Tonal Alignment in Two Pisa Italian Peak Accents

Barbara Gili Fivela

Università degli Studi di Lecce, Lecce
Scuola Normale Superiore, Pisa
gili@alphalinguistica.sns.it

Abstract

A read speech corpus was collected in order to study the tonal targets alignment of two Pisa Italian peak accents. The production of the two accents was forced by inducing a broad and a contrastive interpretation of the target words inserted in small paragraphs. Results of latency and F0 measurements reveal that the contrastive accent has a low target at a 'fixed' distance after the peak whereas the broad focus accent has a target which is influenced by the postaccentual context. Moreover, both broad and contrastive accents are preceded by a low target at a 'fixed' difference from the peak. The interpretation of this target remains provisional and the resulting analysis is [L+]H* for the broad focus accent and [L+]H*+L for the contrastive accent, both followed by a low edge tone.

1. Introduction

In the autosegmental-metrical theory of intonational phonology, the notions of association and alignment play a major role in the phonetic and phonological analysis of the intonation patterns of various languages. Particularly in works following Pierrehumbert approach (see [1] for an overview), the alignment details of the F0 track preceding and following the starred tone determine whether the turning points represent a tonal target (i.e. do not belong to an interpolation phase among other tones) and the type of target (i.e. whether it is a leading/trailing tone -part of the pitch accent- or it is an edge tone).

A number of works in the literature aim at finding the factors that influence the alignment details of the starred tones, which represent the most stable part of the pitch accent [2], but some works also focus on the differences in the alignment of leading and trailing tones. For instance, [3] shows that onset and rime duration have an influence on the location of leading and trailing tones.

Moreover, it is not always possible to clearly categorize a pitch accent as being mono- or bitonal. As reported in the literature, it is not unusual to find pitch accents which show, for instance, both the leading and the starred tone aligned with the nuclear syllable [4,5,6]; or pitch accents showing none of them aligned with the stress position [7,1]. Even if at least one target appears to be clearly aligned with the nuclear syllable, the status of the other target may not be completely understood. For instance, the discussion upon the differences between H* and L+H* pitch accents in some varieties of English (see discussion in [8]) is well known.

As far as Italian is concerned, it is interesting to notice that various analyses have been given for the pitch accent exploited in narrow focus contexts. Due to the number of varieties of Italian that have been analyzed, it may not be surprising that different labels have been proposed. For instance, Avesani labels the narrow focus pitch accent as H* in Florentine Italian, Grice adopts H*+L for the variety spoken in Palermo - as does Savino for the Bari variety - while D'Imperio labels it as L+H* for Neapolitan Italian [9]. Nevertheless, it interesting to notice that some of the differences in these analyses are due to the observation of a systematic dip in front of the peak aligned with the nuclear syllable (see [4]). In Pisa Italian, something similar to the latter situation is observed: a systematic dip is implemented before the peak. Nevertheless, the accent seems correctly analyzable as involving a trailing tone, i.e. as a H*+L pitch accent.

The aim of this work is to describe an experiment performed in order to study two pitch accents both of which show an high peak aligned with the nuclear syllable and a previous low turning point. The hypothesis is that the two pitch accents differ as for the position of the peak in the nuclear syllable and as for the presence of a trailing tone in one case but not in the other case. Nevertheless, the observation of a previous low turning point at a 'fixed' distance from the peak, in both cases, may cause problems in the categorization of the two accents - at least at a first stage of analysis - as both appear to be potentially characterized by a low leading tone.

Therefore, the experiment described below does not aim to study the factors that determine the differences in tonal alignment, rather it aims to point out the ambiguity connected with some characteristics associated with bitonal pitch accents.

2. Corpus

Small paragraphs were created in order to force the production of the two pitch accents under investigation on specific target words (see table 1). The two pitch accents were produced as a result of a broad and a contrastive interpretation of the relevant sentences.

The target words were made up of sonorant segments and had stress in initial position, as the word boundary may influence the pitch accent implementation itself. The three possible nuclear syllables (all open) are followed by up to three postnuclear syllables, from 'postnuclear'=1 to 'postnuclear'=3. The reason no final stressed words were considered is twofold. First, glottalization is often found in final stressed words ([10]), and may cause problems in extracting the F0 values and therefore difficulties in the measurements; second, tonal crowding may be expected in those cases and may have an influence on the alignment of the tonal targets. Notice that, independently from the number of postnuclear positions, the segmental make up of the first postnuclear syllable is kept as constant as possible for each set.

Table 1: Target words grouped by nuclear syllable type and number of postnuclear syllables

\ Nuclear s. n.postncl. \	LA	NO	BE
1	LAva	NOme	BEvi
2	LAvalo	NOmina	BEvilo
3	LAvaglielo	NOminalo	BEviglielo

As one of the goals of the experiment was to look for the alignment of a possible preceding low target, particular attention was also devoted to controlling for the number of interstress syllables (and the segmental context as well). For this reason, the targets words were preceded by two different contexts: *'proNUNcia di ...'* and *'SILlaba di ...'*, allowing to differentiate the case of two unstressed syllables preceding the target word's stress position, i.e. 'interstress'=2 condition, and the case of three unstressed syllables preceding it, i.e. 'interstress'=3 (see table 2). As the syntactic structure is kept constant, it was not possible to add more inter stress syllables.

The target words were final in a clitic left dislocated constituent which is usually separated from the matrix by a low target that can be analyzed as a low phrase accent [11]. Therefore, any difference among the two pitch accents with regards to the postnuclear target has to be connected with a characteristic of the pitch accents themselves as a prosodic boundary is expected in both cases.

Another factor considered in the construction of the corpus is the position of the target word in the sentence. Results of a pilot study showed that the peak in both the pitch accents was preceded by a low turning point that seemed to move closer to the peak itself when preceded by an higher number of syllables. Although the aim of the present work is not to determine which are the factors influencing target alignment, if a condition shows an influence on both the targets this could be considered as pointing to the need for a homogeneous treatment of those targets (for instance treating both of them as leading tones). Therefore, this factor was considered in the construction of the corpus and a different amount of syllables/words was added before the target words.

Table 2: Carrier sentences in the corpus

Inter stress	Phrase	Carrier sentence
2	Short	La proNUNcia di LAvaglielo non la ricordo mai
3	Short	Una SILlaba di LAvaglielo non la pronuncio mai
2	Long	La difficiLISsima e comPLESsa proNUNcia di LAvaglielo non la ricordo mai
3	Long	Una comPLESsa e difficiLISsima SILlaba di LAvaglielo non la pronuncio mai

In the 'phrase'=short condition, the number of prenuclear syllables from the beginning of the sentence was 5 for 'interstress'=2 and 6 for 'interstress'=3 condition; in the 'phrase'=long condition they were 15 and 16, respectively (see table 2).

The number of carrier sentences was 36 for each pitch accent, i.e. 72 in total.

3. Method

Three Pisa Italian speakers, one male and two females, read the paragraphs four times. The recordings were performed in a sound-treated room in the Linguistic Laboratory of Scuola Normale Superiore, in Pisa.

Three out of four repetitions of the carrier sentences were chosen for further analysis, giving a total of 648 stimuli. They were digitized at 22050 Hz in wave format on an Alpha Station using DecSound - Digital Equipement. The measurements (see below) were performed using ESPS Waves+.

On the basis of wide band spectrograms, the pitch accents context was segmented at the beginning of the nuclear syllable onset (c1) and its nucleus (v1); the end of the nucleus (v2) and the end of the first postnuclear syllable onset, i.e. beginning of the postnuclear nucleus (c2); the target word boundary was also labeled. As for the F0 track, the peak value[1] and the preceding and following low values were measured. The two low elbow positions were automatically inserted[2].

On the basis of the labels, the following measurements were performed:

- Latency (ms) between elbowL, c1 and v1; peak and v2; elbowR, v2, c2 and word boundary;
- Latency (ms) between elbowL, peak and elbowR;
- F0 slope, calculated as relation between the F0 excursion from elbowL to peak and the elbowL-peak latency (LHslope); and between the F0 excursion from elbowR to peak and the peak-elbowR latency (HLslope).

The main results of the measurements are discussed in the following section. Particular attention is devoted to the effect of the 'number of interstress syllables' and 'phrase length' factors on the alignment of elbowL and peak; the effect of 'number of postnuclear syllables' is considered in connection to the timing of elbowR.

4. Results

- Latency (ms) between elbowL, c1 and v1; peak and v2; elbowR, v2, c2 and word boundary;

One way Anovas performed on the data with context (Broad vs. Constrastive) as an independent variable, showed a significant effect on all the latencies. The influence of the other factors were then studied keeping separated the two data sets.

The elbowL-c1 latency mean values show that elbowL is aligned after the nuclear onset in both B (0,025717) and C contexts (0,00194), but later in the segment in the first case – see Figure 1. The one way Anova with number of interstress syllables as an independent variable showed that this was not a significant factor for the elbowL-c1 latency in any data set (for broad $[F_{(1,322)}= 1,314; p>0,001]$; for contrastive $[F_{(1,322)}= 0,184; p>0,001]$); nor was the phrase length (for broad $[F_{(1,322)}= 0,248; p>0,001]$; for contrastive $[F_{(1,322)}= 0; p>0,001]$). Coherently, no effect of either the interstress distance or the phrase length was observed on the elbowL-v1 latency for the two sets of data.

The peak-v2 latency mean values show that the peak is

---

[1] When a sort of plateau was observed, the two edges were measured and the peak position and F0 value were calculated as being half a way between the two edges values.

[2] The program exploited was originally written by M.Beckman and later modified by M.D'Imperio [4].

aligned slightly after the nuclear vowel in the broad pitch accent (0,002339) and within the vowel in the contrastive pitch accent (-0,096331) – see Figure 1. The one way Anova showed no effect of the interstress distance for either the broad [$F_{(1,322)}$= 3,933; p>0,001] or the contrastive accent [$F_{(1,322)}$= 5,807; p>0,001]; no effect of the length of the phrase (for broad [$F_{(1,322)}$= 0,036; p>0,001]; for contrastive [$F_{(1,322)}$= 0,005; p>0,001]) and, similarly, no significant influence of the number of postnuclear syllables was observed (for broad [$F_{(2,321)}$= 0,671; p>0,001]; for contrastive [$F_{(2,321)}$= 1,228; p>0,001]).

The v2-elbowR and c2-elbowR latency mean values show that elbowR is aligned within the postnuclear syllable onset in the contrastive (-0,010379 from c2) and after it in the broad pitch accent(0,133499 from c2) – see Figure 1. The number of postnuclear syllables was a significant factor in the v2-elbowR latency in the broad pitch accent, while it was not significant in the contrastive one (respectively, [$F_{(2,321)}$= 32,335; p<0,001]; for contrastive [$F_{(2,321)}$= 1,266; p>0,001]). The Tukey-Kramer post-Hoc comparison showed a significantly shorter latency in the case of 1 postnuclear syllable. Coherently, the number of postnuclear syllables was a significant factor in the c2-elbowR latency for the broad pitch accent [$F_{(2,321)}$= 28,157; p<0,001], while it was not for the contrastive pitch accent [$F_{(2,321)}$= 1,259; p>0,001].
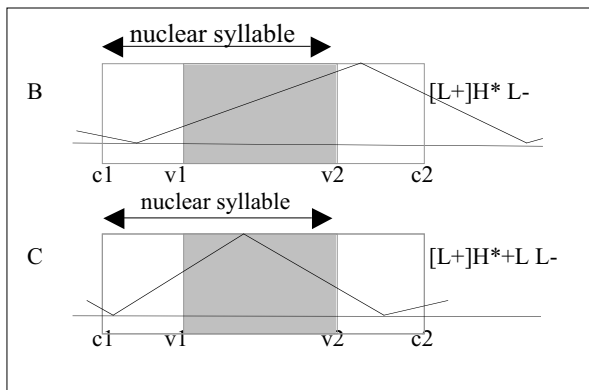

Figure1: Schematic representation of the two pitch accents according to the measurement results

The elbowR-word boundary latency mean values show that elbowR is aligned before the word boundary in both broad (-0,037531 from the word boundary)[1] and contrastive (-0,166989 from the word boundary), although earlier in the last case. The one way Anovas show that the number of postnuclear syllable is a significant factor in both the broad and contrastive pitch accent (respectively, [$F_{(2,321)}$= 325,414; p<0,001] and [$F_{(2,321)}$= 648,028; p<0,001]).

- Latency (ms) between elbowL, peak and elbowR;
The latencies between both elbowL-peak and peak-elbowR are greater for the pitch accent in broad than for the one exploited in context C. A one way Anova was performed on the data with context (B vs. C) as independent variable, showed a significant effect on both the elbowL-peak

---

[1] Looking at the mean values for each class, elbowR appears to be aligned after the word boundary in case of one postnuclear syllable in broad pitch accent.

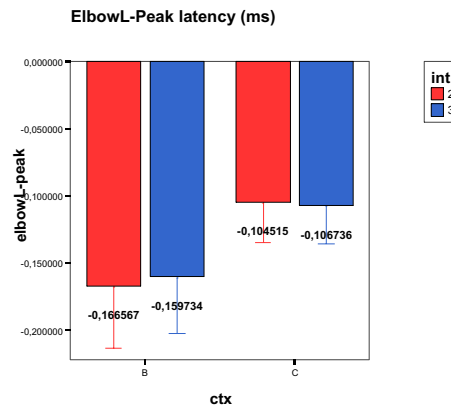[$F_{(1,646)}$= 376,547; p<0,001] and the peak-elbowR [$F_{(1,646)}$= 218,040; p<0,001] latency.

Figure 2: ElbowL-Peak latency mean in B and C contexts for number of interstress syllables(+/-1 SD)

As the measurements are significantly different in these two patterns, the Anovas on the other factors were run on two separated data sets. Neither for broad nor for contrastive was the number of interstress syllables (2 vs. 3) a significant factor for elbowL-peak latency (for broad [$F_{(1,322)}$= 1,906; p>0,001]; for contrastive [$F_{(1,322)}$= 0,465; p>0,001]) – see Figure 2. Similarly, no significant effect of phrase length (short vs. long) was observed (for broad [$F_{(1,322)}$=0,443; p>0,001]; for contrastive [$F_{(1,322)}$=2,346; p>0,001]).

The number of postnuclear syllables (1, 2 or 3) was a significant factor in the peak-elbowR latency in the broad pitch accent, while it was not significant in the contrastive one (respectively, [$F_{(2,321)}$= 29,541; p<0,001] and [$F_{(2,321)}$= 1,547; p>0,001]) – see Figure 3.
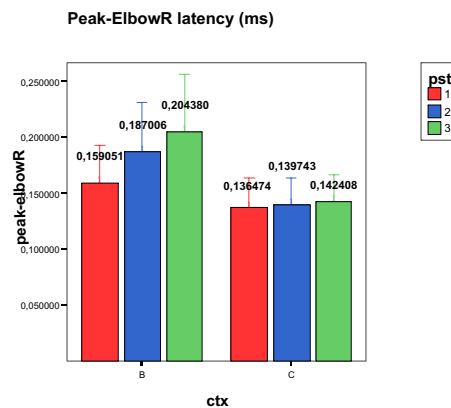

Figure 3: Peak-ElbowR latency mean (ms) in B and C contexts for number of postnuclear syllables (+/-1 SD)

In context B, the peak-elbowR latency appears to increase as the number of postnuclear syllables increases, although the Tukey-Kramer post-Hoc comparison shows no significant difference between two and three postnuclear syllables.

- F0 slope;
One way Anovas performed on the data with context (B vs.

C) as independent variable, showed a significant effect on the slope value. The influence of the other factors were then studied keeping separated the two data sets.

One way Anova performed on the broad and the contrastive data showed that the number of interstress syllables plays no significant role in determining the LHslope (for the B data $[F_{(1,322)}= 0,058; p>0,001]$; for the C data $[F_{(1,322)}= 0,069; p>0,001]$). Similarly, the number of postnuclear syllables is not significant for the HLslope (for the broad data $[F_{(2,321)}= 1,292; p>0,001]$; for the contrastive data $[F_{(2,321)}= 0,707; p>0,001]$.

5. Discussion

The two pitch accents are significantly different in the alignment of all targets. The latency mean values show that the tonal targets are aligned earlier in contrastive than in broad pitch accent.

The number of interstress syllables does not appear to be a significant factor in differentiating the elbowL-peak, the elbowL-c1 or the elbowL-v1 latency time for the two pitch accents (the same holds true for the phrase length factor). Therefore, a low target whose timing appears to be quite stable precedes the high peaks in both pitch accents under investigation. This would point to an analysis such as L+H* for both the contexts. On the other hand, the number of postnuclear syllables appears to play a significant role for the peak-elbowR, the v2-elbowR and the c2-elbowR latency in the broad focus pitch accent. The target does not appear to belong to the pitch accent in this case, rather it may be due to the topic boundary. In the contrastive pitch accent, the number of postnuclear syllable does not appear to influence the peak-elbowR latency, pointing to the need for including the low target in the pitch accent analysis. Therefore, this would point to an analysis such as L+H* L- for broad pitch accent and L+H*+L L- for the contrastive pitch accent.

Interestingly enough, the F0 slope data do not show an influence of the factors considered in any case. Further analysis of the F0 measurements in connection to the targets considered is required in order to fully understand the reason of such a result.

Therefore, on the basis of the measurements, either a tritonal analysis is embraced - because of the phonetic details - or the first low target could be considered a structural property of peak accents in Pisa Italian. In the first case, the well known problem of overgeneration must be faced; in the latter, the argument of tonal stability of targets belonging to the same pitch accent would be weakened.

Different conclusions have been reached in the literature on the alignment of leading tones [12,13]. Moreover some approaches in the autosegmental-metrical theory do not even explicitly label the leading tone [1]. The best option seems to be to leave open the question as to whether the low target preceding the peak is a structural property of the peak accent, leaving examination of different intonation contours to clarify this question [14]. The analysis proposed is thus [L+]H* L- for the broad interpretation and [L+]H*+L L- for the contrastive.

6. Conclusions

Two Pisa Italian peak accents were studied with particular attention to the alignment of their targets. Results of latency and F0 measurements show that the two patterns have significantly different values for all targets. Among the distinctive features is the presence of a low target at a 'fixed'

distance after the peak which characterizes the contrastive accent. This second low target belongs to the pitch accent used for contrast while it is independent of the broad focus accent.

However, in both the contrastive and the broad accent, the high tone is preceded by a low target that does not show differences in timing due to the modification of the preaccentual context. This result is not considered strong enough to force the adoption of a tritonal analysis. Nevertheless, further examination is needed before considering the preceding low target a structural property of both the peak accents in Pisa Italian.

7. References

[1] Ladd, D.R., 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

[2] Frota, S., 2000. *Prosody and Focus in European Portuguese. Phonological Phrasing and Intonation*. New York: Garland.

[3] van Santen, J.P.H.; Hirschberg, J., 1994. Segmental effects on timing and height of pitch contours. In *Proc. of the ICSLP'94*. Japan: Yokohama, vol.2, 719-722.

[4] D'Imperio, M.P., 1999. Tonal structure and pitch targets in Italian focus constituents. In *Proceedings of ICPhS'99*.USA:S.Francisco.

[5] Besana, S., 1999; *Towards an analysis of Turinese Italian intonation and theoretical implications for intonational phonology*. PhD dissertation, MIT, USA.

[6] Marotta, G., 2000. Allineamento e trascrizione dei toni accentuali complessi: una proposta. In *Atti del GFS*, Napoli, Italy.

[7] Arvaniti, A.; Ladd, R.; Mennen, I., 2000. What is a starred tone? Evidence from Greek. In *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, Broe, M.; Pierrehumbert J.B. (eds.). Cambridge: Cambridge University Press.

[8] Beckman, M., Ayers, G.M, 1994. Guidelines for ToBI labelling.http://ling.ohio-state.edu/Phonetics/EtoBI

[9] Grice, M.; D'Imperio, M.; Savino, E.; Avesani, C. (in press). Towards a strategy for labelling varieties of Italian. In Jun, S.-A. (ed.) *Prosodic typology*. OUP.

[10] Vayra, M., 1992. Phonetic explanations in phonology: laryngealization as the case for glottal stops in Italian word-final stressed syllables. In *Proc. 7th International Phonology Meeting*. Austria: Krems, 275-294.

[11] Gili Fivela, B., 1999. The prosody of left-dislocated topicalized constituents in Italian read speech. In *Proceedings of Eurospeech'99*, vol.1, 531-534.

[12] Grice, M., 1995. Leading tones and downstep in English. *Phonology*, 12, 183-233.

[13] Pierrehumbert, J.; Beckman, M., 1988. *Japanese Tone Structure*. Cambridge: MIT Press.

[14] Grice, M.; Savino, M., 1995. Low tone versus 'sag' in Bari Italian intonation; a perceptual experiment, *Proc. XIII International Congress of Phonetic Sciences*, Stockholm.