Alignment perception of high intonational plateaux in Italian and German

Mariapaola D'Imperio, Barbara Gili Fivela**, and Oliver Niebuhr*[+]

[*] Université Aix-Marseille I et Laboratoire Parole et Langage, CNRS, Aix-en-Provence, France
**Centro di Ricerca Interdisciplinare sul Linguaggio & University of Lecce, Italy
[+] Department of General & Comparative Linguistics, University of Kiel, Germany
mariapaola.dimperio@lpl-aix.fr; barbara.gili@unisalento.it; niebuhr@linguistik.uni-kiel.de

Abstract

This paper addresses the issue of tonal perception as it relates to special configurations, i.e. fundamental frequency (F0) plateaux. We here review a series of perceptual experiments in two different languages, Italian (Naples and Pisa variety) and German. A subset of the auditory stimuli employed in these studies contained a high F0 plateau, which had to be either identified for a specific tonal category or matched to a previous context. The results show a tendency, for all languages, to match a pitch accent category having a late H peak target to plateau stimuli, which might be due to a universal auditory integration mechanism. This has consequences for intonation models, since the relationship between dynamic characteristics of accentual contours and tonal target location is complex and not always immediately identifiable with turning points.

Index Terms: tonal alignment, pitch perception, pitch accent category.

1. Introduction

Autosegmental-metrical theories of intonation are based on the assumption that intonation contours can be derived from the interpolation of a relatively small number of "target tones", which are more or less strictly identifiable with F0 peaks and valleys (or *elbows*) in the contour itself. By postulating the existence of such targets, whose nature is both static (in the sense of being identifiable with a simple temporal value) and acoustic (in the sense of being identifiable with a particular F0 value), a new area of intonation studies has developed in the last 20 years, where the notion of tonal association in intonation languages is confronted with the notion of *tonal alignment* (see [1]).

Starting with the work of Ladd et al. [2], it has been often reported that when right-hand prosodic effects are excluded (e.g., when the target tone under investigation is not in the vicinity of a pitch accent or a boundary tone), *the alignment of f_0 peak targets is quite systematic, and possibly governed by "segmental anchoring" principles* [2]. In spite of the volume of scientific activity and publications dedicated to the problems of tonal alignment in the last two decades, we note all the same that certain important issues are still in search of definitive answers. Amongst other things, we do not know if a synchronization between melody and text should be modelled in relation to either acoustic, articulatory, or auditory principles; and secondly, if alignment principles are universal or rather language-specific.

Though the idea of "strict" segmental anchoring is controversial, since syllable structure and segmental effects on alignment have been found for various languages [3,4,5] still most of current alignment studies implicitly assume some sort of perceptual equivalence, for a specific pitch accent category, between accentual contours showing comparable initial and final f_0 target alignments, for both rises and falls.

It is a well known fact that in many languages displacing (High or/and Low) target alignment can induce a perceptual shift between pitch accent categories signalling pragmatic contrast, such as question vs. statement, or even discourse structure (such as *shifting topic* vs. already mentioned topic). Most of these studies have employed resynthesized auditory stimuli in different languages (see [6] for English, [7] for German, [8] for Italian). More recently, a few studies have also showed that the dynamic characteristics of rises and falls showing the same alignment of their f_0 extrema can affect the perception of a tonal category [3,9,6,10]. Here we will concentrate on results of experiments in which the duration of the H target in rising-falling configurations was manipulated so that a high plateau would obtain.

In languages such as English [11], it appears that plateaux are employed in place of high peaks in certain contexts, and that they are perceived as possessing a higher pitch value than peak stimuli having the same F0 value. This regularity might be related to an auditory integration property that might be universally valid. A question may then arise as to the effect of plateau vs. peak stimuli in determining temporal alignment identification for specific pitch accent categories.

2. Experimental Results

2.1. Italian

In the two experiments presented here, the alignment of either a rise-fall peak configuration or of a plateau configuration was shifted in time in order to cross a category boundary (Fig. 1). In Neapolitan Italian, the relevant shift is between an early peak, L+H*, employed in narrow focus statements, and a late peak accent, L*+H, which is typical of yes/no questions. In Pisa Italian, the shift is between an early peak H*+L used mostly for contrastive focus and a late peak H* employed, for instance, mostly for utterance-initial broad focus. The central working hypothesis in both studies [3,12] was that the shape of the accent peak would modify the coordinates of the perceived targets and, consequently, the identification of the pragmatic contrast. Specifically, we hypothesized that plateau configurations would favour later peak identification for both languages.

2.1.1. Neapolitan Italian

After determining the range of variability of alignment and scaling, the alignment of the three targets of a L*+H rise-fall (L1, H and L2) was modified through resynthesis of an

original question. In the present study [3], the F0 contour of a question, uttered by a female speaker of Neapolitan Italian, was stylized and resynthesized by means of PSOLA. Two sets of stimuli were created, a peak and and a plateau series (see Fig. 1) in which the tonal alignment of the rise-fall was varied, while F0 height was kept constant. Hence, apart from the temporal alignment, an additional variable was the shape of the accent peak, which could be either flat (creating a short plateau) or sharp. Note that alignment was shifted only within a limited section of the stressed syllable, so that peak locations would all be included within a critical region for question and statement peaks (i.e., starting from 60 ms into the stressed vowel). This way, the creation of stimuli whose alignment could be mapped onto an additional and unrelated pitch accent category, such as H+L*, was avoided, as well as displacing the percept of accent peaks on the syllable preceding the accented one. Thirty Neapolitan subjects listened to the stimuli and identified each as a question or a statement.

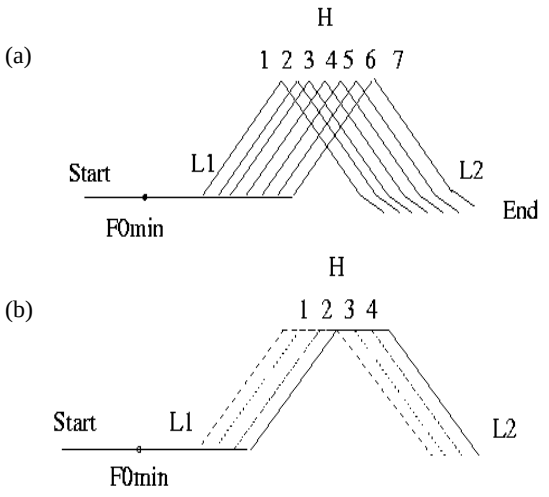

Figure 1: *Schematic diagram of the alignment manipulation for the peak series (upper) and the plateau series (lower) in Neapolitan. From [5].*

First, we tested the hypothesis 1 that when the LHL configuration is moved backwards within the stressed vowel, a higher percentage of statement responses should be obtained. Then we tested the hypothesis 2 that the number of question responses will be the same for plateau stimuli and peak stimuli whose peak is timed at plateau onset or offset. In other words, we tested if the "target" of the rising transition of plateau stimuli is timed to occur at the end of the LH rise (which was implicitly assumed in [8]) or at the beginning of the HL fall. The hypotheses were tested by comparing question scores for peak stimuli with scores for peak stimuli timed at plateau onset or offset.

To test hypothesis 2, we created a "plateau" continuum, which consisted of a series of 4 stimuli whose shape was characterized by a 45 ms F0 plateau (Fig. 1, lower). Plateau duration was therefore equal to three steps of the primary continuum. Figure 2 (left) shows mean question scores for peak stimuli pooled for all listeners (y axis) across stimulus Time Step (x axis). As expected from the results for the "inter-peak" continuum reported in [8], shifting the L1-H-L2 configuration backwards within the accented vowel decreased the number of question responses. Figure 2 (right) presents mean question results for the plateau continuum. These results appear immediately strikingly different from those of the peak continuum. What we notice is that a great number of question responses was already obtained early in the continuum. Stimulus 2, for instance, presents a score that is already well above chance (0.75), while this is not the case for the peak series.

We then tested the hypothesis that the shape of the peak can affect the perception of target location, again by assuming that this perceptual difference would translate into a greater or smaller percentage of question responses on the part of the listeners. Here, we found a difference between the results of primary continuum stimuli timed at plateau onset and plateau stimuli. A much smaller difference was found when a comparison was made between the results of primary continuum stimuli timed at plateau offset and plateau stimuli.

We interpreted these results to show that shape of the accent peak does indeed affect the perception of target location, and that the perceived target for plateau stimuli must be displaced somewhat towards plateau offset, though not exactly timed with this location. Above all, we take this result to mean that the perceived target of plateau stimuli cannot be identified with the end of the LH rise within the rise-fall.

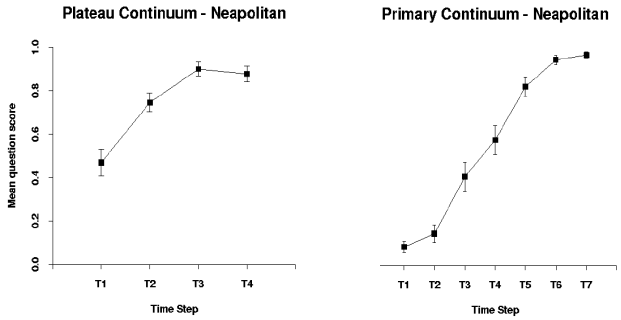

Figure 2: *Mean question scores for the peak series (left) and the plateau series (right). From [5].*

2.1.2. Pisa Italian

A perception test was performed [13] in order to find out whether the H* (followed by a low edge tone) and H*+L were 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 peak and following low (i.e., changing the rise-to-peak slope), and the scaling of all the target tones. The alignment and scaling interplay found in the perception of these two pitch accents appeared to be well-suited for comparing subjects perception of peak and plateaux stimuli.

In [12] we employed the same base utterance used for previous investigations on Pisa Italian and created two series of stimuli in order to get a gradual shift from H*+L to H* L-, both with high plateaux and high peaks. Two main manipulations were performed through PRAAT [17], with PSOLA resynthesis (same manipulation as in Fig. 1).

.In the peak-configuration series, the alignment of the H*+L rise-fall was forward shifted in 8 steps (rising and falling slopes were kept unvaried) and, for each alignment step, the rise-fall was shifted upwards on the Hz scale in 2 steps (13, 17, and 6 Hz for each F0 turning point), creating two additional continua. In the other series, a plateau of 45 ms

was created, so that the duration of the H tone corresponded to three steps of the alignment manipulation for the peak series.

The alignment of the plateaux stimuli was forward shifted in 6 steps (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: in particular, the plateau rise was always matched to a peak stimulus rise and its fall corresponded to the fall of a peak stimulus that was three steps away (at lest for 6 steps); for each alignment step, all turning points were shifted upwards on the Hz scale in 2 steps, as in the other series.

The hypothesis was that scores for plateaux would be similar to scores for late and higher peaks, showing a category shift [3,11]. An identification test was performed asking 14 subjects to listen to 3 repetitions of the stimuli. The results relevant to the discussion at hand are the following: Plateau-configuration results were similar to late peak results in the case stimuli showed the same scaling, in line with the findings on Neapolitan. The mean of subject answers for the two series in the same scaling condition showed that the score obtained for plateau stimuli was similar to the score obtained for peak stimuli matching as to fall alignment. For instance, in the intermediate scaling condition ('Pr1') stimuli were identified as not contrastive since step '4' in the peak series and step '1&4' in the plateaux series (that is, when the plateaux were showing the rising phase matching peak stimulus '1' and the falling phase matching peak stimulus '4', see Figure 3). On the other hand, plateaux were perceived as higher in pitch than peak stimuli aligned at rise beginning (see [12] for further discussion).

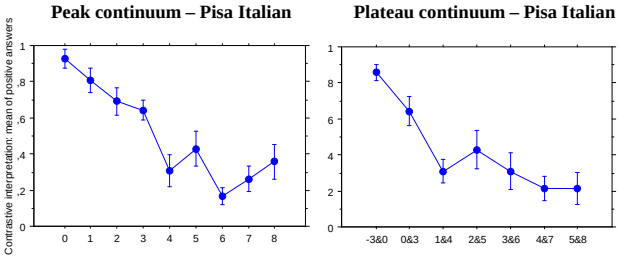

Figure 3: Mean contrastive scores for the peak series (left) and the plateau series (right) in the intermediate scaling condition. From [12].

2.2 Northern Standard German

Different from Neapolitan and Pisa Italian, pitch accents of Northern Standard German are not directly used for distinguishing questions and statements or focus type, although recent studies revealed that shape variations within the rising or falling movements caused by pitch accents can contribute to the question-statement distinction [10]. Rather, the pitch-accent paradigm of Northern Standard German is involved in conveying discourse structure. In this, H* is one of the most frequent accent categories [14] and signals that the associated piece of information is to be taken as new.

The relationship between F0 peaks and H* perception has been intensively studied by [7]. The author claimed that the identification of H* (or the medial peak in terms of the Kiel Intonation Model, KIM, [15]) requires an F0 rise peaking within the accented vowel. Later, [9,16] refined this basic picture by showing that not only peak alignment, but also scaling and slope of the peak movements determine H* identification. However, F0 peaks having their rise offset and hence their high target within the accented vowel remained clear H* indicators. The present perception experiment continues this line of research and adds a further variable to this interplay: F0 plateaux.

As indicated by the contrastive peak-plateau analyses for Italian, the high target in F0 plateaux is not located at the rise offset, but around the *fall onset*. Moreover, in terms of scaling, longer plateaux correspond to higher F0 peaks. These findings suggest that the perception of high targets is governed by a *universal* integration-time mechanism, which should hence also hold for German. If this is the case, then a rise beyond the vowel onset is not the only means to place a high target within the vowel and to signal German H*. Extending an F0 plateau and hence the fall onset into the accented vowel should also yield H* identifications.

Testing this hypothesis did not require two parallel stimulus continua in which entire rising-falling peak and plateau contours are shifted into the accented vowel, as in Neapolitan and Pisa Italian. Instead, the stimulus continuum that was resynthesized with PSOLA in PRAAT [17] started from a stylized rising-falling F0 peak aligned at the onset of the accented vowel. Then, solely the *fall* was shifted into the vowel in 10 steps of 10 ms, while the *rise* remained constantly aligned at the vowel onset. So, across the stimuli, the sharp F0 peak was changed into a flat and successively extended plateau contour. The longest (i.e. the 10th) plateau covers almost the entire vowel. The peak-to-plateau continuum was created in the continuously voiced utterance *"über Langeland"* ('via Langeland') with the (only) accent on *"Lan-"* ([laŋ]).

The stimuli were integrated into an indirect identification task. That is, instead of asking the subjects directly to identify the pitch-accent category or a particular metalinguistic label, they judged whether the stimuli matched semantic-pragmatically with a constant context precursor. In the present experiment this was the question *"Und wie willst du fahren?"* ('And which way do you want to go?'). This signals that the route of the dialogue partner was neither predetermined by a third person nor mentioned before in the discourse. Hence, the context precursor was designed to be only compatible with the following stimuli if the latter contain an H* pitch accent (which opens a new argument). In a previous pilot experiment, note that we verified that the phonetically adjacent pitch-accent categories, H+L* and L*+H, did not match with the context precursor (cf. [16]). Five copies of the context-stimulus pairs were arranged in a randomized order and judged by 18 native speakers of German (average age 26.7 years).

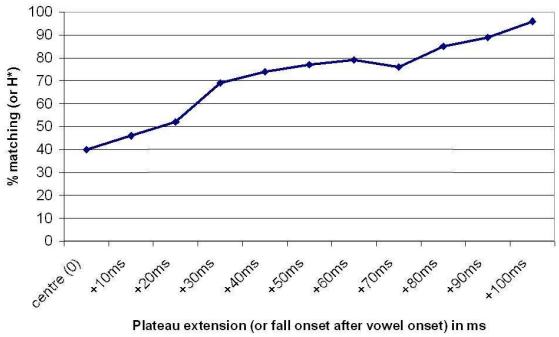

Figure 4: Percentages of matching judgements, representing H identifications in the 10 stimuli in which the onset of the F0 fall and hence the end of the high F0 plateau is shifted successively into the accented vowel (n=90).*

The results in Figure 4 show, in line with the hypothesis, that the plateau extension into the vowel leads to a clear transition in the majority of judgements from 'not matching' to 'matching', which represents a perceptual change to H* identifications. Thus, the results clearly contradict the claim in [7] that for H* there *must* be an F0 rise into the accented vowel. High F0 plateaux which show no rise, but a late fall in the accented vowel, have a similar perceptual effect. So, as in the case of the Italian pitch accents, plateaux must also be taken into account in the interplay of F0 variables signalling German pitch accent categories.

3. Discussion

In this paper we report results for a series of perception experiments in two varieties of Italian and in German assessing the effect of plateaux on the perception of contrast between pitch accents exhibiting different alignment properties, i.e. an early vs. a late peak accent. In both Italian varieties, both peak and plateaux stimuli were created. As already shown, the peak series successfully shifted the perception from one category to the other (question vs. statement for Neapolitan, contrastive vs. utterance-initial broad focus for Pisa Italian) when the alignment of the three targets within a LHL rise-fall was forward or backward shifted within the stressed vowel. As for Neapolitan, this confirms results for the inter-peak stimuli shown in [8], while Pisa Italian results confirm findings in [13].

Crucially, we also showed that by modifying the shape of the peak within the rise-fall one can affect identification scores for a certain category in the Italian varieties. That is, when the accent peak is in the shape of a pitch plateau, either more question responses (for Neapolitan) or less contrastive responses (Pisa Italian) were recorded relative to stimuli with a sharp peak timed at plateau onset. On the other hand, much more similar results were obtained when the plateau stimuli scores were compared to those for stimuli with sharp peak timed at plateau offset. We interpret this result as an effect of peak shape on perceived target location. That is, the perceived target location appears to be closer to plateau offset than onset. This phenomenon (which was replicated for other stimuli in [3]), cannot be explained in terms of language-specific constraints on prosodic parsing.

The experiment reported for German, though slightly different in methodology (since plateau duration was not kept constant and a context-matching study was employed), again shows that the longer the duration of the H tone, the later is its perceived target. This suggests that the observed effect is psychoacoustic in nature and is potentially cross-linguistically shared. Taken together, our results are problematic for tonal center of gravity approaches [18], since the prediction would be that plateau stimuli would match with peak stimuli timed at the center of the plateau, which was not the case for any of the languages studied here.

4. Conclusion

The relationship between acoustic tone targets and perceived targets was tested in two varieties of Italian and German through a set of perception experiments. The experiments show that details of temporal alignment of target tones as well as the shape of the peak within a rising accent affect the identification of the contrast, for both languages. These findings have implications for both intonational phonology and speech recognition. We propose that both psychoacoustic and "learned", language-specific, constraints on perception must shape the intonational phonological categories of a language, and cannot be merely relegated to lower-level, quantitative effects. Finally, our results do not support a simple "tone center of gravity" approach intended to explain the perception effects.

5. References

[1] D'Imperio, M., Arvaniti A. and S. Frota (forthcoming). "Prosodic Representations". In Cohn A.C., Fougeron C. and M. Huffman (Eds.) *The Oxford Handbook of Laboratory Phonology*, ch. 9. Oxford University Press.

[2] Ladd, D.R., Faulkner, D., Faulkner, H. & A. Schepman (1999). 'Constant "segmental anchoring" of F0 movements under changes in speech rate'. *Journal of the Acoustical Society of America*, 106(3): 1543-1554.

[3] D'Imperio, M. (2000). *The role of perception in defining tonal targets and their alignment.*Doctoral dissertation, Ohio State University.

[4] Gili-Fivela, B., & Savino, M. (2003). 'Segments, syllables and tonal alignment: A study on two varieties of Italian'. In M. J. Sole´ ,D. Recasens, & J. Romero (Eds.), *Proceedings of the XVth International Congress of Phonetic Sciences* (pp. 2933–2936). Barcelona: Causal Productions.

[5] Prieto, P. & Torreira, F. (2007). "The segmental anchoring hypothesis revisited. Syllable structure and speech rate effects on peak timing in Spanish". *Journal of Phonetics* 35.4, p. 473-500.

[6] Pierrehumbert, J.B. & S. Steele (1989). 'Categories of tonal alignment in English'. *Phonetica*, 46 : 181-196.

[7] Kohler, K. (1987). 'Categorical pitch perception'. In *Proceedings of the XIth International Congress of Phonetic Sciences*, Tallin, Estonia, August 1-7, 1987, pp. 331-333.

[8] D'Imperio, M., & D. House. (1997). "Perception of questions and statements in Neapolitan Italian". In Kokkinakis, G., Fakotakis, N. et Dermatas, E. (eds.), *Proceedings of Eurospeech'97*, Rhodes, Greece, vol. 1, pp. 251-254.

[9] Niebuhr, O. (2003). Perceptual study of timing variables in F0 peaks. Proc of the 15th ICPhS, Barcelona, Spain, 1225-1228.

[10] Petrone, C., & Niebuhr, O. (2008). The role of the prenuclear F0 region in the identification of questions and statements. Paper presented at *Phonetics & Phonology in Iberia*. *www.linguistik.unikiel.de/downloads/PAPI_Petrone_Niebuhr_final.pdf*

[11] Knight, R. and F. Nolan (2006). "The effect of pitch span of intonational plateaux. JIPA, 36(1), 1-28.

[12] Gili Fivela B., D'Imperio M. (2010). "High peaks versus high plateaux in the identification of two pitch accents in Pisa Italian", Proceedings of Speech Prosody, this volume.

[13] Gili Fivela, B. 2005. "La percezione degli accenti: il ruolo dell'allineamento e dello 'scaling' dei bersagli tonali". In *Proceedings of Convegno Nazionale AISV (Associazione Italiana di Scienze della Voce) "Misura dei parametri"*. Padova. Italy. December 2-4. 313-326.

[14] Peters, B., Kohler, K.J., & Wesener, T. (2006) Melodische Satzakzentmuster in prosodischen Phrasen deutscher Spontansprache – statistische Verteilung und sprachliche Funktion. *AIPUK 35a*, 7-54.

[15] Kohler, K.J. (1991). A model of German intonation. *AIPUK 25*, 295-360.

[16] Niebuhr, O. (2007). The signalling of German rising-falling intonation categories - Interplay of synchroniz-ation, shape, and height. *Phonetica 64*, 174-193.

[17] Boersma, P. & Weenick, D., "Praat: doing phonetics by computer" (Version 4.5.18), 2007, http://praat.org/.

[18] Veilleux, N., Barnes, J., Shattuck-Hufnagel, S. and Brugos, A. (2009). "Perceptual Robustness of the Tonal Center of Gravity for Contour Classification", poster presented at the "Workshop on Prosody and Meaning", Sept. 17-18, Univeristy Pompeu Fabra, Barcelona, Spain.