ON THE IMPORTANCE OF FINE ALIGNMENT AND SCALING DIFFERENCES IN PERCEPTION: THE CASE OF TURIN ITALIAN

Barbara Gili Fivela, Grazia M. Interlandi*, Antonio Romano*
CRIL-DReAM, Università del Salento-Lecce, *LFSAG, Università di Torino
barbara.gili@unisalento.it, intermic@tin.it, antonio.romano@unito.it

1. ABSTRACT

The present paper addresses a very specific issue that, despite being very particular, allows to draw attention to some key topics concerning the phonetic and phonological analysis of intonation within a laboratory phonology approach. In particular, the specific issue regards the way alignment and scaling differences in stimuli produced in the Turin variety of Italian are perceived by Turin Italian listeners, pointing to the need of differentiating a L*+H and a L*+>H pitch accent, who shows an anticipated peak. The general statements relate to findings in perception experiments that usually suggest or confirm which elements phonologically contrast within a system.

A perception experiment involving 10 subjects, speakers of Turin Italian, is described to check the need to differentiate phonological categories in the analysis of pitch accents found in exclamations and yes-no questions, both showing a pitch accent that involves a low target aligned within the nuclear syllable, followed by a rise to a high target. As it is often the case in the literature on phonological categories in intonation, the experiment aims to shed light on the presence or absence of a change in participants' linguistic interpretation of utterances depending on changes in intonation events and in particular on specific intonational features. Interpretation is in fact the cue to the need of proposing different phonological analyses and labels. However, while nice minimal pairs are usually investigated in the literature, in which it is often the case that even only one correlate needs to be changed to obtain different interpretations, in the case considered here the experimental design is particularly complicated as more than one correlate is crucial and, more importantly, the minimal pairs are not ideal for perception check. Indeed the patterns under investigation, showing possibly different pitch accents both resembling a L*+H, differ as for the edge tone composition (H- vs. L-) and, theoretically, such difference could account for eventual differences in the phonetic properties of the preceding pitch accent (L*+H). Nevertheless the theoretical and easy way to account for the differences observed in production has to be very carefully checked as soon as cross-linguistic comparison needs to be preserved. Indeed, the point of view assumed here is that it may be important to highlight any relevant pitch accent differentiation *per se*, especially as it may seem to be redundant in a variety in which, say, the difference appears to be already conveyed by phrase accents, but it may be clearly pertinent in another one in which, say, there is no difference in phrase accent composition. Thus, it is important to always code the presence of a pitch accent phonological differentiation to avoid misinterpretations in the comparison of phonological systems.

In fact, the perception experiment described in the paper involves not ideal minimal pairs, which imply the need to perform a composite set of manipulations on stimuli and, therefore, to offer listeners more than one series of stimuli, 4 in particular, in which various pitch characteristics of the contour under investigation are separately manipulated. The quite complicated perception results, observed as a whole, point out the need to differenti-

ate phonological analyses, as different meanings are found to be associated to a low edge tone specification depending on an earlier and lower vs. a later and higher, or more variable, alignment of the peak. Thus the pitch accent differentiation is taken to be phonologically relevant, even though it could seem to be redundant on the basis of production data analyzed so far.

2. INTRODUCTION

The recent work performed within the *Interactive Atlas of Romance Intonation* (IARI - http://prosodia.upf.edu/iari/) clearly highlighted the need of proposing phonological analyses of intonation that are consistent both inter- and intra-system. The project aimed at collecting and analyzing with the same methods data on various sentence types produced by speakers of different Romance languages. Within each language, speakers of different areas have been interviewed, collecting very precious data from a sociolinguistic point of view. One of the methods used to collect the linguistic *corpora* is the Discourse Completion Task (Blum-Kulka et al. 1989) that allows to gather spontaneous renditions which speakers produce to fit a given context. As for the analysis, a set of 31 sentence types collected in Italian and other Romance languages has been analyzed according to the Autosegmental-Metrical framework (Bruce 1977, Lieberman & Prince 1975, Pierrehumbert 1980).

It is well known that analyses performed within such framework are phonological, that is they may be suggested only after considering phonological contrasts within the system. On the other hand, it is also well known that the observation of high and low phonetic targets on the main phonetic correlate of intonation, that is the fundamental frequency (F0), often suggests the main information coded in the labels (e.g., a rising pitch accent showing two phonetic targets, high and low, that are both phonologically relevant is labelled as L+H rather than, say, H+L). The analysis of materials belonging to different languages, as well as to different varieties or to different vernaculars of the same language, clearly implies that the choices of specific labels and analyses for given intonational events have to be made keeping in mind both system-internal and cross-system aspects. It is indeed clear that system-internal choices that are inconsistent with others may be well justified by phonological features of a specific system. Nevertheless, it is also clear that cross-system inconsistencies that are not due to strong phonological arguments are not desirable, as they represent an obstacle to the cross-variety and cross-language comparison.

The work performed on Italian within the IARI project, which is partially described in Gili Fivela et al. (in press), clearly shows the effort to offer analyses and transcriptions while keeping in mind cross-variety as well as cross-language comparison. It is the first work on Italian in which this effort is systematic and, moreover, is applied to the analysis of a wide number of varieties and a representative set of sentence types[1]. One debated problem regards rising pitch accents showing both a low and a high target aligned within, or very close to, the nuclear syllable and labelled as either L+H* or L*+H. As usual within the Autosegmental-Metrical framework and as already mentioned, the analysis given and the choice of a label strongly depend on the contrasts identified within the specific system, that is, in our case, on system-internal aspects related to varieties of Italian. This is the reason why in the inventory of some varieties it is possible to find both a L+H* and a L*+H pitch

[1] Thirteen varieties are considered in the forthcoming paper, in which, then, interesting discussion is offered in relation to the consistency issue.

accent that actually show very similar phonetic characteristics. For instance, this is the case of the system described for the variety spoken in Naples, where the difference between the two pitch accents is very small (that is tonal targets differ in average alignment for few milliseconds), although crucial as to differentiate questions and narrow focus statements. However, in other varieties a more clear phonetic difference between the two accents is found. Thus, when considering rising bitonal pitch accents and their main phonetic characteristics, the overall situation is that depicted in Figure 1.

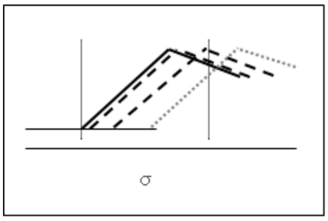

Figure 1: Schematic representation of alignment in L+H* and L*+H pitch accents (see text for discussion).

As the Figure shows, L+H* pitch accents usually shows a rise through the syllable starting from around the beginning of its onset and a peak that aligns either around the middle or in the second half of the syllable (see black filled line in Figure 1); the L*+H accent may be both very similar to the L+H*, showing a slightly later alignment (as in the Neapolitan Italian case, e.g. as shown by the first black dotted line from the left), or be quite different, with a clear low target within the first part of the syllable and a peak aligned even within the very same syllable. This also applies to the varieties spoken in Pescara or Turin whose typical pitch accents are indicated by second black dotted line from the left in Figure 1[2]. Moreover, consider that in other Romance languages, such as Catalan and Spanish (Frota & Prieto, in press), the label L*+H usually corresponds to pitch accents showing a much later aligned peak (see grey dotted line on the right in Figure 1).

A question may then arise as to the listener's perception of similar variations in pitch patterns when the patterns are contrasting within the specific variety s/he speaks. Some works in the literature deeply analyze the perception of acoustic differences in F0 characteristics in relation to phonological categories. Such investigations take into account findings related to perceptual thresholds, such as the fact that differences in pitch height in speech seem quite accurately perceived, while the perception sensitivity to alignment differences may vary considerably, for instance depending on spectral environment (House 1990). However, they rather concentrate on the perceptual boundaries that may be found between linguistic categories, often in order to check the relevance of specific features in relation to the identity of intonational categories (e.g., the role of tone target height in differentiating yes-no and wh-questions in Majorcan Catalan – Vanrell 2006). Notice that experimental investigations on these issues are usually performed by manipulating specific acoustic features to check the perceptual impact of their gradual phonetic variation on listeners' capability of identifying different linguistic categories and/or on their skill in stimuli discrimination (Gili Fivela 2008; 2009).

[2] Scaling features are not considered in the schema. As for the Turin variety, they are highlighted by Interlandi & Romano (2004); see below.

Along these lines, the present paper reports findings on the perception of phonetic variations in pitch patterns that are contrasting within a specific variety of Italian, that is the variety spoken in Turin, which have been previously described and tested in various contributions (Besana 1999, Romano & Interlandi 2002, Interlandi 2003, 2004, Interlandi & Romano 2004, Gili Fivela et al. in press). However, the very final goal of the paper is not discussing the perceptual boundaries between contrasting patterns. Rather, the goal is finding perceptual cues to the composition of contrasting pitch patterns. Indeed, the idea is that, though phonological analyses and labels have not to be necessarily transparent with respect to phonetic characteristics, they usually code the phonologically relevant features of a given category in terms of targets that compose it and, as already mentioned, on the relevance of specific features in relation to the identity of intonational categories. Finding out what is perceptually relevant is taken to shed light on which are such relevant features. Thus, similarly to what is done when, say, a rising-falling pitch pattern is given different phonological analyses depending on the fact that either the rise or the fall is considered as relevant to speakers[3], we try to get hints from the systematic investigation of speaker interpretation of acoustic variation. In order to do this, we try to identify the pertinent features speakers/listeners pay attention to when they encode/decode the phonological category.

In particular, the present paper investigates the variety of Italian spoken in Turin which shows a quite interesting phonological contrast between exclamations and yes-no questions that seems to be conveyed by complex phonological and phonetic differences. The pitch accent involved in both cases is rising and is characterized by a clear low target in the first part of the syllable, previously analyzed as L*+H. However the two sentence types differ as for the edge tone combination, in that the accent is followed by a low edge tone in exclamations and by both a high and a low edge tone in questions (that is, the analyses proposed are L*+H L% vs. L*+H HL%; see Interlandi 2003, *pace* Besana 1999, Gili Fivela et al. in press). Such phonological analyses could account for possible, even expected, phonetic differences in peak position, such as peak retraction in the case of an immediately following low target. Nevertheless, pitch accents features related to the peak position could be relevant in differentiating linguistic interpretation by themselves, and their role could just be masked by the edge tone specifications or be redundant with respect to them (e.g., an analysis could correspond to L+>H* vs. L*+H, that is anticipated vs. not anticipated peak). This point is even more important when considering that there are varieties of Italian in which exclamation and questions are given the same phonological analysis and in which, therefore, there is no difference in edge tone specification. At least in such varieties, a specific pitch accent feature could clearly be pertinent, rather than being, or seeming to be, redundant in conveying an important difference in linguistic interpretation[4]. This suggests the importance of checking the phonological analyses. It may indeed be important to study pitch accent differentiation, indipendenlty of the apparent redundancy vs. pertinence of a specific feature. Indeed, it is essential to explicitly code the presence of a pitch accent differentiation to avoid a wrong impression in the comparison of phonological systems.

[3] E.g., in the L+H* vs. H*+L analysis of the narrow contrastive focus in Italian or other languages (see Frota in press, and Gili Fivela et al. in press).
[4] One such variety is Pescara Italian (Gili Fivela et al. in press), which would indeed offer a nice minimal pair for perceptual investigation and hopefully it will be soon investigated in this direction.

3. TURIN ITALIAN: EXCLAMATIONS, QUESTIONS AND DECLARATIVE LISTS

The variety of Italian spoken in Turin shows a rising accent, characterized by a low tone associated to the nuclear syllable and followed by a high target, which may be followed by either a low or a high edge tone: the former is found in exclamations and the latter is found in yes-no questions (see the productions by two female Turin Italian speakers in Figure 2 and 3, upper and lower panel respectively, where the relative changes in duration and fundamental frequency range may be appreciated too). Given that the exclamation and yes-no question contours differ as for the presence of a low vs. high tone following the pitch accent, a question arises as to the pitch accents category preceding such tonal specification. Indeed, by observing the examples reported in Figure 2 and 3, it is clear that the end of the pitch accent rise, that is the peak, seems to be variably affected by the presence of the following tone specification. In particular, the peak alignment may either seem to be quite stable and early, irrespectively of the presence of a following low vs. high tone target (Figure 2, though it may be debatable where the pitch accent high target is) or be drastically anticipated in the case of a low rather than a high-low following target (Figure 3), that is in exclamations rather than in questions.

Notice that the examples reported in the figures are produced by two female speakers of Turin Italian, however other speakers suggest similar observations in that the high peak tends to align earlier in exclamations than in questions, where it is followed by a short plateau. Of course, tonal repulsion effects induced by a following low tone or speaker dependent strategies in general could account for the issue. On the other hand, a peak realized as the beginning of a F0 plateau – as it happens in the question contour – may be actually perceived as a later peak (D’Imperio 2000, D’Imperio et al. 2010, Gili Fivela & D’Imperio 2010); it would then be possible to argue that an early peak followed by a fall and a peak representing the beginning of a plateau make patterns very different from a perceptual point of view, ensuring different interpretations as expected for different phonological categories. Thus, one may wonder whether the pitch accent peak alignment affects the speaker perception of the utterance and is therefore a feature that is worth to code as phonological.

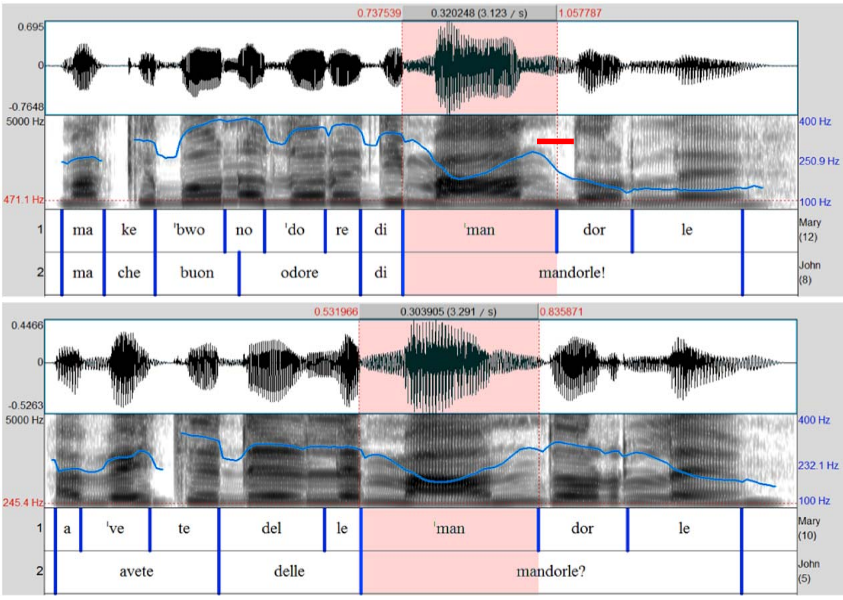

Figure 2: Exclamation (upper panel) and yes-no question (lower panel) by speaker F4.

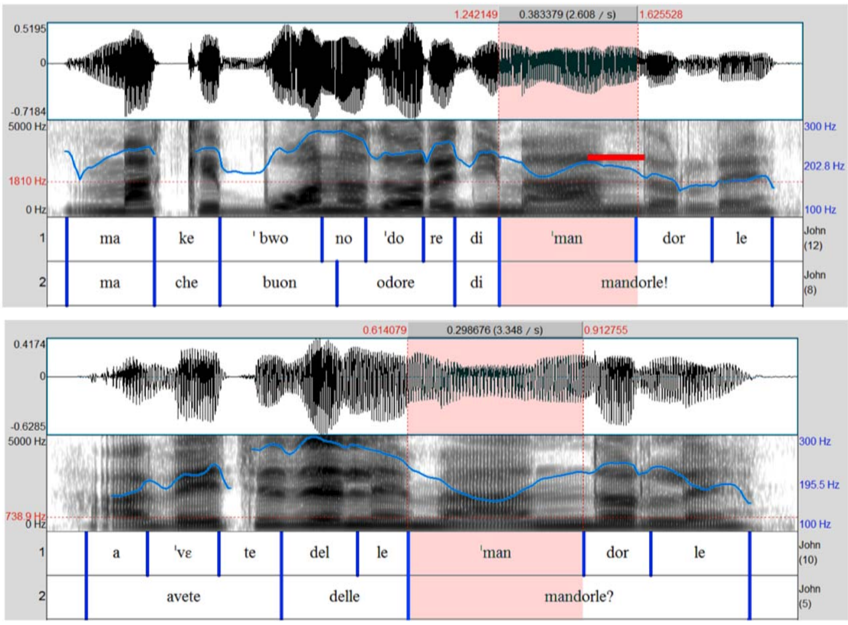

Figure 3: Exclamation (upper panel) and yes-no question (lower panel) by the speaker F5.

As already mentioned, the pitch accent under investigation is analyzed as L*+H, and it shows a clear low phonetic target by the beginning of the syllable followed by a rise. As Gili Fivela et al. (in press) discuss, it contrasts with L+H*, conveying for instance contrastive focus and showing earlier low and high tone targets (L+H* L%), and it is observed in exclamations (L*+H L%), yes-no questions (L*+H HL%, where L+H* LH% is also found), incredulity and counter-expectation yes-no-questions (for which the analysis is still provisional, though it seems to involve L*+H L!H%). In some cases, the pitch accent is also observed in declarative lists (where the accent shows a quite late peak alignment, more similarly to questions than to exclamation – see Figure 4; as in lists low tones may be found too, this observation is exploited in the experimental design, to try to point out the possible contribution of peak anticipation to conveying exclamation).

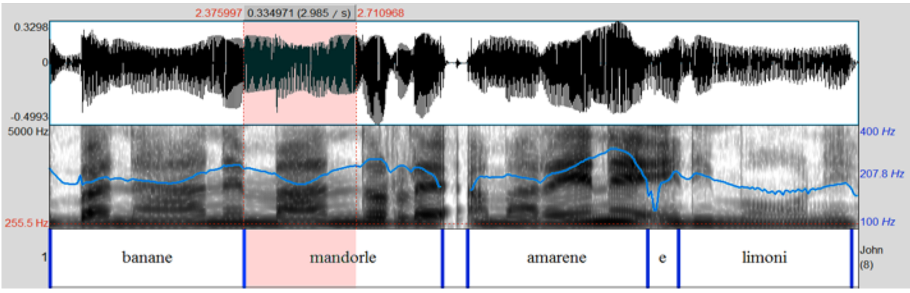

Figure 4: L*+H in a declarative list by speaker F5.

4. THE PERCEPTION EXPERIMENT

4.1. Goal and hypothesis

The experiment aims to investigate the perception of exclamation and yes-no questions in Turin Italian, trying to understand whether the difference between the two is conveyed by edge tone specification only (e.g., L*+H L% vs. L*+H HL%) or also by pitch accent features relating to the peak position (e.g., L*+>H vs. L*+H, that is anticipated vs. not anticipated peak). An implicit goal is then to tease apart the pitch accent and the edge tone contribution to check whether some pitch accent features are a side effect of differences in edge tone composition or not. Of course, as a consequence, the study aims to clarify the more general issue of perception of tonal targets and movements in Italian, and in the variety spoken in Turin in particular, with respect to changes in pitch height and timing.

The hypotheses are that peak alignment and scaling offer a relevant contribution to the perception of pitch pattern functions and that, in the case under investigation, differences in pitch accent alignment and scaling with respect to the overall pattern drive in fact the listener's interpretation; that is, a differentiation between L*+H and L*+>H is useful. The working assumption is that checking the pitch accent contribution independently from the edge tone contribution is possible if/when either the pitch accent and/or the edge tone change may convey other interpretations, as it happens in the case of Turin Italian. In such case, apart from the two interpretations under investigation, others may be available to carefully check the specific contribution of pitch accents and edge tones.

4.2. Methods

4.2.1.Base utterances and acoustic measures

The base utterances were chosen among those produced within the Atlas project by 4 Turin Italian speakers, who produced 4 renditions for each given context within the DCT, accordingly to the methods defined for data collection in Italian (Gili Fivela et al. in press). In particular, utterances produced by a female speaker were chosen and the set of stimuli was created by manipulating the acoustic characteristics of two base utterances, a yes-no question and an exclamation. Before being manipulated as for their acoustic characteristics, the two base utterances (see a) were shortened to the minimum to be potentially consistent with both an exclamation and a question interpretation, with no lexical material that could favor either interpretation (see corresponding sentences in b):

Exclamation: *Ma che buon odore di mandorle!* ‘What a good smell of almonds!’
Yes-no question: *Avete delle mandorle?* ‘Do you sell almonds?’

a. Exclamation: *odore di mandorle!* ‘smell of almonds!’
Yes-no question: *le mandorle?* ‘almonds?’

The acoustic differences between exclamations and questions that were considered in order to plan the needed manipulations are shown in Table 1. They correspond to differences of average values in exclamations and questions for measurements performed on productions by two female speakers (values on the right in the right column are those concerning the productions by the speaker who realized the utterances used as base for acoustic manipulation).

Parameter	**Excl. vs. Quest.**	**Difference (Δ)**
Syllable duration	E > Q	5-20%
F0 L1 (register)	E > Q	15-15 Hz
F0 H	E < Q	25-30 Hz
F0 Rise-range (L1toH)	E < Q	40-45 Hz
F0 Post-tonic to H	E < Q	100-40 Hz
F0 L2 (and final stretch)	E < Q	25-20 Hz
F0 in the posttonic	E < Q	80-140 Hz
F0 in the downdrift	E < Q	110-65 Hz
Latency ons-L1	E < Q	17-25 ms
Latency H-off (and % of syllable)	E < Q	71-195 ms (25-55%) (86to110% - 65to120%)
Latency H-L2	E > Q	100-80 ms;
Latency off-L2:	E < Q	165-170 ms

Table 1: Acoustic differences between exclamations and questions (see text for discussion).

Measurements relate to syllable duration, pitch accent target alignment[5] and both pitch accent target and edge tone scaling[6] and represented the base information for calculating

[5] Measurements were taken for the first low target (L1), the peak (H) and the second low target (L2) latency with respect to either syllable onset or syllable offset (ons-L1, H-off, off-L2) and with respect to each other (H-L2 latency).

[6] Measurements regarded the first low target (L1), the peak (H), the second low target (L2) and the final stretch fundamental frequency values (F0 L1, F0 H0, F0 L2 considered for final stretch too), the range of the F0 rise from the first low target to the peak (L1toH), the F0 difference between the peak and the posttonic syllable average F0 values (posttonic to H), the F0 as measured in the posttonic (F0 in posttonic), and the difference in F0 level measured at 20 ms after the onset of the final syllable (F0 in downdrift).

parameters for manipulation (see Figure 5, left and right schema respectively). Measurements and differences regarding the productions by the speaker who uttered the base utterance were given priority, although they were always checked with reference to the other speaker values, in order to ensure, whenever possible (but see footnote 9), that manipulated stimuli represented a continuum including, rather than excluding, values found in the other speaker production.

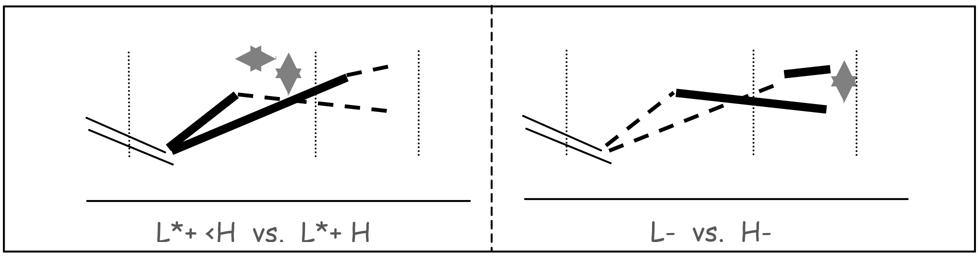

Figure 5: Schema of manipulation of peak alignment (left) and edge tone scaling (right).

4.2.2. Acoustic manipulations

Relevant acoustic characteristics were then manipulated, by considering the differences between their mean values in the exclamation and question interpretation and by calculating a number of steps of manipulation that allowed the realization of a gradual shift from one interpretation to the other. The number of steps was set to 8 and, thus, 8 manipulated stimuli were created starting from both the yes-no question and the exclamation base utterance[7]. Four series were created (see Figure 6), to specifically investigate the role of different parts of the contour and to try to tease the contribution of the peak features and of the following edge events apart. Given that duration was always kept unvaried, that is consistent with either the exclamation or the question rendition, manipulations were performed as follows.

- Series I (see Figure 6, top line, left schema): to investigate the effect due to the rise (and peak position) together with the following stretch, as a whole. Given the manipulation performed, peak alignment changes implies posttonic pitch height changes, meaning that early peak implies lower posttonic.
    - H was shifted in order to move from 65 to 120% of the syllable duration for the exclamation base, and from 120% to 65% for the question base utterance; peak height was shifted of 4 Hz for each step, increasing the peak height in the exclamation to question manipulation, and decreasing it in the other direction. This meant realizing 8 steps of 25 ms X 4 Hz difference from the exclamation base

[7] Creating stimuli from two base utterances was done in order to check for the contribution of other factors, to start with syllable duration, without explicitly investigating them. Indeed manipulating all acoustic correlates possibly related to the exclamation vs. question distinction was not possible. On the other hand, neutralizing them, for instance by imposing values that were intermediate between the exclamation and question reading, would have meant creating ambiguous stimuli, whose interpretation could then be problematic because of acoustic properties that were not directly related to the crucial features of these patterns.

and 8 steps of 20 ms X 4 Hz from the question base (differences in steps are due to differences in syllable duration in the base utterances);

- L2 was shifted from the peak by 12 ms for each step and both the target and the following stretch were shifted in 3Hz as for their frequency height;
- L1 was kept fixed with respect to the syllable onset, at a distance that was half of the average difference found for the two speakers in questions and exclamations (that is, 11 ms)[8] and it was shifted of 2 Hz for each step.

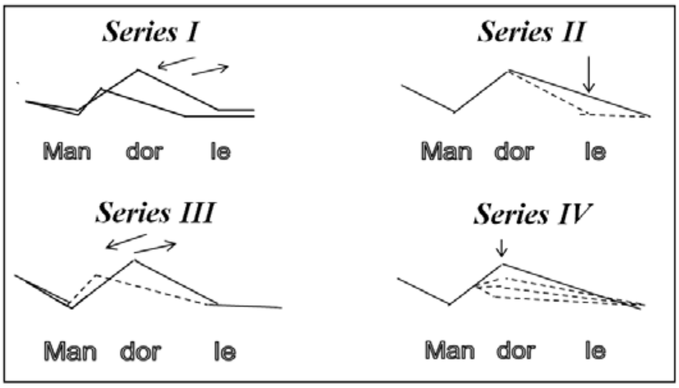

Figure 6: Schematic representation of acoustic manipulations performed in each series.

- Series II (see Figure 6, top line, right schema): to check the role of the F0 height of the very final syllable independently of the peak position, that is, to control if the final contour is enough to shift the perception from question to exclamation, given a late alignment of the peak; indeed, the final syllable is definitely low in exclamation while it has not to be so in questions, where the F0 in the final vowel may be falling rather than low. Manipulations were performed starting from a question-like contour (the original contour in question base stimuli and the final step of peak manipulation in series I in exclamation base stimuli).
  - H and L1 were kept in the position found in questions, for both alignment and scaling;
  - an artificial/hypothetical L2 target in the downdrift was identified 20 ms after the onset of the final syllable and its F0 height was shifted by 8 Hz[9] for each step of manipulation (the very final stretch was kept consistent); thus, a low F0 level for the final syllable in the final steps of manipulation was obtained, corresponding to an earlier low edge tone target.
- Series III (see Figure 6, bottom line, left schema): to check the role of peak alignment, given a low specification on the very final syllable, that is, to control if an early or

[8] However, L1 latency was calculated with respect to the syllable onset, but its position with respect to the offset is very different in the two utterance bases.

[9] In this case, the actual measures of the base utterances were used to avoid unnatural pitch excursions. In fact, the speaker who produced the actual base stimuli shows smaller variation than the other one; therefore this manipulation did not cover the variation observed in the other speaker too.

late alignment of the peak affect the utterance interpretation given an earlier low edge tone target.

- H and L1 were manipulated as in series 1;
- L2 and the final stretch correspond to the final step of manipulation in series II, that is the final syllable was completely low (as usually found in exclamation and differently from neutral questions).

- Series IV (see Figure 6, bottom line, right schema): to check the role of the height of the posttonic syllable, taken as a cue of both edge tone specification and peak alignment. Indeed, the posttonic syllable may be much lower in exclamation than in questions (although inter-speaker variation is strong in this respect) and this may have a relevant impact on speech perception. However, manipulating the posttonic F0 height offers a chance to check the perception of items in which peak alignment is early in the nuclear syllable: the interpolation between the low target at the beginning of the nuclear syllable and the high target in the posttonic, in fact, ends up in being not linear, as a high target within the nuclear syllable is made detectable, showing a lower/higher F0 frequency than the posttonic syllable depending on the step (the high target in the nuclear syllable allows to independently vary the characteristics of the nuclear and postnuclear syllables). Manipulations were performed starting from a question-like contour (the original contour in the question base stimuli and the final step manipulation in series I in the exclamation base stimuli).
  - The H position was identified on the basis of the latency between peak position and syllable onset as measured in exclamations; due to the impossibility of recovering in the corpus direct measurements of early peak realization in questions, the same latency was used to identify the hypothetical early peak position in the question base utterance, which gave a slightly later peak alignment in questions due to the shorter syllable duration. The L1 position was not manipulated;
  - The artificial/hypothetical L2 target was identified as the end of the rise of the question-like contour and its F0 height was shifted by 10 Hz for each step of manipulation (see also footnote 9) in order to obtain a low F0 level for the posttonic syllable in the final steps of manipulation; the very final stretch was kept consistent.

4.2.3. Subjects and recordings

Each series was considered as a single block and randomization was applied both inter- and intra-block. For each block, subjects were offered 5 repetitions of LPC synthesized stimuli, showing either manipulated (see above) or original acoustic correlates. Each block was preceded by a short training, in which subjects judged two repetitions of 4 stimuli representing the extremes of manipulations, and, during the test phase, subjects could interrupt the block for a pause, if they needed. Subjects participated on a voluntary basis and received no payment for their collaboration.

Twelve subjects participated in the perception experiment, although two could not be considered for analysis because they gave less than 90% correct answers for original items, used as controls. Subjects were all native speakers of the Turin variety, aged between 21 and 50, and were asked to listen and judge the stimuli offered in each series. In particular, subjects were asked to identify the categories the stimuli belonged to among 4 options, that is, exclamation (*Exclamation*), question (*Question*), statement (*Statement*) and question suggesting incredulity and counter-expectation (*QuestionS*).

4.2.4. Statistics

Two-way multivariate ANOVAs were run to check the influence of the Base utterance factor (2 levels), the Step of manipulation factor (8 levels) and their interaction on each answer option; one-way ANOVAs were carried out to investigate the interaction, by evaluating the effect of Step of manipulation for each answer option. Moreover, the effect of Step of manipulation was further investigated by means of the Tuckey post-hoc test and by one sample *t*-tests that were carried out to test the mean percent of selected option for each stimulus against chance level (25%). In all cases, the alpha level was set at p=.05. In the next section, rather than reporting and discussing the effect of Step of manipulation on each single option, results of the *t*-tests are presented in tables and discussed in the text, together with the corresponding effect size (Cohen's *d*)[10].

4.3. Results

- Series I

Figure 7 shows the average number of answers in favor of the given options for each stimulus, for both Question base (left) and Exclamation base stimuli (right plot).

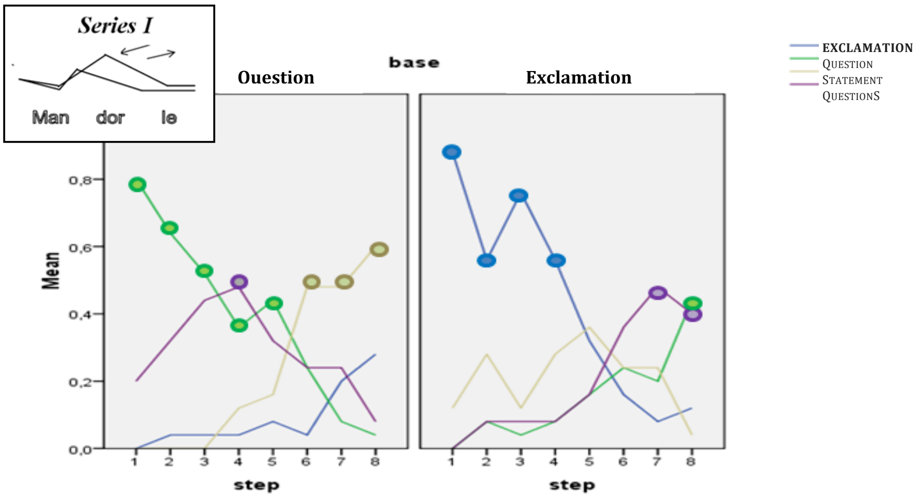

Figure 7: Average number of answers (equivalent to percentages) in favor of each given option (lines/colors) for each stimulus (x axis), created from the Question base (left plot) and the Exclamation base utterance (right plot) – bullets show significant *t*-test results.

---

[10] The effect size was computed as the Cohen's *d* at http://www.uccs.edu/lbecker/index.html#Calculate%20d%20and%20r%20using%20t%20values%20%28separate%20groups%29, using the *t*-test value for between subjects *t*-test and the degrees of freedom; Cohen's *d* values were interpreted on the basis of the table found at http://www.uccs.edu/lbecker/effect-size.html and reported in appendix.

Results of two-way multivariate ANOVAs showed that, apart from the Base factor with respect to the *Statement* option (n.s.), all factors are strongly significant, together with their interaction (*Exclamation* option: Base [F(1,784)=190.721; p=.000], Step [F(7, 784)=12.067; p=.000], Interaction [F(7, 784)=30.074; p=.000]; *Question* option: Base [F(1, 784)=74.649; p=.000], Step [F(7, 784)=4.552; p=.000], Interaction [F(7, 784)=25.910; p=.000]; *QuestionS* option: Base [F(1, 784)=8.698; p=.000], Step [F(7, 784)=3.504; p=.000], Interaction [F(7, 784)=11.381; p=.000]; *Statement* option: Base [F(1, 784)=.563; p=.453], Step [F(7, 784)=10.943; p=.000], Interaction [F(7, 784)=15.046; p=.000]).

a) Step Question base	Option (mean)	*t*-test result	Effect size
1	Question (M=.80)	[t(49)= 9.625, p= .000]	2.75
2	Question (M=.64)[11]	[t(49)= 5.688, p= .000]	1.62
3	Question (M=.52) QuestionS (M=.44)	[t(49)= 3.783, p= .000] [t(49)= 2.679, p= .010]	1.08 0.76
4	QuestionS (M=.48)[12]	[t(49)= 3.223, p= .002]	0.92
5	Question (M=.44)[13]	[t(49)= 2.679 p= .010]	0.76
6	Statement (M=.48)	[t(49)= 3.223 p= .002]	0.92
7	Statement (M=.48)	[t(49)= 3.223 p= .002]	0.92
8	Statement (M=.60)[14]	[t(49)= 5.001 p= .000]	1.42

b) Step Exclamation base	Option (mean)	*t*-test result	Effect size
1	Esclamation (M=.88)	[t(49)= 13.571, p= .000]	3.87
2	Esclamation (M=.56)[15]	[t(49)= 4.372, p= .000]	1.24
3	Esclamation (M=.76)	[t(49)= 8.359, p= .000]	2.38
4	Esclamation (M=.56)[16]	[t(49)= 4.372, p= .000]	1.24
5	not significant[17]	-	-
6	not significant[18]	-	-
7	QuestionS (M=.48)	[t(49)= 3.223 p= .002]	0.92
8	QuestionS (M=.40) Question (M=.44)	[t(49)= 2.143 p= .037] [t(49)= 2.679 p= .010]	0.61 0.76

Table 2a,b: Series I. Results of one sample *t*-tests for each step of manipulation (selected answer against chance level - 25%) and corresponding effect size: Question base stimuli (table a) and Exclamation base stimuli (table b).

One-way ANOVAs carried out to investigate the interaction showed that, for both the question base and the exclamation base, the step factor is always significant (for the question base: *exclamation* option [F(7,392)=6.315; p=.000]; *question* option [F(7,392)=19.524;

[11] *QuestionS* (M=.32) was not significant.
[12] *Question* (M=.36) was not significant.
[13] *QuestionS* (M=.32) was not significant.
[14] *Exclamation* (M=.28) was not significant.
[15] *Statement* (M=.28) was not significant.
[16] *Statement* (M=.28) was not significant.
[17] Exclamation (M=.32) and *Statement* (M=.36) were not significant.
[18] *QuestionS* (M=.36) was not significant.

p=.000]; *questions* option [F(7,392)=4.305; p=.000]; *statement* option [F(7,392)=25.026; p=.000]; for the exclamation base: exclamation option [F(7,392)=27.693; p=.000]; *question* option [F(7,392)=8.565; p=.000]; *questions* option [F(7,392)=11.906; p=.000]; *statement* option [F(7,392)=3.554; p=.001]).

Results of *t*-tests checking the mean percent of selected options for each stimulus against chance level (25%) are reported in table 2a for the Question base stimuli (see also bullets in Figure 7, left). They show that the mean for the *Question* option was significantly above the chance level for stimuli at early stages of manipulation, that is for later and higher peaks, while at later stages of manipulation, that is for earlier and lower peaks, it was the mean for the *Statement* option to be significantly chosen (apart significance and large effect size for *QuestionS* at intermediate step 4). Moreover, notice that effect size is large, though, consistently with average scores, it gradually decreases for the *Question* option and then gradually increase for the *Statement* option. Results for the Exclamation base stimuli are shown in table 1b (see also bullets in Figure 7, right) and show that the mean for the *Exclamation* option was significantly above the chance level for stimuli at early stages of manipulation, that is for earlier and lower peaks, while at later stages of manipulation, that is for later and higher peaks, it was the mean for the *QuestionS* option and, at the very final step, for the *Question* option to be significantly chosen. Moreover, notice that effect size is large, though, quite consistently with average scores, it tends to gradually decrease for the both the *Exclamation* and the *Question* option.

- Series II

Figure 8 shows the average number of answers in favor of the given options for each stimulus, for both Question base (left) and Exclamation base stimuli (right plot).

Results of two-way multivariate ANOVAs showed that, apart from the Step factor with respect to the *QuestionS* option (n.s.), all factors are strongly significant, together with their interaction for the *Exclamation* option (*Exclamation* option: Base [F(1,784)= 46.284; p=.000], Step [F(7,784)=4.824; p=.000], Interaction [F(7,784)=3.373; p=.000]; *Question* option: Base [F(1,784)=11.041; p=.001], Step [F(7,784)=7.039; p=.000], Interaction [F(7,784)=.926; p=.486]; *QuestionS* option: Base [F(1,784)=116.595; p=.000], Step [F(7, 784)=1.339; p=.2], Interaction [F(7, 784)=1.574; p=.1]; *Statement* option: Base [F(1, 784)=99.625; p=.000], Step [F(7,784)=6.128; p=.000], Interaction [F(7,784)=1.643; p=.1]).

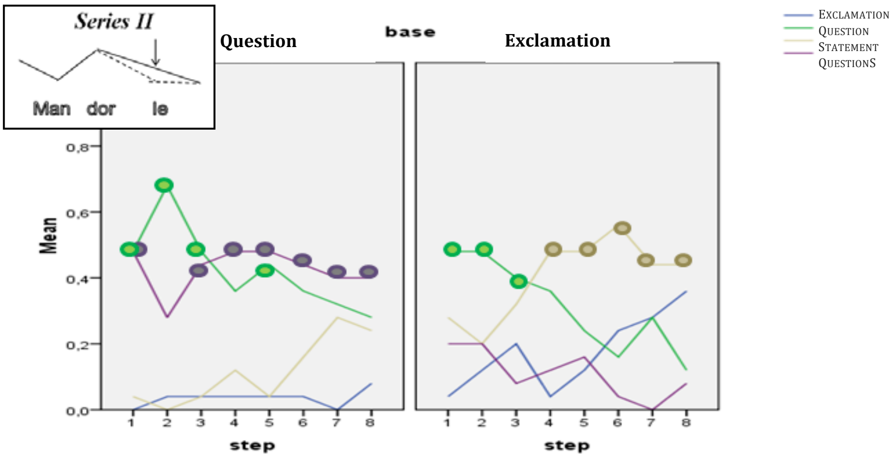

Figure 8: Average number of answers (equivalent to percentages) in favor of each given option (lines/colors) for each stimulus (x axis), created from the Question base (left plot) and the Exclamation base utterance (right plot) – bullets show significant *t*-test results

One-way ANOVAs carried out to investigate the interaction showed that, for the Question base, the Step factor is significant for the *Question* and *Statement* options, while for the Exclamation base it is always significant (for the Question base: *Exclamation* option [F(7,392)=.970; p=.453]; *Question* option [F(7,392)=3.397; p=.002]; *QuestionS* option [F(7,392)=.926; p=.486]; *Statement* option [F(7,392)=5.682; p=.000]; for the Exclamation base: *Exclamation* option [F(7,392)=4.881; p=.000]; *Question* option [F(7,392)=4.660; p=.000]; *QuestionS* option [F(7,392)=2.824; p=.007]; *Statement* option [F(7,392)=3.155; p=.003]).

a) Step Question base	Option (mean)	*t*-test result	Effect size
1	Question (M=.48) QuestionS (M=.48)	[t(49)= 3.223 p= .002] [t(49)= 3.223 p= .002]	0.92 0.92
2	Question (M=.68)[19]	[t(49)= 6.453, p= .000]	1.84
3	Question (M=.48) QuestionS (M=.44)	[t(49)= 3.223 p= .002] [t(49)= 2.769, p= .01]	0.92 0.76
4	QuestionS (M=.48)[20]	[t(49)= 3.223, p= .002]	0.92
5	Question (M=.44) QuestionS (M=.48)	[t(49)= 2.769, p= .01] [t(49)= 3.223 p= .002]	0.76 0.92
6	QuestionS (M=.44) [21]	[t(49)= 2.769, p= .01]	0.76
7	QuestionS (M=.40) [22]	[t(49)= 2.143 p= .03]	0.61
8	QuestionS (M=.40) [23]	[t(49)= 2.143 p= .03]	0.61

[19] *QuestionS* (M=.28) was not significant.
[20] *Question* (M=.36) was not significant.
[21] *Question* (M=.36) was not significant.
[22] *Question* (M=.32) was not significant.
[23] *Question* (M=.28) was not significant.

b) Step Exclamation base	Option (mean)	*t*-test result	Effect size
1	Question (M=.48)[24]	[t(49)= 3.223, p= .002]	0.92
2	Question (M=.48)	[t(49)= 3.223, p= .002]	0.92
3	Question (M=.40) [25]	[t(49)= 2.143 p= .03]	0.61
4	Statement (M=.48)	[t(49)= 3.223, p= .002]	0.92
5	Statement (M=.48)	[t(49)= 3.223, p= .002]	0.92
6	Statement (M=.56)	[t(49)= 4.372, p= .000]	1.24
7	Statement (M=.44) [26]	[t(49)= 2.769, p= .01]	0.76
8	Statement (M=.44) [27]	[t(49)= 2.769, p= .01]	0.76

Table 3a,b: Series II. Results of one sample *t*-tests for each step of manipulation (selected answer against chance level - 25%) and corresponding effect size: Question base stimuli (table a) and Exclamation base stimuli (table b).

Results of *t*-tests checking the mean percent of selected options for each stimulus against chance level (25%) are reported in table 3a for Question base stimuli (see also Figure 8 left). Results of *t*-tests show that only the mean for the *Question* and *QuestionS* options were significantly above the chance level, however the latter is the statistically significant choice at later stages of F0 manipulation, that is for late peak and lower F0 values of the final syllable (corresponding to earlier low edge tone alignment). Consistently, the effect size is larger for *Question* in steps 2 and 3, while it is larger for *QuestionS* at later steps. Results for the Exclamation base stimuli are shown in table 3b (see also bullets in Figure 8, right) and show that the mean for the *Question* option was significantly above the chance level for stimuli at early stages of manipulation, that is for late peak and not low F0 values of the final syllable (corresponding to later low edge tone alignment), while the mean for the *Statement* option was significantly above the chance level at later stages of manipulation, that is for late peak and lower F0 values of the final syllable (corresponding to earlier low edge tone alignment). Moreover, notice that effect size is large and, consistently with average scores, it tends to gradually decrease for the *Question* option and then increase for the *Statement* option, though in the very final steps it gradually decreases again (possibly due to the increasing number of *Exclamation* choices)[28].

[24] *Statement* (M=.28 was not significant.
[25] *Statement* (M=.32) was not significant.
[26] *Question* and *Exclamation* (M=.28) were not significant.
[27] *Exclamation* (M=.36) was not significant.
[28] The *Statement* option receives always an insignificant number of choices for the question base stimuli and, even if the option should be fine with high boundaries too, it receives an insignificant number of choices in early steps of manipulation for exclamation base stimuli too. This suggests that in these cases there may be other cues interfering with the statement interpretation.

- Series III

Figure 9 shows the average number of answers in favor of the given options for each stimulus, for both Question base (left) and Exclamation base stimuli (right plot).

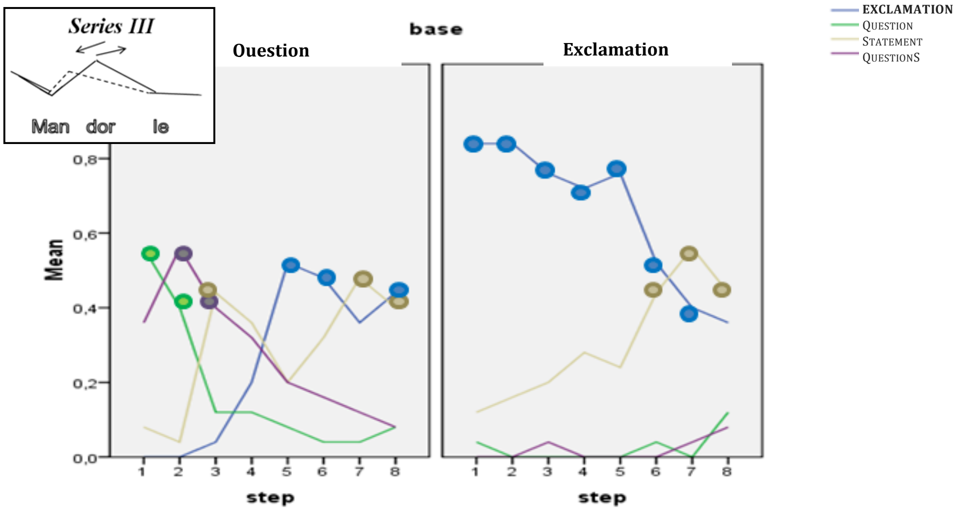

Figure 9: Average number of answers (equivalent to percentages) in favor of each given option (lines/colors) for each stimulus (x axis), created from the Question base (left plot) and the Exclamation base utterance (right plot) – bullets show significant *t*-test results.

Results of two-way multivariate ANOVAs showed that, apart from the Base factor and the Interaction with respect to the *Statement* option and the Step factor factor and the Interaction with respect to the *QuestionS* option (n.s.), all factors are strongly significant, together with their interaction (*Exclamation* option: Base [F(1,784)= 179.255; p=.000], Step [F(7, 784)=4.132; p=.000], Interaction [F(7, 784)=20.906; p=.000]; *Question* option: Base [F(1, 784)=68.245; p=.001], Step [F(7, 784)=13.543; p=.000], Interaction [F(7, 784)=14.030; p=.000]; *QuestionS* option: Base [F(1, 784)=130.851; p=.000], Step [F(7, 784)=5.742; p=.2], Interaction [F(7, 784)=8.157; p=.1]; *Statement* option: Base [F(1, 784)=.237; p=.627], Step [F(7, 784)=11.756; p=.000], Interaction [F(7, 784)=1.921; p=.1]).

One-way ANOVAs carried out to investigate the interaction showed that, for both the Question and the Exclamation base, the Step factor is always significant (for the Question base: *Exclamation* option [F(7,392)=16.470; p=.000]; *Question* option [F(7,392)=15.750; p=.000]; *QuestionS* option [F(7,392)=7.438; p=.000]; *Statement* option [F(7,392)=7.354; p=.000]; for the Exclamation base: *Exclamation* option [F(7,392)=9.535 p=.000]; *Question* option [F(7,392)=3.868; p=.000]; *QuestionS* option [F(7,392)=2.383; p=.007]; *Statement* option [F(7,392)=6.346; p=.021]).

a) Step Question base	Option (mean)	*t*-test result	Effect size
1	Question (M=.56)[29]	[t(49)= 4.372, p= .000]	1.24
2	Question (M=.40) QuestionS (M=.56)	[t(49)= 2.143 p= .03] [t(49)= 4.372, p= .000]	0.61 1.24
3	QuestionS (M=.40) Statement (M=.44)	[t(49)= 2.143 p= .03] [t(49)= 2.769, p= .01]	0.61 0.76
4	not significant[30]	-	-
5	Exclamation (M=.52)	[t(49)= 3.783, p= .000]	1.08
6	Exclamation (M=.48)[31]	[t(49)= 3.223, p= .002]	0.92
7	Statement (M=.48)[32]	[t(49)= 3.223, p= .002]	0.92
8	Exclamation (M=.44) Statement (M=.40)	[t(49)= 2.769, p= .01] [t(49)= 2.143, p= .03]	0.76 0.61

b) Step Exclamation base	Option (mean)	*t*-test result	Effect size
1	Esclamation (M=.84)	[t(49)= 11.265, p= .000]	3.21
2	Esclamation (M=.84)	[t(49)= 11.265, p= .000]	3.21
3	Esclamation (M=.76)	[t(49)= 8.359, p= .000]	2.38
4	Esclamation (M=.72)[33]	[t(49)= 7.327, p= .000]	2.09
5	Esclamation (M=.76)	[t(49)= 8.359, p= .000]	2.38
6	Exclamation (M=.52) Statement (M=.44)	[t(49)= 3.783, p= .000] [t(49)= 2.679, p= .010]	1.08 0.76
7	Statement (M=.56) Exclamation (M=.40)	[t(49)= 4.372, p= .000] [t(49)= 2.143, p= .03]	1.24 0.61
8	Statement (M=.44)	[t(49)= 2.769, p= .01]	0.76

Table 4a,b: Series III. Results of one sample *t*-tests for each step of manipulation (selected answer against chance level - 25%) and corresponding effect size: Question base stimuli (table a) and Exclamation base stimuli (table b).

Results of *t*-tests checking the mean percent of selected options for each stimulus against chance level (25%) are reported in table 4a for the Question base stimuli (see also bullets in Figure 9, left). At early stages of manipulation, that is with later peak alignment and a low pitch final syllable (corresponding to earlier low edge tone alignment), the number of choices are significantly in favor of the *Question* option and then, from step 2, the *QuestionS* option is chosen significantly above the chance level and the *Statement* option appears to be significant for step 3; however, for stimuli at later stages of manipulation, that is with earlier peak positions, only the *Exclamation* and the *Statement* options are chosen significantly above chance. The effect size is larger for *Question* and *QuestionS* at early stages and *Exclamation* at later stages. Results for the Exclamation base stimuli are shown in table 4b (see also bullets in Figure 9, right) and show that the mean for the *Exclamation* option was significantly above the chance level for stimuli at early stages of manipulation, that is with early peak alignment and a low pitch final syllable, while the *Statement* option becomes significant at later stages of manipulation, with *Exclamation* that is still signifi-

[29] *QuestionS* (M=.36) was not significant.
[30] *QuestionS* (M=.32) and *Statement* (M=.36) were not significant.
[31] *Statement* (M=.32) was not significant.
[32] *Exclamation* (M=.36) was not significant.
[33] *Statement* (M=.28) was not significant.

cantly chosen in step 6 and 7, though in a drastically lower number of cases, and is no more significantly chosen in step 8. Moreover, notice that effect size is large and, consistently with average scores, it tends to gradually decrease for the *Exclamation* option and then increase for the *Statement* option.

- Series IV

Figure 10 shows the average number of answers in favor of the given options for each stimulus, for both Question base (left) and Exclamation base stimuli (right plot).

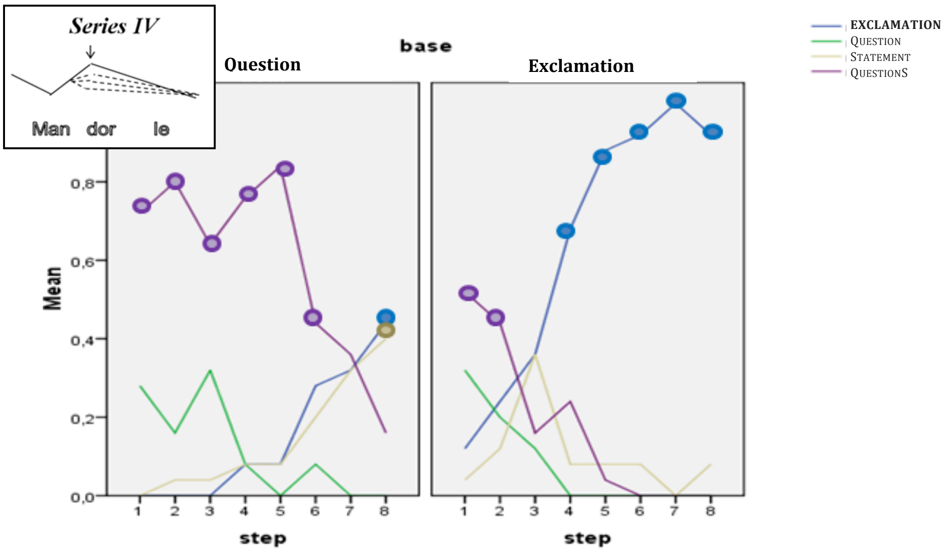

Figure 10: Average number of answers (equivalent to percentages) in favor of each given option (lines/colors) for each stimulus (x axis), created from the Question base (left plot) and the Exclamation base utterance (right plot) – bullets show significant *t*-test results.

Two-way multivariate ANOVAs were run to check the influence of the Base utterance factor, the Step of manipulation and their interaction on each answer option. Results showed that, apart from the Base factor with respect to the *Question* and *Statement* options and the Step factor with respect to the *QuestionS* option (n.s.), all factors are significant (*Exclamation* option: Base [F(1,784)= 417.936; p=.000], Step [F(7, 784)=54.085; p=.000], Interaction [F(7, 784)=11.911; p=.000]; *Question* option: Base [F(1,784)=3.245; p=.072], Step [F(7,784)=18.531; p=.000], Interaction [F(7, 784)=2.185; p=.034]; *QuestionS* option: Base [F(1,784)=227.774; p=.000], Step [F(7, 784)=27.315; p=.2], Interaction [F(7,784)=6.646; p=.000]; *Statement* option: Base [F(1,784)=3.322; p=.06], Step [F(7,784)=5.547;p=.000], Interaction [F(7,784)=11.746; p=.000]).

One-way ANOVAs carried out to investigate the interaction showed that, for both the Question and the Exclamation base, the Step factor is always significant (for the Question base: *Exclamation* option [F(7,392)=14.276; p=.000]; *Question* option [F(7,392)=9.062; p=.000]; *QuestionS* option [F(7,392)=15.297; p=.000]; *Statement* option [F(7,392)=9.994; p=.000]; for the Exclamation base: *Exclamation* option [F(7,392)=48.388; p=.000]; *Question* option [F(7,392)=12.238; p=.000]; *QuestionS* option [F(7,392)=19.987; p=.000]; *Statement* option [F(7,392)=6.950; p=.021]).

a) Step Question base	Option (mean)	*t*-test result	Effect size
1	QuestionS (M=.72)[34]	[t(49)= 7.327, p= .000]	2.09
2	QuestionS (M=.80)	[t(49)= 9.625, p= .000]	2.75
3	QuestionS (M=.64)[35]	[t(49)= 5.688, p= .000]	1.62
4	QuestionS (M=.76)	[t(49)= 8.359, p= .000]	2.38
5	QuestionS (M=.84)	[t(49)= 11.265, p= .000]	3.21
6	QuestionS (M=.44)[36]	[t(49)= 2.679, p= .01]	0.76
7	not significant[37]	-	-
8	Exclamation (M=.44) Statement (M=.40)	[t(49)= 2.769, p= .01] [t(49)= 2.143, p= .03]	0.76 0.61

b) Step Exclamation base	Option (mean)	*t*-test result	Effect size
1	QuestionS (M=.52)	[t(49)= 3.783, p= .000]	1.08
2	QuestionS (M=.44)	[t(49)= 2.679, p= .01]	0.76
3	not significant[38]	-	-
4	Esclamation (M=.68)	[t(49)= 6.453, p= .000]	1.84
5	Esclamation (M=.88)	[t(49)= 13.571, p= .000]	3.87
6	Esclamation (M=.92)	[t(49)= 17.288, p= .000]	4.93
7	Exclamation (M=1)	-[39]	-
8	Esclamation (M=.92)	[t(49)= 17.288, p= .000]	4.93

Table 5a,b: Series IV. Results of one sample *t*-tests for each step of manipulation (selected answer against chance level - 25%) and corresponding effect size: Question base stimuli (table a) and Exclamation base stimuli (table b).

Results of *t*-tests checking the mean percent of selected option for each stimulus against chance level (25%) are reported in table 5a for the Question base stimuli (see also bullets in Figure 10, left). From the very first step of manipulation, as soon as a high target independent of the end of the rise was introduced, participant's perception significantly favored the *QuestionS* option; only at the very final stage of manipulation, that is when both a clear anticipated peak and a very low posttonic syllable were audible (corresponding to earlier low edge tone alignment), listeners significantly shifted their choices towards the *Exclamation* and the *Statement* options. The effect size for *QuestionS* is very large for quite a high number of steps of manipulation and it is then slightly higher for *Exclamation* than for *Statement*. Results for the Exclamation base stimuli are shown in table 5b (see also bullets in Figure 10, right) and show again that participant chose the *QuestionS* option as soon as a high target independent of the end of the rise was introduced. However, for stimuli obtained from the Exclamation base, it is clear that subjects shift towards another option at earlier stages in comparison to what happens for Question base stimuli. In particular, they clearly opt for the *Exclamation* option since the 4th step of manipulation. This suggests that,

[34] *Question* (M=.28) was not significant.
[35] Question (M=.32) was not significant.
[36] *Exclamation* (M=.28) was not significant.
[37] *QuestionS* (M=.36), *Exclamation* (M=.32) and *Statement* (M=.32) were not significant.
[38] *Exclamation* and *Statement* (M=.36) were not significant.
[39] The *t* value could not be computed because the standard deviation is 0, as all subjects always chose the *Exclamation* option.

apart from the height of the posttonic, other features, such as syllable duration, may play a very relevant role in affecting subjects perception.

4.4. Discussion

Results obtained for the four series are schematically summarized in Table 6.

Series I shows that the late or early alignment of the peak, its higher or lower F0 height and a later or an early end of the following fall appear to be crucial to shift subject perception from a question or an incredulous question interpretation, mainly depending on the base stimulus, to a statement or exclamation interpretation, depending again on the base stimulus. Thus, schematically, a late vs. early alignment of the peak configuration affects subjects' interpretation of the stimuli with a quite clear association of later configuration to questions. However, given the manipulation characteristics and the resulting intonation contours, results concerning this series do not allow to tease pitch accent (and in particular its peak position) and postaccentual contribution apart**.** Moreover, the relevance of the base stimulus underlines that differences concerning parameters which were not varied (such as syllable duration or the first low target position with respect to the end of the nuclear syllable) seem to be important in affecting the neutral question vs. incredulous question choice as well as the exclamation vs. statement choice.

Consistently, results for Series II, where stimuli had always a late peak alignment (in the postnuclear syllable) and changed as for the final syllable height (corresponding to earlier/later low edge tone alignment), clearly showed that the low specification on the final syllable is not enough to shift the interpretation towards exclamation, while it is sufficient to shift it from question to incredulous questions or statement, depending on the base utterance. Actually, results for this series (as well as for the question base stimuli in series III) show a higher degree of ambiguity, as manifested by the overall fewer choices in favor of one specific option. Nevertheless, what is particularly important to stress here is that a late peak and a late low in the final syllable always suggest the question interpretation, whatever the base utterance is, while a late peak is never interpreted as conveying exclamation, independently of the low pitch height on the final syllable. Interestingly, if the final syllable shows a lower F0 level, that is an earlier low target, the interpretation shifts towards statement in the case of the Exclamation base and towards incredulous question in the case of a Question base utterance. In particular, for the Question base stimuli the difference between incredulous and neutral question choices is statistically significant, though it is not big, suggesting that the boundary specification signals a difference within the question category.

In line with the existence of a strong base effect, which though, does not mask the presence of differences in subject's interpretation, results for series III showed that, when listening to stimuli showing an early low target in the final syllable and a peak which is variably aligned and scaled, subjects appear to interpret stimuli from the exclamation base more as statements when the final low syllable is preceded by a later (in postnuclear) and higher peak, and clearly as exclamations when the peak is earlier and lower. For stimuli from the question base, answers show a higher degree of ambiguity, as manifested by the overall fewer choices in favor of one specific option; however, subjects clearly interpret as exclamations and statements, though with higher effect size values for exclamation, stimuli showing earlier peak positions; they interpret as incredulous questions and neutral questions stimuli showing a late peak alignment.

	Early peak				Late peak			
	series I	*series II*	*series III*	*series IV*	*series I*	*series II*	*series III*	*se-ries IV*
Early Low edge	Excl. (E b.) Stat.(Q b.)	-	Excl. (E, Q b.) *Stat.(Q b.)*	-	-	Quest S (Q b.) Stat. (E b.)	Quest Quest S(Q b.) Stat.(E,Q b.)	-
Low/ not high Posttonic	-	-	-	Excl.(E, Q b.) *Stat. (Q b.)* / QuestS (E, Q b.)	-	-	-	-
Late Low edge	-	-	-	-	Quest (Q b.) Quest S (E b.)	Quest (Q,E b.) *Quest S* *(Q b.)*	-	-
High Posttonic	-	-	-	Quest S (E, Q b.).	-	-	-	-

Table 6: Summary of results for the four series. Options that were chosen significantly above chance are shown for each combination of peak and low edge tone position (early/late; for series IV, high/low posttonic is reported, rather that early/late low edge tone) – base utterances mainly associated to reported results are shown in brackets.

Finally, quite clear results are offered by Series IV, where the height of the posttonic syllable was varied and taken as a cue of both edge tone specification and peak alignment; indeed, due to the manipulation performed, a lower posttonic implies the presence of an early peak alignment, that is, due to the experimental design, a peak in the nuclear syllable. Results for this series clearly show that, independently of the base utterance, participants opt for the incredulous question option (rather than the neutral question option) as soon as a high target, earlier and independent of the end of the rise, is introduced in the nuclear syllable and while the posttonic is not very low. Indeed, when the posttonic is low (almost as low as the starting point of the rise) and, of course, the peak is early, participants switch to the exclamation and the statement interpretations, in the case of the Question base stimuli (with a slightly higher effect size for Exclamation than for Statement). For Exclamation base stimuli, they switch towards the exclamation option, and they do that at much earlier stages of manipulation in comparison to what observed for the Question base stimuli (basically as soon as the posttonic is lower than the peak); moreover they do not choose the Statement option at all. Thus, in general, answers show again a strong base effect, which suggests that, apart from the height of the posttonic and the peak alignment and scaling, other features, such as syllable duration or the rise F0 range, may play a very relevant role in affecting subjects perception. Nevertheless, they also show that an early peak pitch accent followed by a high posttonic and a fall, is perceived as an incredulous question and, at least for Question base stimuli, this interpretation is possible even in the case the posttonic is lower than the preceding peak. On the other hand, a very low posttonic excludes any question interpretation.

Interpreting the results for the four Series is a not easy task, first of all because of the strong base effect emerged in various cases, due for instance to differences in syllable duration for stimuli created from different base utterances. Moreover, the manipulation of the final low alignment in series II and III, that all show a slightly retracted low edge tone target, seem to create quite ambiguous stimuli, as shown by the lower number of subject's choices in favor of one specific option. Nevertheless, results indicate that:

1) Question (neutral) is never conveyed by an early peak position, which in the manipulation performed here means that the peak is never in the nuclear syllable for a stimulus to be perceived as a neutral question (see results for all series); moreover, even though stimuli are identified as neutral questions mainly in the case of a late low edge tone alignment, in very few cases they are still identified as such when the low edge was slightly earlier, though always in the case of a very late peak position (see series III). These results are in line with the labeling proposed in the literature, that is L*+H HL%.

2) Incredulous questions may be signaled by an early (and lower) peak, with both a later and an earlier final low (thought the F0 level of the low has to be not very low - see results for Series IV). However, the peak does not have necessarily to be early, as a late peak and an earlier low target on the final syllable, as well as a quite low postnuclear syllable, may be enough to convey incredulous question (see respectively results for Series II, where the neutral question interpretation is disfavored, and for Series IV). Thus, the incredulous question interpretation is observed for stimuli showing a greater variability in the early/late peak position, often with the presence of an earlier low edge tone specification. These results are considered consistent with the labelling proposed in the literature, that is L*+H L!H (even though in the present experiment the very final syllable F0 level was not under investigation and we cannot make observations on the !H specification).

3) Statement interpretations are never conveyed by a late low edge target or high posttonic, while they appear to be possible in the case of both early and late peak position. These results would be consistent with a L*+H L- transcription, though cases in which a high edge tone following the L*+H accent have been observed in list intonation (see Figure 4); however it can also be the case that the greater variability in the high target alignment and scaling accounts for the pattern observed in connected speech, involving a H- boundary. Thus, the transcription L*+H L- is consistent with perception results.

4) Exclamation is never conveyed by a late peak position, a late edge tone specification or a high posttonic condition (see all series). In the manipulation performed here, this means that, for subjects to perceive an exclamation, the peak cannot be in the postnuclear syllable and the low edge tone has to be early, to the extent that the posttonic syllable has to be already very low (series IV). This shows that the peak position is more constrained than in the two previous conditions (incredulous questions and statements) as it needs to be aligned (early) in the nuclear syllable. The presence of a following low edge tone phonological specification (L*+H L-) is not enough to account for such difference in the degree of allowed variability. Thus, a specific label (L*+>H) is introduced to highlight the necessary early alignment of the pitch accent peak that opposes the exclamation contour to others.

Thus, what the previous results tell us about the intonation system in Turin Italian? The clear difference observed between neutral question intonation and the other options is taken to be perfectly in line with the phonological analysis proposed in the literature, that is L*+H HL%. This analysis clearly represents the contrasts between this contour and the others. However, the presence of an earlier low edge tone specification which was already suggested in phonological analyses proposed in the literature in the other cases (that is in incredulous questions, exclamations and lists) is not considered enough to account for the results obtained in the perception experiment. In particular, the low edge tone specification is taken to be consistent with the variability observed in list and incredulous question contours, where the peak may indeed be either early or late aligned. Conversely, the more constrained situation observed in exclamations, where the peak has necessarily to be early aligned for a stimulus to be perceived as an exclamation, is taken to correspond to a specific feature of the pitch accent. Thus, a specific transcription, that is L*+>H L%, is proposed for exclamation, to contrast the pattern with the L*+H L% one found in lists and the L*+H L!H% found in incredulous questions. Theoretically, another option would have been to label such pitch accent as L+H*, rather than L*+>H. However such option is not considered as a real one, as L+H* is already exploited for other functions within the system (e.g., for contrastive accents) and the accent found in exclamatives is quite different from those usually labeled as L+H* within the Turin variety.

CONCLUSION

The experiment organized to investigate the perception of exclamation and yes-no questions in Turin Italian showed that the difference between the two is not only conveyed by edge tone specification (e.g., L*+H L% vs. L*+H HL%), but also by pitch accent features relating to the peak (e.g., L*+>H vs. L*+H, that is anticipated vs. not anticipated peak, though scaling differences are also expected). Teasing apart the pitch accent and the edge tone contribution to check whether some pitch accent features are a side effect of differences in edge tone composition or not was a not easy task. Given that some parameters were left unvaried (see footnote 7), the manipulation performed implied the realization of stimuli whose function was apparently problematic to identify, probably because of ambiguous/unusual acoustic correlates. However, in line with the working assumption, the pitch accent and the edge tone contribution was finally identifiable because the L*+H pitch accent and the L edge tone could convey other interpretations, such as incredulity yes-no questions and list statements. Such interpretations were then used to tease the contribution of pitch accent and edge tones apart and, in particular, to show that a differentiation between L*+H and L*+>H is useful as pitch accent alignment and scaling differences affect listener's interpretation.

A key point to be made here is that perception experiments appear to be an important tool for suggesting which are the features contrasting within a system, even when they involve quite complicated designs. In some cases, a similar depth in the analysis could seem to be marginally important, as the main contrasting contours are already identified in the system, for instance by means of edge tones combination. Nevertheless, it is important to identify and highlight contrasting features even if they seem to be redundant in a variety. Indeed, at a careful check they may convey differences in meaning and, moreover, in a different variety they may be overtly pertinent rather than apparently redundant features. The indication is that it is important to code the presence of a contrasting feature in any case, to avoid misunderstandings in the comparison of phonological systems of two such varieties.

The perception experiment described in this paper, then, confirms the importance of perception to point out the relevance of specific features in the phonological coding of intonational events. Moreover, it offers an encouraging example of how perception data can be gathered even in the case of not straightforward comparisons, related to pitch patterns that differ as for both scaling and alignment and that are not directly comparable because they cannot be considered as minimal pairs.

Appendix

The interpretation of Cohen's *d*

Cohen's Standard	Effect Size	Percentile Standing	Percent of Nonoverlap
	2.0	97.7	81.1%
	1.9	97.1	79.4%
	1.8	96.4	77.4%
	1.7	95.5	75.4%
	1.6	94.5	73.1%
	1.5	93.3	70.7%
	1.4	91.9	68.1%
	1.3	90	65.3%
	1.2	88	62.2%
	1.1	86	58.9%
	1.0	84	55.4%
	0.9	82	51.6%
LARGE	0.8	79	47.4%
	0.7	76	43.0%
	0.6	73	38.2%
MEDIUM	0.5	69	33.0%
	0.4	66	27.4%
	0.3	62	21.3%
SMALL	0.2	58	14.7%
	0.1	54	7.7%
	0.0	50	0%

Cohen (1988) hesitantly defined effect sizes as "small, *d* = .2," "medium, *d* = .5," and "large, *d* = .8", stating that "there is a certain risk in inherent in offering conventional operational definitions for those terms for use in power analysis in as diverse a field of inquiry as behavioral science" (p. 25).

Effect sizes can also be thought of as the average percentile standing of the average treated (or experimental) participant relative to the average untreated (or control) participant. An ES of 0.0 indicates that the mean of the treated group is at the 50th percentile of the untreated group. An ES of 0.8 indicates that the mean of the treated group is at the 79th percentile of the untreated group. An effect size of 1.7 indicates that the mean of the treated group is at the 95.5 percentile of the untreated group.

Effect sizes can also be interpreted in terms of the percent of nonoverlap of the treated group's scores with those of the untreated group, see Cohen (1988, pp. 21-23) for descriptions of additional measures of nonoverlap.. An ES of 0.0 indicates that the distribution of scores for the treated group overlaps completely with the distribution of scores for the untreated group, there is 0% of nonoverlap. An ES of 0.8 indicates a nonoverlap of 47.4% in the two distributions. An ES of 1.7 indicates a nonoverlap of 75.4% in the two distributions.

Cohen's *d* values. Table found at http://www.uccs.edu/lbecker/effect-size.html.

REFERENCES

Besana, S. (1999). Towards an analysis of Turinese Italian intonation and theoretical implications for Intonational Phonology. MIT, Cambridge, Mass., *Master Thesis in Linguistics, manuscript*.

Blum-Kulka, S., House, J., Kasper G.. (1989). “Investigating cross-cultural pragmatics: An introductory overview”. In Blum-Kulka, S., House, J., Kasper G. (eds.), *Cross-cultural pragmatics: Requests and apologies*. Norwood, NJ: Ablex, 1-34.

Bruce G. (1977). *Swedish Word Accent in Sentence Perspective*. CWK: Gleerup.

D’Imperio M. (2000). *The role of perception in defining tonal targets and their alignment*. PhD thesis, Ohio State University, USA.

D’Imperio M., Gili Fivela B., Niebhur O. (2010). “Alignment perception of high intonational plateaux in Italian and German”. In: *Proceedings of Speech Prosody 2010*, Special Session on Shape, Scaling, and Alignment Effects in the Production and Perception of F0 Events, 11-14 May, Chicago, 1-4.

Frota S. (1998). *Prosody and Focus in European Portuguese*. PhD thesis, Lisbon.

Frota S. (in press). Surface and Structure: Transcribing Intonation within and across Languages. Presented at the *Workshop Advancing Prosodic Transcription for Spoken Language Science and Technology* and submitted for review.

Frota S. & Prieto P. (in press). *Intonational variation in Romance*, Oxford: Oxford University Press.

Gili Fivela B. (2008). *Intonation in Production and Perception: The Case of Pisa Italian*, Alessandria: dell'Orso.

Gili Fivela B. (2009). "From production to perception and back: An analysis of two pitch accents". In: S. Fuchs et al. (eds.), *Some aspects of speech and the brain*, München: Peter Lang, 363-405.

Gili Fivela B. (2013). "Varietà di italiano e differenze nella percezione degli accenti intonativi". In: A. Romano & M. Spedicato (eds.), Sub voce Sallentinitas: *Studi in onore di G.B. Mancarella*, Lecce: Grifo, 289-303.

Gili Fivela B., Avesani C., Barone M., Bocci G., Crocco C., D'Imperio M., Giordano R., Marotta G., Savino M., Sorianello P. (in press). "Varieties of Italian and their intonational phonology". In: S. Frota & P. Prieto (eds.), *Intonational variation in Romance*, Oxford: Oxford University Press.

Gili Fivela B. & D'Imperio M. (2010). "High peaks versus high plateaux in the identification of two pitch accents in Pisa Italian". In: *Proceedings of Speech Prosody 2010*, Special Session on Shape, Scaling, and Alignment Effects in the Production and Perception of F0 Events, 11-14 May, Chicago, 4-8.

House D. (1990). *Tonal Perception in Speech*. Lund: Lund University Press.

IARI - Interactive Atlas of Romance Intonation (http://prosodia.upf.edu/iari/, last accessed 30 Nov. 2014)

Interlandi, G.M. (2003). "L'intonazione delle interrogative polari in italiano torinese tra varietà di italiano regionale e nuova koiné". Università degli Studi di Pavia, *PhD dissertation, manuscript*.

Interlandi, G.M. (2004). "Il continuum della variazione pragmatico-espressiva nell'intonazione dell'italiano parlato a Torino". In: F. Albano Leoni et al. (eds.), *Il parlato italiano*, Naples: D'Auria (*CD-Rom*).

Interlandi G. & Romano A. (2004). "Le continuum intonatif de l'italien parlé à Turin : résultats d'un test d'identification". *Proc. of MIDL 2004* « Identification des langues et des variétés dialectales par les humains et par les machines » (Paris, 2004), Paris: École Nationale Supérieure des Télécommunications, 157-160.

Liberman M., Prince A. (1977). "On stress and linguistic rhythm". In *Linguistic Inquiry*, 8, 2, 249-336.

Pierrehumbert J. (1980). The phonology and phonetics of English intonation. MIT, *PhD Dissertation*, MIT – published in 1988 by IULC.

Romano A. & Interlandi G. (2002). "Quale intonazione per il torinese?". In: A. Regnicoli (ed.), *La fonetica acustica come strumento di analisi della variazione linguistica in Italia*, Roma: Il Calamo, 117-122.

Vanrell M. (2006). "A scaling contrast in Majorcan Catalan interrogatives". In: *Proceedings of Speech Prosody 2006*, Dresden, Germany, 807-810.