlighted by the box in both figures). In Fig. 2 we find a gradual fall, while in Fig. 3 we find a slight rise followed by a quite rapid fall. In other words, the f0 peak (H in the figures) occurs slightly before the vowel onset in the sentence, but is found later (vowel-internal) in the question, where is also visibly higher. Following the usual terminology, the H belonging to the last pitch accent is *aligned* (in time) and *scaled* (in frequency) differently in the two contexts.

Tone alignment and scaling are the indices usually employed in AM to define the phonetic properties of different phonological entities (e.g., of different pitch accents). But if we concentrate on the intonation contour of the first word in the sentence (*Milena*, isolated from the rest of the utterance by the vertical line in Figures 2 and 3), we notice that the rising movement associated with the stressed syllable has a different *shape* in the two contexts. This difference, though, does not seem to be related either to the alignment or to the scaling of the two tones: both Ls are in the first half of the stressed syllable onset, and around 225Hz; both Hs are at the end of the stressed syllable nucleus, and around 350Hz.

This work aims to investigate some acoustic differences considered in the AM framework as phonetic detail without phonological relevance, such as dynamic (i.e. in shape) differences. Their efficacy as indices to differentiate pragmatic contexts will be compared to that of traditional cues such as the alignment and scaling of the tones which compose the first nuclear accent of our sentences[3]. In the conclusions we will discuss the implications of our results with respect to a fine-tuning of the existing phonological model.

[3] For a discussion about the nuclearity of the first pitch accent in the (partial topic) statement utterances, see D'Imperio & Cangemi (2009).

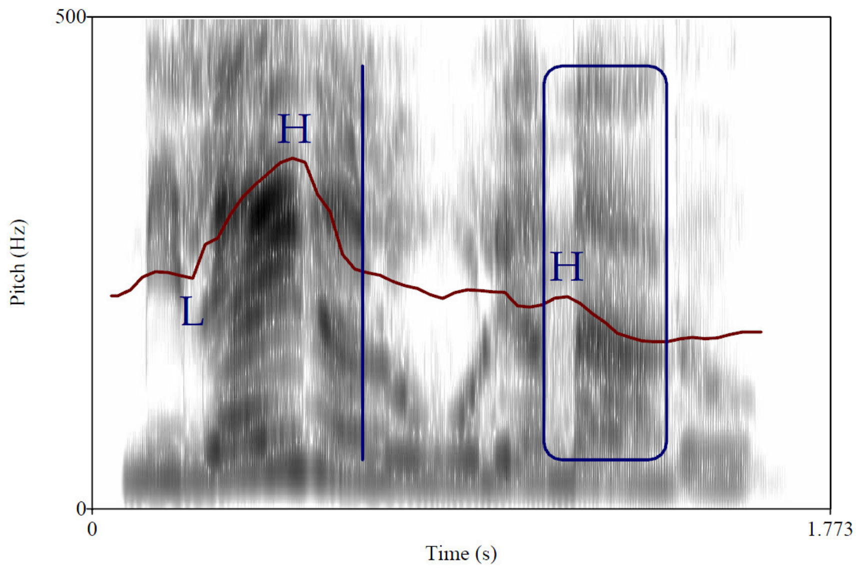

Figure 2. Utterance in (partial topic) statement context.

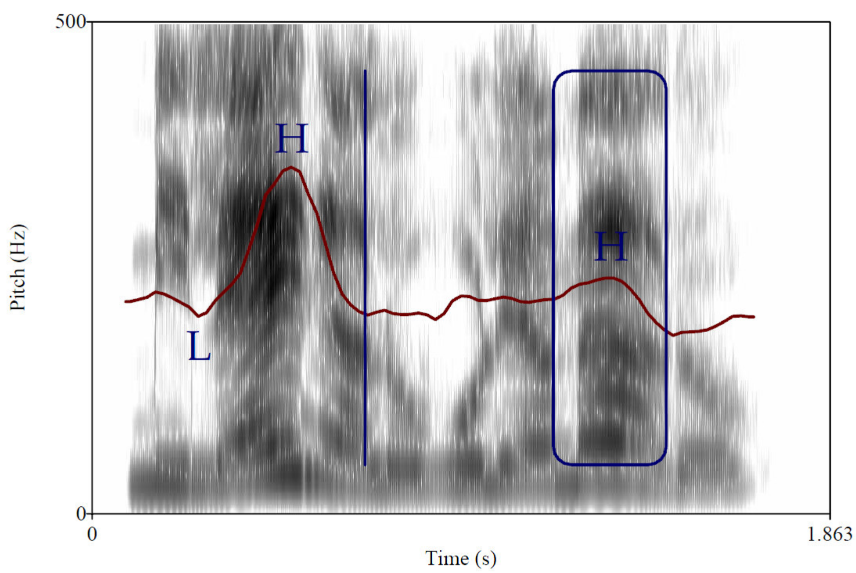

Figure 3. Utterance in (narrow focus) question context.

2. MATERIALS AND METHOD

2.1 Corpus

For our study we used a subset of the corpus described in (D'Imperio *et alii*, 2008). Three native speakers of Neapolitan Italian read 30 experimental stimuli and 70 fillers in a silent room. The stimuli consisted of five repetitions of three sentences designed without voiceless plosives, which were semantically plausible and syntactically quite similar: "Amelia dorme da nonna" (Amelia sleeps at grandma's), "Valeria viene alle nove" (Valeria arrives at 9) and "Milena lo vuole amaro". The target words were all feminine proper names, agents, subjects, trisillabic, and paroxitones, with the same syllabic structure (CV) for the tonic syllable and the same quality for its nucleus (/ɛ/). The sentences were presented together with a context paragraph, which had to be read silently; this made possible the elicitation of every sentence with two different pragmatic meanings. For example, the sentence "Milena wants her coffee unsweetened" would be interpreted (and uttered) by speakers as a Narrow Focus Question (*QNF*, meaning "Is it Milena, the one who drinks unsweetened coffee?", see Figure 3,) if preceded by the context:

> *After a family lunch, you're preparing coffee. You know that one of your cousins is on a diet and stays away from sugar, but you don't remember which one. You ask your aunt:...*

On the other hand, sentences preceded by the context:

> *In the afternoon, among friends, your brother is preparing coffee. He asks you whether your friends would like it sweetened or not. You don't know everybody's preferences, but only your girlfriend's. You answer:...*

would be interpreted (and uttered) as Partial Topic[4] Statements (*SPT*, meaning "As for Milena, she drinks it unsweetened; as for the others, I couldn't tell", see Figure 2,).

The experimental material consisted of 3 subjects x 3 sentences x 2 pragmatic contexts x 5 repetitions = 90 items in total.

2.2 Measures

Target words were manually labelled in syllables using PRAAT (Boersma & Weenink, 2009). The stressed syllable, which always had a CV structure, was also labelled in segments: the labels were *Os* for the beginning of the onset (and of the entire syllable), *Ns* for the beginning of the nucleus (or the end of the onset) and *Ne* for the end of the nucleus (and of the entire syllable)[5].

The rising f0 movement in the stressed syllable was characterized by measuring the height (in Hz) and the position in time of its starting and ending points (L and H)[6]. Hs were located at f0 maxima inside the stressed vowels, while the detection of Ls proved more challenging. A widely used automatic procedure is based on the detection of the local minima in the stressed syllable's onset, but we found this method too sensitive to microprosodic perturbations, which were irrelevant for our analysis. We determined that another strategy for the detection of Ls, the *two lines fitting* used for example in (D'Imperio, 2000), was not suited for our goals.

---

[4] For the notion of Partial Topic, see Büring (1997).
[5] See Figure 4: Os, Ns and Ne on x-axis.
[6] See Figure 4: y(L) and y(H) on y-axis for height, and x(L) and x(H) on x-axis for position.

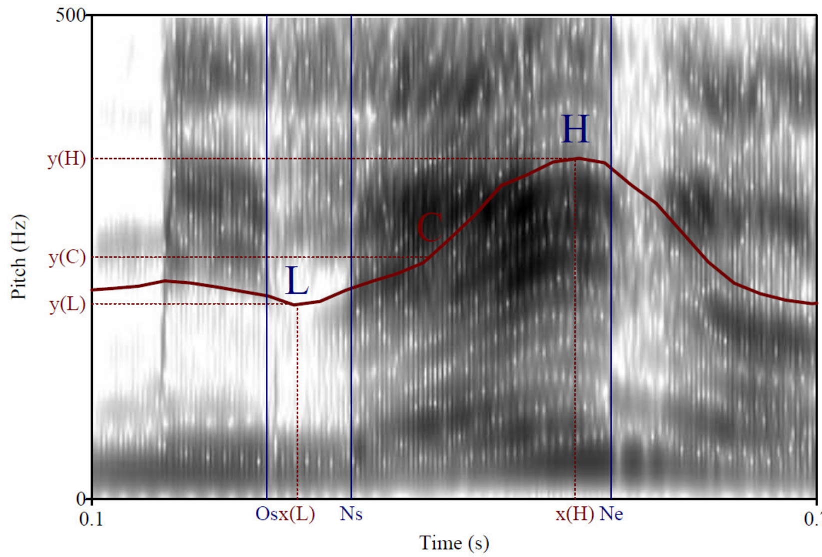

Figure 4. Measures.

With this technique, the region in which the L must be found (in our case, the f0 stretch from utterance start to H) is divided into steps. For each point, two straight lines are fitted with a linear regression to the contour on its left and on its right. The L is chosen as the point associated with the pair of lines leading to the smallest modelling error. Since the differences between a concave and a convex rise have consequences on modelling and errors, the algorithm often locates Ls away from the *elbow*, the point in which the f0 curve visibly bends upwards. Concave shapes tend to be associated to an L on the left of the real elbow, and for convex ones the L is detected on its right. This means of course that we would still have an index to express our differences in interpolation, but in this case the information is conveyed in an implicit and indirect way: different shapes are translated into different position of a same tonal target.

We decided to use a method which would ignore the specifically local features of the f0 contour (such as microprosodic minima) and at the same time avoid the implicit encoding of the global proprieties we were trying to characterize explicitly (as in the case of the two lines regression). Trying to find a compromise between these two constraints, we decided to locate the elbows at the point of maximal acceleration of the curve. Through the elaboration of an automated procedure in R (R Development Core Team, 2005), the L was located by inspecting the f0 second derivative, looking for sufficiently wide local maxima.

Although the L detection procedure is innovative, height and position of tone targets remain traditional measures. Besides these, we also calculated the height of the mid-point in time between L and H (*C*)[7]. This allowed us to calculate an index (based on Dombrowski & Niebuhr, 2005) which could express the type of interpolation between the two targets in a simple and explicit way; see §2.3.

[7] See Figure 4: y(C) on y-axis.

In conclusion, for every experimental item we measured the coordinates of L, C and H in the (time, f0) plane.

2.3 Indices

We used these coordinates to calculate various indices (see Table 1), and we ran a comparison of the ability of these indices to express the contrast between the aforementioned pragmatic categories. In addition to the traditional indexes of scaling (height of L and H) and alignment (distance of L and H from both start and end of, respectively, stressed syllable onset and nucleus), we calculated a curve index, expressed as the ratio of the difference between the heights of the intermediate and the starting points, and the difference between the heights of the end and starting points of the rise.

Index	**Description**	**Formula**
sL	**L s**caling	y(L)
aLs	**L a**lignment to **s**tart of stressed vowel onset	x(L) – Os
aLe	**L a**lignment to **e**nd of stressed vowel onset	Ns – x(L)
sH	**H s**caling	y(H)
aHs	**H a**lignment to **s**tart of stressed vowel nucleus	x(H) – Ns
aHe	**H a**lignment to **e**nd of stressed vowel nucleus	Ne – x(H)
sC	**C** (intermediate point in time between L and H) **s**caling	y(C)
Ci	**C**urve **i**ndex	$\frac{y(C)-y(L)}{y(H)-y(L)}$

Table 1. Indices.

3. RESULTS

The results show, for all subjects, weak or no correlations between the two pragmatic contexts (narrow focus question, QNF, and partial topic statement, SPT) and the indices usually employed in AM-based studies (alignment and scaling of tones). Two-sample Welch-Satterthwaite t-tests show that H scaling tends to be significantly different only for some speakers, while in other subjects only L scaling is significantly correlated to the two pragmatic contexts (see Figure 5).

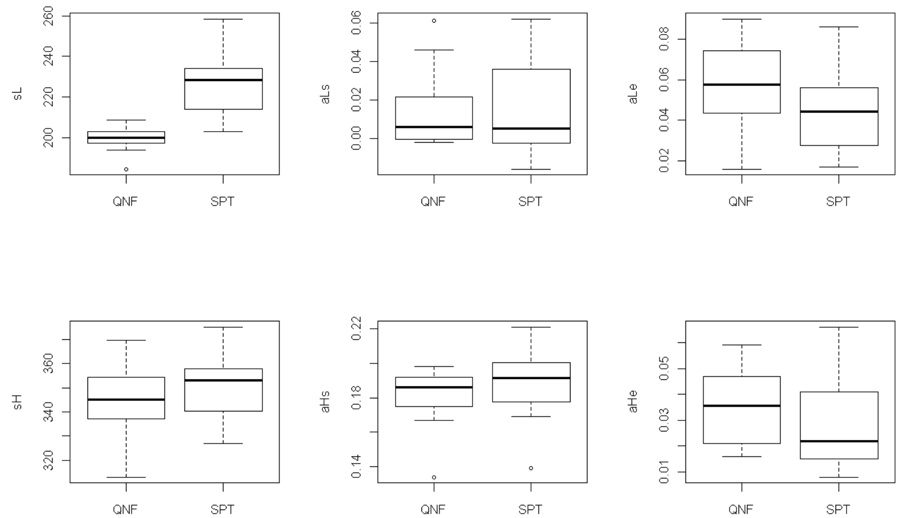

Figure 5. Box-plot for indexes sL, aLs, aLe, sH, aHs, aHe; speaker WP.

On the other hand, the curve index (and, consequently, the scaling of the midpoint in time between L and H) shows a strong correlation for all subjects with the pragmatic contexts ($p<0.001$). Moreover, considering that $Ci = 0.5$ would indicate a linear interpolation, we note a trend towards a convex interpolation for QNF contexts ($Ci < 0.5$), and a slight trend towards a concave interpolation for SPT contexts ($Ci > 0.5$); see Figure 6.

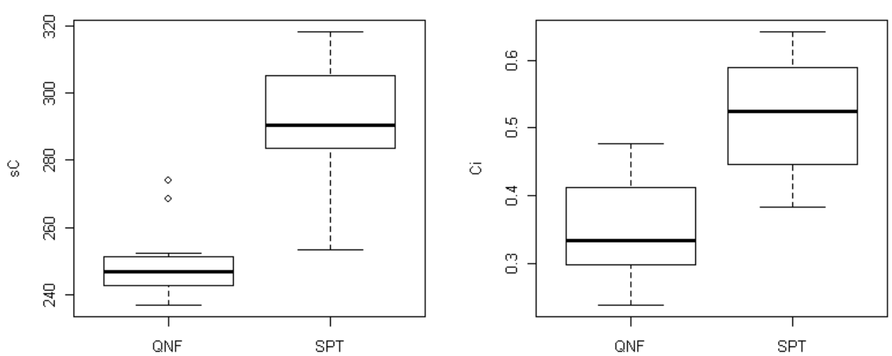

Figure 6. Box-plot for indexes sC and Ci; speaker WP.

4. DISCUSSION AND CONCLUSION

Our results confirm that two different kinds of interpolation between the two targets which compose a rising pitch accent (specifically, concave and convex) are correlated to two different pragmatic contexts (specifically, partial topic statement and narrow focus question). More generally, we can state that the analysis of the dynamic proprieties of the f0 contour allows for a better description of post-lexical meaning. If, as we mentioned at the beginning of §1, intonational phonology investigates the relationship between suprasegmental features and post-lexical meaning, we suggest that the AM model needs to be revised in order to a give proper place to the phonological value of these dynamic proprieties. This claim seems to be supported by other studies, as (Petrone & D'Imperio, 2008) on NI and (Petrone & Niebuhr, 2009) on German, in which the authors examine the importance of dynamic factors outside pitch accents.

In any case, we believe that such a revision cannot be proposed before an examination is made of the perceptual relevance of the contrasts found in this production experiment. As we said in §1 (see Figures 2 and 3), even if we only take into account the f0 contour, the most striking acoustic difference between the two utterances lies in the f0 movement corresponding to the last stressed syllable (i.e., the last pitch accent). In order to correctly retrieve the pragmatic meaning of these utterances, listeners could rely mainly or exclusively on this cue. The patterns we found in production, even if robust, could prove perceptually irrelevant.

In addition, the nature of the pragmatic contrast used in this experiment is another factor that could affect the usefulness of our results. The two contexts were chosen for the acoustic features of their realizations, i.e. for the clear differences in the interpolation between the targets of the first pitch accent, which still were equally aligned and scaled. We acknowledge that from a pragmatic point of view, our two contexts are far from being prototypically contrastive. Even if the modality value of the two contexts is clearly different (question *vs* statement), both share a "inconclusiveness" or "openness" feature. This feature is self-evident in the question context, but it can also be retrieved in the partial topic statement. In this case, a question about the properties of a set ("How would your friends like their coffee?") is answered to by providing information on the proprieties of a subset ("As for Milena, she drinks it unsweetened…"), implicitly excluding from the predication the properties of the complement subset ("…as for the others, I couldn't say").

Should we want to use a perceptual study to evaluate the phonological importance of the phonetic opposition highlighted in the present production study, we will need to take into account these pragmatic aspects too.

ACKNOWLEDGEMENTS

I would like to thank Pauline Welby and two of the anonymous reviewers for useful comments.

REFERENCES

Anderson, M., Pierrehumbert, J. & Liberman, M. (1984), Synthesis by rule of English Intonation patterns, in *Proceedings of IEEE-ICASSP 84*, 2.8.1-2.8.4.

Black, A. & Hunt, A. (1996), Generating F0 contours from ToBI labels using linear regression, in *Proceedings of ICSLP 96*, vol 3, 1385-1388.

Boersma, P. & Weenink, D. (2009), Praat: doing phonetics by computer, URL http://www.praat.org/.

Büring, D. (1997), *The Meaning of Topic and Focus – the 59th Street Bridge Accent*, London: Routledge.

D'Imperio, M. & Cangemi, F. (2009), Phrasing, register level downstep and partial topic constructions in Neapolitan Italian, in *Hamburger Studies on Multilingualism 10* (C. Gabriel & C. Lléo, editors), Amsterdam: John Benjamins.

D'Imperio, M. (1999), Tonal structure and pitch targets in Italian focus constitutents, in *Proceedings of ICPhS XIV*, San Francisco, U.S.A., Vol. 3, 1757–1760.

D'Imperio, M. (2000), *The role of perception in defining tonal targets and their alignment*, PhD Dissertation, The Ohio State University, U.S.A.

D'Imperio, M., Cangemi, F. & Brunetti, L. (2008), The phonetics and phonology of contrastive topic constructions in Italian, in *Third Conference on Tone and Intonation in Europe*, Lisbon, Portugal, 15-17 September 2008.

Dombrowski, E. & Niebuhr, O. (2005), Acoustic patterns and communicative functions of phrase-final F0 rises in German: activating and restricting contours, *Phonetica*, 62, 176-195.

Face, T. L. (2001), Focus and early peak alignment in Spanish intonation, *Probus*, 13, 223–246.

Hualde, J. I. (2000), Intonation in Spanish and the other Ibero-Romance languages:overview and status quaestionis, in *Romance phonology and variation* (C. Wiltshire & J. Camps, editors), Amsterdam: John Benjamins, 101–116.

Ladd, D. R. & Schepman, A. (2003), "Sagging transitions" between high pitch accents in English: experimental evidence, *Journal of Phonetics*, 31, 81-112.

Ladd, D. R. (2008), *Intonational Phonology, 2nd edition,* Cambridge: Cambridge University Press.

Petrone, C. & D'Imperio, M. (2008), From tones to tunes: the role of f0 prenuclear region in intonation identification in Italian, in *Third Conference on Tone and Intonation in Europe*, Lisbon, Portugal, 15-17 September 2008.

Petrone, C. & Niebuhr, O. (2009), The role of the prenuclear F0 region in the identification of German questions and statements, in *Fourth Conference on Phonetics and Phonology in Iberia*, Las Palmas, Spain, 17-19 June 2009.

Pierrehumbert, J. (1980), *The phonology and phonetics of English intonation*, PhD Disseration, Massachussets Intitute of Technology, U.S.A.

Pierrehumbert, J. (1981), Synthesizing Intonation, *Journal of the Acoustical Society of America*, 70, 985–995.

R Development Core Team (2005), R: A language and environment for statistical computing, URL http://www.R-project.org.