

High peaks versus high plateaux in the identification of two pitch accents in Pisa Italian

Barbara Gili Fivela, Mariapaola D'Imperio*

Università del Salento & CRIL – Lecce, Italy

* Université Aix-Marseille I et Laboratoire Parole et Langage, CNRS, Aix-en-Provence, France

barbara.gili@unisalento.it; mariapaola.dimperio@lpl-aix.fr

Abstract

The role of pitch pattern shape and perceptual target location is investigated here by means of an identification test involving two pitch accent categories in Pisa Italian. The perception test concerned the pitch accents used in contrastive and utterance-initial broad focus and was performed by asking subjects to identify two continua composed of peak and plateau stimuli whose alignment and scaling characteristics were manipulated. In line with previous findings, results show that alignment and scaling both play a role in accent identification and that pitch shape affects subject perception. In particular, plateau stimuli are perceived as late peak stimuli having the same fundamental frequency height or else they are perceived as early peak stimuli realized at a higher frequency.

Index Terms: tonal alignment, scaling, pitch perception.

1. Introduction

In a seminal work, Bruce [1] showed that the difference between two lexically governed pitch accents in Swedish was captured by observing the fundamental frequency (F0) level reached at a specific moment in time, with respect to the stressed syllable. Since Bruce's work, intonologists have relied on pitch targets and their alignment in order to distinguish tonal events within intonational contours. The whole autosegmental-metrical framework [14,11] is based on the idea that, whatever the segmental composition of the sentence, the pitch contour corresponds to a sequence of tonal events realized as pitch targets, connected by interpolation lines. The focus on pitch targets, their alignment and scaling characteristics is clear by a quick inspection of the literature: alignment and scaling of pitch targets have been investigated in the acoustic, the perceptual and the articulatory domain (among the others, [15,3,4]), and attention has also been paid to their interplay [12].

Nevertheless, some works over the past years, and more recent work, pointed out that pitch pattern shape, apart from tone target alignment, may play an important role both in the production and the perception of intonational events: in production, different acoustic shapes seem to be related to different pragmatic contexts; in perception, pattern shapes have been found to differently affect subject perception in a complex way. In Neapolitan Italian, for instance, two rising pitch accents expressing two different pragmatic contexts (narrow focus question and partial topic statement) have been found to be acoustically characterized by a different (concave vs convex) rising phase [2]. On the other hand, the perception of H* in northern standard German has been shown to be due to the interplay of various factors, among which the slope of F0 movements [12]; to make another

example, variations in the rise or fall shape in pitch accents has been shown to contribute to the question-statement distinction in German [13].

Among pattern shape variability, the one differentiating peaks and plateaux has also been investigated. In particular, it has been shown that in Neapolitan Italian scores for plateaux in a question vs statement identification task are similar to scores for late peak stimuli [3]. Moreover, in English, plateaux are reported to be perceived as higher in pitch and more prominent in comparison with peak stimuli having the same F0 value and being aligned at the fall [10].

In this paper the perception of peaks and plateaux is investigated in Pisa Italian, in which the scaling of peaks has been found to interact with their alignment in the identification of two different pitch accents. Pisa speakers' perception is an interesting test for the results in [3,10].

2. Previous investigations on Pisa Italian

In Pisa Italian two pitch accents showing a rise to a peak are found, phonetically described as [L+]H* and [L+]H*+L. The former accent can be found in broad focus, sentence initial position - both prenuclear and nuclear - and in narrow focus; the latter is usually found in contrastive focus, in either "checks" or questions characterized by disbelief. Detailed measurements performed on the two accents produced phrase finally in utterance initial position, revealed their average alignment and scaling characteristics [6,8]. Results show that a) a low target, representing the starting point of the rise, is aligned at syllable onset in both pitch patterns (with a difference of few ms); b) the peak is aligned significantly earlier in the [L+]H*+L than in the [L+]H* accent and c) it is lower in [L+]H*+L than in [L+]H* (fig.1).

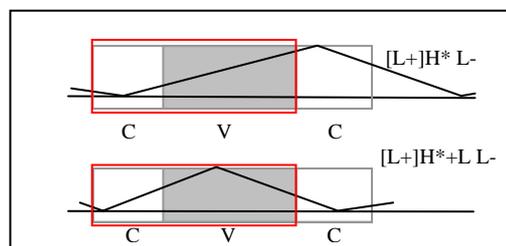


Figure 1: Representation of the two pitch accents.

A perception test was later performed [7,8] in order to find out whether the two accents are identified as belonging to different categories and to tease apart the contribution of alignment and scaling to their identification. Two series of stimuli were created by manipulating the alignment of both the peak and the following low (i.e., changing the rise slope), and the scaling of all the target tones.

An identification test was carried out by asking subjects to listen to stimuli that belonged to a continuum of gradual changes between the two patterns and to assign them to one of two pitch accent categories, contrastive and utterance-initial broad focus, on the basis of the stimulus interpretation. Results show that the F0 height of tonal targets affects subject judgements, playing a role in terms of 'at what step' and 'how much' a specific alignment pattern is identified as being a member of a given category (see [8] for discussion).

The alignment and scaling interplay in the perception of these two pitch accents called for a check on the perception of peak and plateau stimuli. Indeed, as already mentioned, scores for plateaux appear to be similar to scores for late peak stimuli [3] and for higher peak stimuli, in that plateaux may be perceived as higher in pitch than peak stimuli aligned at their fall, but having the same F0 value [10].

3. Corpus and method

A corpus of stimuli was created (see also [9]), starting from the same base utterance employed for previous investigations on Pisa Italian, that is, *ho detto velava velocemente* 'I said veiled quickly', produced by a female speaker of Pisa Italian with a contrastive accent on *velava* 'veiled' [7,8]. The utterance was used to generate two series of stimuli corresponding to a gradual shift from [L+]H*+L to [L+]H* L-: in the first series the pitch accents corresponded to high peaks (see § 3.1) while in the second series they corresponded to high plateaux (see § 3.2). Stimuli were generated in PRAAT with PSOLA resynthesis.

An identification test was performed asking 14 Pisa Italian subjects to listen to 3 repetitions of the stimuli and to answer the following question: does the stimulus correct a preceding utterance in a peremptory and conclusive way? (for discussion see [7,8]). The hypothesis was that plateaux would score similarly to late and higher peaks, showing a category shift.

3.1. Peak-configuration series

In the first series, the following manipulations were made:

a. Alignment manipulation (steps '0', '1', '2', etc.): the low-high-low targets were forward shifted in 8 steps of 15 ms each (rising/falling slopes were kept unvaried), fig. 2, left;

b. Scaling manipulation (steps 'Pr0', 'Pr1', 'Pr2'): for each alignment step, all the targets were shifted upwards on the Hz scale in 2 steps (13, 17, and 6 Hz, respectively), fig. 2, right;

c. Moreover, two stimuli were inserted in order to check whether the slope of the rise made any difference (stimuli 'peak&slope'): the stimulus at the 'Pr2' scaling had peak alignment and scaling typical of [L+]H* L- (i.e. it corresponded to the final step of both alignment and scaling manipulations), and showed the original rising slope - see fig. 1, upper. The other stimulus showed peak scaling and rise slope of the [L+]H*+L pattern - fig. 1, lower - which is actually the same throughout the series.

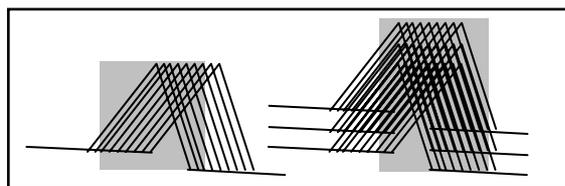


Figure 2: Scheme of manipulations in the peak series.

3.2. Plateau-configuration series

In the second series, a plateau of 45 ms was created so that the duration of the H target corresponded to three steps of the alignment manipulation in the peak-configurations series. Alignment and scaling manipulations were then carried out.

a. Alignment manipulation (steps '0&3', '1&4', etc.): the low-high-high-low targets were forward shifted in 6 steps of 15 ms each (rising and falling slopes were kept unvaried) in such a way that the rise and the fall of a plateau stimulus were aligned to the rise and the fall of two different peak stimuli: the plateau rise was matching a peak stimulus rise and its fall corresponded to the fall of the peak stimulus that was three steps away (e.g., in step '1&4' the rise matched that of step '1' in the peak-configuration and the fall matched that of step '4'); moreover, an extra step (step '-3&0') was added in order to match its fall with that of step '0' in the peak-configuration series, that is, the rise was anticipated of 45 ms - see fig. 3, left;

b. Scaling manipulation (steps 'Pr0', 'Pr1', 'Pr2'): exactly as in the other series, for each alignment step all targets were shifted upwards on the Hz scale in 2 steps (13, 17, and 6 Hz, respectively) - see fig. 3, right;

c. To further verify that plateaux would be perceived as late peak stimuli, a further step was added: for each scaling condition, a long plateau was created whose rise corresponded to the rise in step '0' and whose fall corresponded to that in step '8' (the high stretch was spanning all the peak positions, having a total duration of 120 ms - see step 'long-0&8').

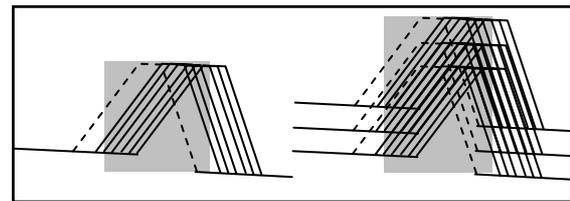


Figure 3: Scheme of manipulations in the plateau series.

4. Results

The mean of positive answers in favour of the contrastive interpretation is given in figs. 4 and 5, with respect to align (x axis) and scaling steps of manipulation (colours/forms).

Consistently with previous results, the F0 height of tonal targets affected subject judgements in both the peak and the plateau-configuration series. Specifically, in both cases the scaling value of the contrastive configuration (see 'Pr0-C' values in the figures) interferes with the perception of the competing category even for the very final steps of the alignment manipulation (step '8' in the peak-configuration and '5&8' in the plateau-configuration series - figs. 4,5). Nevertheless, results for peak and plateau stimuli cannot be equated.

4.1. Peak-configuration series

Two-ways ANOVAs performed on the data, with alignment and scaling steps as main factors, showed a significant effect of both factors and no interaction ($[F(351,8)=20.633; p<0.0001]$ for alignment, $[F(351,2)=38.090; p<0.0001]$ for scaling, $[F(351,16)=1.136; p>0.05]$ for the interaction). A Fisher's post-hoc test showed that step '0' and '1' in the alignment manipulation correspond to different answers in comparison to the other stimuli; moreover, answers for step '2' were different from those for steps from '4' to '8'. As for scaling, every step gives scores different from the others.

Thus, subjects gave different answers depending on scaling, but basically after step '4' in alignment, scores in favour of the contrastive interpretation were lower.

No significant difference was found between scores for peak stimuli with unvaried slope and those with varied slope (for 'Pr0' condition, $[F(26,1)=1.333; p>0.05]$; for 'Pr2', $[F(26,1)=0.324; p>0.05]$). This is particularly interesting for stimuli belonging to the 'Pr2' scaling condition, as their shape was drastically different (see 'Pr2, peak-&-slope' in fig. 4).

Thus, scaling information interacts with alignment in the identification of the two accents, that is, the steps of alignment that (tend to) correspond to one category or the other vary depending on the scaling condition (i.e. chance level is crossed at step '3' in condition 'Pr2' and step '4' in condition 'Pr1'; for 'Pr0', answers for stimulus '7' correspond to the chance level, which is basically never crossed).

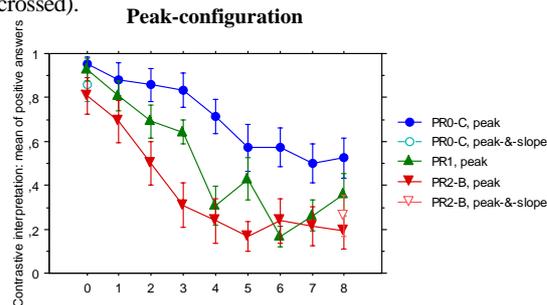


Figure 4: Mean of contrastive scores: alignment (x axis) and scaling steps (colors/forms).

Data show that the (low) scaling characteristics of the contrastive accent lower the identification rate of the other pattern even when the alignment is late in the syllable (the score is around chance level for steps '7,8' in the 'Pr0,peak' condition); however high scaling did not have a similar effect (see steps '0,1' in the 'Pr2,peak' condition). Interestingly, when rise slope was varied [7,8] high scaling interfered with the identification of the contrastive accents (the score was around chance level for steps '0,1' at 'Pr2' scaling).

4.2. Plateau-configuration series

Two-ways ANOVAs performed on the data with alignment and scaling steps as main factors were significant, while there was no interaction ($[F(273,6)=17.152; p<0.0001]$ for alignment, $[F(273,2)=90.496; p<0.0001]$ for scaling, $[F(273,12)=1.703; p>0.05]$ for the interaction). A Fisher's post-hoc test showed that step '-3&0' in the alignment manipulation elicited answers different from those given to other stimuli; moreover, answers for '0&3' and '1&4' were similar to each other and those for the latter were also similar to answers given to '2&5'. As for scaling, every step yielded scores different from the others. Thus subjects gave different answers depending stimuli scaling, but basically after step '1&4', or '2&5' at the most (basically in condition 'Pr0'), stimuli corresponded to significantly lower contrastive scores.

ANOVAs and post-hoc performed including the long-plateau stimuli gave similar results ($[F(310,7)=14.393; p<0.0001]$ for alignment, $[F(310,2)=90.799; p<0.0001]$ for scaling, $[F(310,14)=1.41; p>0.05]$ and for their interaction). Specifically, scores for long-plateaux (stimuli 'long-0&8') corresponded to scores for stimuli starting from alignment step '1&4'. Indeed, long-plateaux yielded responses similar to late aligned plateaux, which were more similar to scores for plateau '5&8' than for plateau '0&3'.

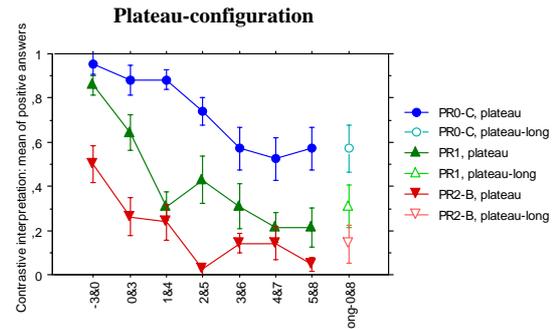


Figure 5: Mean of contrastive scores: alignment (x axis) and scaling steps (colors/forms).

However, in the plateau-configuration series, similarly to what was observed in the other series, scaling information interacts with alignment in the identification of one category or the other. However, the steps of alignment that (tend to) correspond to one category or the other vary depending on scaling (i.e. the second step - '0&3' is above chance level for condition 'Pr2', which is the first step involving a plateau extending to the right of the starting peak position; chance level is crossed at step '1&4' for condition 'Pr1', which is actually the only scaling allowing a clear identification of two different categories; in condition 'Pr0', answers for step '3&6' and later steps are around chance level).

Again, the scaling characteristics of one pattern interfere with identification of the other one. Data show that the scaling characteristics of the contrastive pattern lower the identification rate of the competitive pattern even when the alignment is late in the syllable (the score is around chance level for steps '3&6' and later steps in condition 'Pr0, plateau'); in the 'Pr2' condition higher scaling interferes with the identification of the contrastive accent from the very first step (similarly to what was observed by varying rise slope [7,8]).

4.3. Peak- vs Plateau-configuration series

By simply comparing the plots for peak- and plateau-configuration results, the correspondence between answers given to stimuli sharing the same fall contour is clear. In fact, stimuli in condition 'Pr2' were identified as non-contrastive from step '3' and step '0&3' for the two series, respectively; for condition 'Pr1' the same holds true for stimuli '4' and '1&4'; for condition 'Pr0', the correspondence is not ideal, although there is a similar tendency (chance level is reached around step '7' and '3&6' in the two series respectively).

These observations, which are in line with previous findings on Neapolitan Italian [3], have been confirmed by statistical analysis, performed by matching plateau stimuli with peak-configuration stimuli as for either rise or fall alignment (e.g., stimulus '0&3' in the plateau series matched stimulus '0' in the peak-configuration series for the rising phase analysis, while it matched stimulus '3' for the falling phase analysis). ANOVAs performed on the data with alignment, scaling steps and peak/plateau-configuration as factors showed the expected influence of alignment and scaling factors (in the rising phase analysis: $[F(468,5)=24.585; p<0.0001]$ for alignment, $[F(468,2)=97.593; p<0.0001]$ for scaling; in the falling phase analysis: $[F(468,5)=10.1744; p<0.0001]$ for alignment, $[F(468,2)=102.314; p<0.0001]$ for scaling). Moreover they showed that the peak/plateau-configuration factor plays a significant role in differentiating only the stimuli matched as for the rising phase. On the contrary, when stimuli were matched for the falling phase no significant difference was found between

the peak and the plateau series (rising phase analysis: $[F(468,1)=77.799; p<0.0001]$; falling phase analysis: $[F(468,1)=0.38; p>0.05]$). The same was true when long-plateaux were compared with both peaks and plateaux, matched for their rising or for their falling phase (rising phase analysis: $[F(115,2)=36.301; p<0.0001]$; falling phase analysis $[F(115,2)=0.673; p>0.05]$). In this case a Fisher's post-hoc showed a significant difference between each configuration, with long-plateaux obtaining the lowest score and peak-configuration stimuli obtaining the highest.

As far as scaling is concerned, this factor interferes with the perception of two distinct categories in the case of 'Pr0' scaling (lower pitch) in both series; in the 'Pr2' condition (higher pitch) the interference is mainly found in the plateau condition. Possibly, given that plateaux are perceived as later peaks (see above), the scaling effect adds to that of alignment and is therefore stronger. Apparently, plateaux and stimuli with varying slope (that is, stimuli that show a longer portion of high or rising F0 throughout the accented syllable – see [7,8]) yield greater sensitivity to scaling, as if subjects were facilitated in perceiving that correlate. Two-way ANOVAs were performed on stimuli matched, again, for either the falling or the rising phase, considering plateau and peak stimuli in the scaling conditions as if they were completely different stimuli. Results show that the peak/plateau-scaling factor plays a significant role when considering both stimuli matched for the rise and stimuli matched for the fall ($F(540,5)=40.953; p<0.0001$, $F(498,5)=38.195; p<0.0001$ respectively). However, a Fisher's post-hoc test showed that stimuli matched for the falling phase differ from each other, apart from stimuli belonging to the same scaling condition (as already shown by the results above). On the other hand, stimuli matched as for their rising phase differ from each other except from peak and plateau stimuli in the lower scaling condition ('Pr0-Peak' and 'Pr0-Plateau') and, more importantly, except from peak stimuli at high scaling condition and plateau stimuli at low scaling condition (that is, 'Pr1-Peak' and 'Pr0-Plateau', as well as 'Pr2-Peak' and 'Pr1-Plateau' correspond to identification scores that are not significantly different). Thus results resemble those in [10] only apparently, in that here plateaux are perceived as higher in pitch than peak stimuli having the same F0 value and being aligned for their rise (rather than for their fall, which was the case in [10]).

The effect is even clearer when results for long-plateau stimuli are considered (and compared with those for rises at step '0' and falls at step '8'). Consistently, results show that the peak/plateau-scaling factor plays a significant role when considering both stimuli matched for the rising and stimuli matched for the falling phase ($F(115,8)=14.689; p<0.0001$, $F(115,8)=4.826; p<0.0001$ respectively). However, as far as the relation between long-plateau and plateau is concerned (no systematic detail on the difference with peak stimuli is given here), a Fisher's post-hoc test showed similarities to what was previously observed as for the relation between plateaux and peaks. In particular, for stimuli matched for the rising phase long-plateaux at scaling 'Pr0' are perceived as plateaux at scaling 'Pr1' (but differently from peak stimuli at scaling 'Pr1', that show a higher score for the contrastive interpretation), and long-plateaux in condition 'Pr1' gave similar results to plateaux in condition 'Pr2'. In stimuli matched for their falling phase, no similar effect was found, since long-plateau stimuli were perceived as different from plateau stimuli (with the exception of long-plateaux and plateaux both belonging to condition 'Pr1', and long-plateaux in condition 'Pr2' and plateau at both 'Pr1' and 'Pr2' conditions).

5. Discussion and conclusions

Alignment and scaling seem to interact in the identification of both peak and plateau stimuli. However, the two stimuli series of this study did not yield exactly the same results. In fact, in line with [3], plateau stimuli appear to be perceived as late peak stimuli (mostly those matching as for their falling phase), scaled at the same height. On the other hand, when different H scaling conditions are taken into account, plateau stimuli appear to be perceived as early peak stimuli having the same rising phase but higher peak scaling. Thus, these results only apparently resemble those in [10]: here, plateaux are perceived as higher in pitch than peak stimuli being aligned for their rising phase, rather than for their falling phase.

Apart from the direct comparison of peak and plateau stimuli, the lower scaling (which is typical of a contrastive interpretation) interferes with the perception of two distinct categories in both series, while the higher scaling mainly interferes in the plateau-configuration series. Apparently plateaux and stimuli with varying slope (see [7,8]), that is, stimuli that show a longer portion of high or rising F0 throughout the accented syllable, allow a more precise perception of scaling characteristics.

These results clearly confirm the interplay of scaling and alignment in perception and the need for carefully comparing pitch patterns with different shapes (see also [5]).

5. References

- [1] Bruce, G. "Swedish word accents in sentence perspective", CWK, Gleerup, 1977.
- [2] Cangemi, F. "Phonetic detail in intonation contour dynamics", in Proc.V AISV Conf., Zurich, Feb.09, in print.
- [3] D'Imperio, M. "The role of perception in defining tonal targets and their alignment", PhD thesis, OSU, 2000.
- [4] D'Imperio, M. "Language-specific and universal constraints on tonal alignment: The nature of targets and "anchors", proc.Speech Prosody, Aix-en-Pr., 101-106, 2002.
- [5] D'Imperio, M., Gili Fivela, B. and Niebuhr O. "Alignment perception of high intonational plateaux in Italian and German", Speech Prosody 2010, this volume.
- [6] Gili Fivela, B. "Tonal alignment in two Pisa Italian peak accents", proc.Speech Prosody, Aix-en-Pr., 339-342, 2002.
- [7] Gili Fivela, B. "La percezione degli accenti: il ruolo dell'allineamento e dello 'scaling' dei bersagli tonali", in Proc. I AISV Conf., Padua, Dec.2004, 313-326, 2005.
- [8] Gili Fivela, B. "Intonation in Production and Perception: The Case of Pisa Italian", Edizioni dell'Orso, Alessandria, 2008.
- [9] Gili Fivela B. and D'Imperio M. "High peak vs high plateau in the identification of peak accents in Italian", poster presented at TIE Conference, Lisbon, Sept. 2008.
- [10] Knight, R. "The shape of nuclear falls and their effect on the perception of pitch and prominence: peaks vs. plateaux", Language and Speech, 51, 3, 223-244, 2008.
- [11] Ladd, D. R. "Intonational Phonology". CUP, 1996/2009.
- [12] Niebuhr, O. "The signalling of German rising-falling intonation categories - Interplay of synchronization, shape, and height". *Phonetica* 64, 174-193, 2007.
- [13] Petrone, C. and Niebuhr, O. "The role of the prenuclear F0 region in the identification of German questions and statements". Paper presented at the PAPI conf., 2008.
- [14] Pierrehubert, J. "The phonology and phonetics of English intonation", PhD thesis, MIT 1980 (pub.1988 by IULC).
- [15] Silverman, K. and Pierrehubert, J. "The timing of prenuclear high accents in English", in J.Kingston, M.Beckman (eds.), Papers in Laboratory Phonology I, CUP, Cambridge, 72-106, 1990.