Coarticulation and stability effects in tonal clash contexts in Catalan

Pilar Prieto

Departament de Filologia Catalana
Universitat Autònoma de Barcelona
pilar.prieto@uab.es

Abstract

This paper discusses the effects of *tonal clash* (or strict adjacency between two accents) on the phonetic realization of H* accents in Catalan. The analysis of the data shows that the adjacency of two rising accents triggers a drastic temporal reorganization of the F_o gestures involved, resulting in anticipation of the first gesture and delay of the second. Specifically, in crowded tonal environments we find that the first peak is placed significantly earlier with respect to the syllable onset and offset and the start of the second rise is located significantly later in the second syllable. Thus, the data reveals that F_o gestures are roughly timed to accented syllables, keeping a more o less floating relationship with the segmentals: both right and left-hand prosodic contexts (including clash situations) determine the alignment patterns of both the valley and the peak. Finally, no significant differences were found on the H scaling in clash vs. non-clash environments.

1. Introduction

The main goal of this study is to describe the effects of tonal clash on alignment and scaling patterns of H* prenuclear accents (or simple peaks) in Catalan.[1] As it is well-known, languages resolve stress clash situations through two different strategies**:** (1) *avoidance*: stress/accent shift is frequently used in English to avoid placing accents on adjacent or near-adjacent syllables (*achromátic léns* > *áchromatic léns*; *thirtéen mén* > *thírteen mén* [16]; also, deaccenting is widely used in Catalan (*déu néns* > *deu néns*; [9], [15])[2] and in Italian clash environments (*la cittá vécchia* > *la citta vécchia*; [6], [8]);[3] (2) *tonal clash*, which can be resolved through tonal overlap or tonal repulsion: for example, in English, the presence of competing pitch movements can also be resolved by anticipation of the first gesture or delay of the second ([2] for Swedish, [17] for English); similarly, Catalan and Italian can optionally use this strategy: *en Pép dórm a la platja* > *en Pép * dórm a la platja* [9], [15]; *la veritá vínce quasi sempre* > *la veritá * vínce quasi sempre* [8]. Thus, Catalan speakers can either choose to delete the first accent (cf. *nen blanc,* top panel of Fig. 1) or pronounce both accent gestures (cf. *nén blanc,* bottom panel of Fig. 1). This study focuses on the second type of strategy, hoping that it will contribute to shed some light on our understanding of the production and coarticulation of F_o gestures and their coupling with the segmentals.

In general, very few studies have dealt with the problem of tonal clash resolution and their effects on alignment and scaling of accent gestures. Previous work on alignment of H and L targets has generally emphasized its anticipatory effects, and more specifically, the influence of the right-hand prosodic context ([17] for English, [12] for Spanish), while carryover effects have been very scarcely analyzed. Some fairly recent experiments on tonal clash ([11] for Spanish, [18] for Mandarin Chinese) reveal that time pressure affects the position of both H and L in both accents, clearly indicating the presence of systematic anticipatory and carryover coarticulation effects. Moreover, d'Imperio [4] shows that H timing in Neapolitan Italian is not influenced by the right-hand prosodic environment (distance to L- and prosodic boundaries), but by the distance of H to the beginning of the word. Regarding LH scaling, previous studies have highlighted that the pitch level of H points is extremely constant across repetitions of the same sentence ([7], [12]). The study reported here is intended to contribute to this line of research and test the effects of tonal pressure on pitch alignment and scaling of LH points and compare the strength of anticipatory vs. carryover coarticulatory effects on tonal realization vastly studied in segmental phonology.

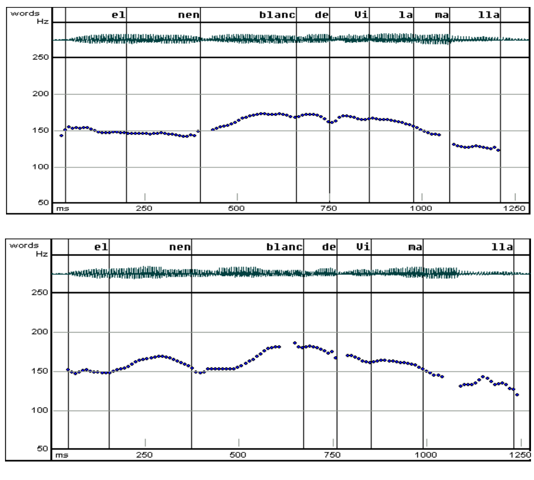


Figure 1: Waveforms and F_o contours of the unaccented (top panel) and accented version (bottom panel) of the clash sequence *El nén blánc de Vilamalla.*

[1] Prenuclear accents in Catalan are characterized by a simple rise, from an F_o *minimum* generally starting at the onset of the accented syllable to a peak generally located on the following unaccented syllable. These accents have also been analyzed as L*H by Estebas-Vilaplana [5].

[2] A recent investigation [15] has shown that clash sequences in spontaneous speech are generally resolved through total deaccenting of the first syllable, making *camí nét* 'clean path' perceptually homophonous with *caminét* 'little path'. Note that throughout the article stress marks correspond to phonetically stressed syllables.

2. Speech materials

The materials under study consist of 240 utterances (24 target sentences x 2 speakers x 5 repetitions). The same clash ((1a) *molí nét* 'clean mill') and non-clash sequences ((1b) *molinét* 'little mill' and (1c) *molí netét* 'little.dim mill') were placed at the beginning of a three-accent utterance, in such a way that one can easily test the effects of the presence vs. absence of a clash environment on both the first and the second accent.[3] (1) illustrates the coding scheme used: S-S=clash; U-S=unstressed-stressed; S-U=stressed-unstressed.[4]

(1) a. *El molí nét no li agrada* (S-S)
b. *El molinét no li agrada* (U-S)
c. *El molí netét no li agrada* (S-U)

The target utterances were read six times by two speakers of Central Catalan, PG and PP. In order to obtain a pronunciation with two adjacent pitch accents, we recommended a slow speech rate and listed (1a) and (1b) together, clearly indicating that the sentences should be distinguished one from the other. Despite such indications, cases of accent deletion and overlap were found in the data, and were discarded later. The following measurements were manually placed in each sound file avoiding errors and segmental effects on the pitch contour: utterance-initial F_o value (In), utterance-final value (Fin), highest F_o peak for every pitch accent (H1, H2), and lowest F_o value between pitch accents (L1, L2). To calculate the timing values of peaks and valleys, the following measurements were made with the help of spectrograms: onset (On1, On2) and offset (Of) of every target syllable. Fig. 2 illustrates the waveform, F_o contour and labeling scheme of the utterance *El molí nét no li agrada*.

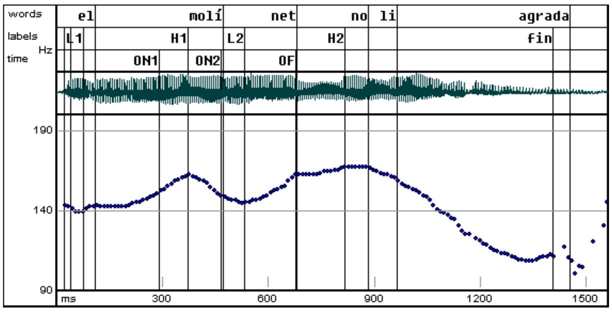


Figure 2: Waveform display, F_o contour and labels corresponding to the utterance *El molí nét no li agrada.*

3. Results

3.1. Pitch scaling of LH target points

Utterance-initial and utterance-final values were almost constant and totally uncorrelated with clash/non-clash conditions for the two speakers. Mean utterance-initial and utterance-final values are shown in Table 1 in the three conditions. Standard deviation values demonstrate the small variability present in the data: SD for starting F_o values are 5.87 Hz for speaker PG and 8.56 Hz for speaker PP; SD for ending F_o values are 6.32 Hz for speaker PG and 7.90 for speaker PP. These results are in agreement with previous results for Spanish ([12], [13]):[5]

Table 1: Mean values (in Hz) of utterance-initial and utterance-final F_o values

Speakers	Initial F_o Value			Final F_o Value		
	S-S	U-S	S-U	S-S	U-S	S-U
PG	117	122	121	97	97	97
PP	157	161	159	125	124	125

Regarding the scaling behavior of H target points, previous investigations on descending intonational contours ([17] for English, [13] for Spanish) have revealed that H peaks are highly stable and they can be predicted quite successfully by a local downstep ratio of constant reduction from the previous' peak value. Our data, though, does not display a decay F_o pattern but a raising trend where H2 is always higher than H1 (cf. bottom panel of Fig. 1). Table 2 shows the mean values (in Hz) of H1 and H2 peaks in three different conditions (SS, U-S, S-U) for the two speakers. Indeed, we compare H1 in the S-S/S-U conditions and H2 in the S-S/U-S conditions —note that H1 is an empty value in the U-S condition. The data show that, despite the fact that we are dealing with a non-descending contour, H1 and H2 continue to be rather constant within speakers and across different clash conditions.

Table 1: Mean values (in Hz) of H1 and H2 peak values in two different conditions (SS, S-U for H1 and S-S, U-S for H2) for the two speakers

Speakers	H1 Value		H2 Value	
	S-S	S-U	S-S	U-S
PG	153	157	171	169
PP	171	171	184	184

What is the effect of clash contexts on L scaling? Table 2 displays the mean values (in Hz) of L1 and L2 points in three different conditions (SS, U-S, S-U) for the two speakers. As in the preceding analysis, L1 is missing in the U-S condition and L2 in the S-U condition. Results in Table 2 clearly show a significant increase in F_o of the second valley when there is no intervening syllable in between the accents, indicating that the falling gesture seems to accommodate to the number of intervening syllables between accents. *T*-tests comparing L2 in these two conditions show that the two populations are highly distinct (at $p < 0.01$) for the two subjects. By contrast, L1 values are near-constant in clash and non-clash environments (indeed, the distance in syllables from the first accented syllable to the beginning of the utterance is the same in every sentence).

[3] The 24 target sequences in the database are the following: *molí nét*, *molinét*, *molí netét*; *llumí nét*, *lluminét*, *llumí netét*; *camí nét*, *caminét*, *camí netét*; *veí nét*, *veïnét*, *veí netét*; *meló nét*, *melonét*, *meló netét*; *remolí nét*, *remolinét*, *remolí netét*; *colomí nét*, *colominét*, *colomí netét*; *violí nét*, *violinét*, *violí netét*.

[4] To obtain a clean F_o contour and minimize errors, we used sonorant consonants throughout the target syllables.

[5] Note that utterance-length and distance in syllables from the beginning of the utterance to the first accent is the same in all the sentences in the database.

Table 1: Mean values (in Hz) of L1 and L2 in two different conditions (SS, S-U for L1 and S-S, U-S for L2)

Speakers	L1 Value		L2 Value	
	S-S	S-U	S-S	U-S
PG	125	126	**145**	126
PP	157	159	**154**	149

As expected, the data display a strong correlation between L2 and the value of the preceding peak (R^2=0,68 for PG; R^2=0,30 for PP) and between L2 and the distance between this point and the preceding peak (R^2=0,63 for PG and R^2=0,55). This tendency has been already reported in languages like English [10], Japanese [3], and Spanish [14]. Pierrehumbert, for example, suggests that the degree of dipping between peaks depends on the amount of time available, as F_o "falls until it is time to start aiming for the next H* level." ([10], p. 71). In our data, the contrast between the scaling behavior of H and L values seem to indicate that H points function as "real" targets and the speaker makes the effort to reach these values despite the time constraints; by contrast, L behaves more as an interpolation function, and, consequently, the speaker cap opt to "undershoot" it under time pressure conditions.

3.2. Pitch timing of LH target points

In this section we analyze the temporal behavior of H and L tonal targets in clash vs. non-clash contexts. Previous investigations on the alignment behavior of LH targets in non-clash contexts have noted that, while H positions are more variable and depend on the right-hand prosodic context, L values are consistently "anchored" with the onset of the accented syllable ([17] for English, [12] for Spanish, [1] for Greek). Do tonal clash contexts affect L point alignment patterns? Does the first pitch accent gesture start significantly earlier in clash environments? And the second, does it start later? The following two graphs in Fig. 3 plot the observed distance (in ms) from the start of the first rise to the onset of the accented syllable (which we call *L1 delay*) as a function of the duration of the syllable in two conditions (S-S and U-S, represented by two different plotting characters). The data show that, even though L1 is generally placed before the accented syllable in both clash and non-clash environments, it also tends to be placed slightly earlier in clash contexts (an average of 2 ms for PG and 40 ms for PP). *T*-tests reveal, though, that such differences are only significant for speaker PP (at $p < 0,0001$).

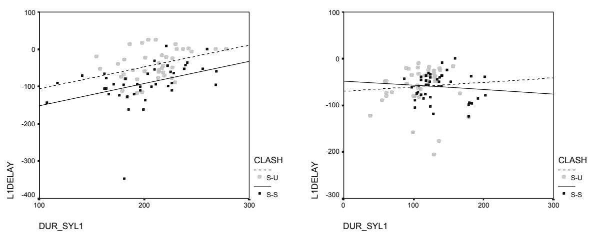


Figure 3: L1 delay values (in ms) as a function of the duration of the first accented syllable in clash/non-clash environments for speaker PP (left-hand panel) and PG (right-hand panel).

The two plots in Fig. 4 demonstrate that L2 delay values are radically different in the two clash conditions for the two speakers. In clash environments the second rising gesture starts significantly later (an average of 38 ms into the accented syllable for speaker PG and 45 ms for speaker PP, roughly 1/3 of the syllable duration) than in non-clash contexts, where L is more or less aligned with the onset position (or even 20 ms. before the onset for the two speakers). Clearly, points separated by the clash condition are located in separate regions along the scaling axis (y-axis). *T*-tests comparing L2-delay values across clash and non-clash environments were highly significant (at $p < 0.001$) for the two speakers. Finally, the two regression lines (represented by dotted and solid lines) summarize the negative correlations between L2 delay and duration of the preceding syllable: that is, the longer the preceding syllable, the closer L2 gets to the onset. This is probably indicating that the speaker can either choose to lengthen the first target syllable (and have more time available) or delay the start of the second gesture.

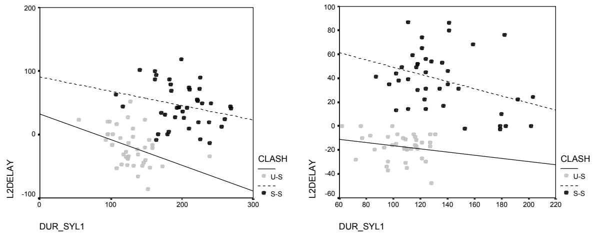
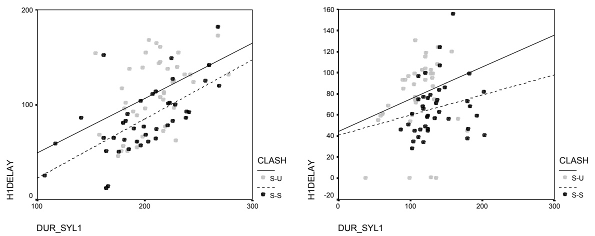


Figure 4: L2 delay values (in ms) as a function of the duration of the first accented syllable in clash/non-clash environments for speaker PP (left-hand panel) and PG (right-hand panel).

Do tonal clash contexts affect H alignment patterns? The four graphs in Fig. 5 plot the distance in time from H1 to the onset of the syllable (*H1 delay*, see top panels) and the distance in time from H1 to the offset of the accented syllable (*H1dist2*, see bottom panels) as a function of the syllable duration in clash vs. non-clash contexts. As expected, the data display a clear effect of clash on H1 alignment. First, *peak delay* is significantly shorter in clash contexts, which means that F_o peaks tend to shift backwards in the presence of a following accented syllable. *T*-tests comparing the values of H1-delay in the two conditions were significant (at $p < 0.0001$) for the two speakers. Second, distance to the offset (*H1Dist2*) increases in clash environments. In our data, H1 is generally reached before the end of the syllable in both clash and non-clash environments: the fact that peaks are not slightly "delayed" in non-clash environments is probably due to the fact that such sentences have only one unstressed syllable in between accents. *T*-tests comparing the values of H1 distance to the offset were significant (at $p < 0.0001$) for the two speakers. The regression lines in the graphs also reveal a positive correlation between *peak delay* and duration of the accented syllable and a negative correlation between *H1dist2* and duration of the accented syllable (that is, as expected, the longer the syllable, the shorter the distance from the peak to the end of the syllable).



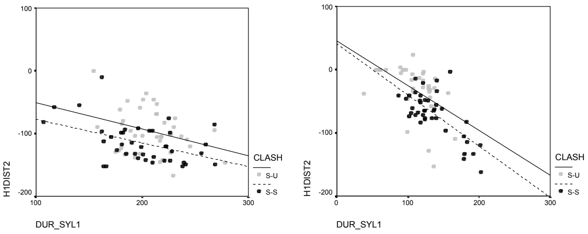


Figure 5: Mean values (in ms) of the distance between the first peak to the onset (top panels) and the offset of the syllable (bottom panels) as a function of the duration of the syllable in two conditions (SS, S-U) for speaker PP (left-hand panels) and PG (right-hand panels).

In contrast with the behavior of H1, no clear effects of pitch clash were found on H2 alignment. In our data, the F_o peak of the second accent was generally synchronized with the offset of the syllable, regardless of the clash condition.

3.3. Effects of clash on rise time

What is the effect of tonal clash on the absolute *rise time* (or temporal distance from the peak to the previous valley) of the H* accents? The results of our data replicate earlier findings by [12] for Spanish and [1] for Greek and show that there is no fixed duration for the rise. Indeed, there is a tendency to have less rise time in pitch clash situations while the duration of the syllable is equal or longer (cf. the two plots in Fig. 6 for rise time values of the second accent and regression lines). *T*-tests are significant (at $p < 0.0001$) for both rise times (first and second accent) for the two speakers.

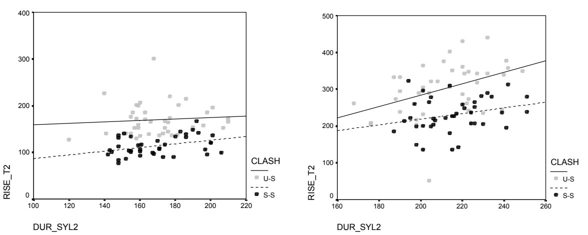


Figure 6: Rise time values (in ms) of the second accent as a function of its duration in clash/non-clash environments for speaker PP (left-hand panel) and PG (right-hand panel).

Further analyses show a significant increase of the velocity of the rising gesture in pitch clash situations, which "accommodates" to the time allocated for the rising gesture. It seems that speakers make the effort of reaching a certain F_o target level both by increasing syllable durations and velocity of F_o gestures (and also by compromising L scaling in between the two Hs), a strategy which confirms the implicit prediction of the pitch target view.

4. Conclusions

The analysis of the data shows consistent differences in the **time domain** between accent gestures in a clash vs. a non-clash context. Clash contexts trigger a drastic timing reorganization of the accents involved, namely, the first peak is placed significantly earlier with respect to the syllable offset and the second valley significantly later with respect to the syllable onset. No clear effects of pitch clash were found on the scaling of H peaks (we find its value to be rather constant across repetitions of the same speaker), showing clear stability effects in the **scaling domain**.

5. References

[1] Arvaniti, A.; Ladd, D.R.; Mennen, I., 1998. Stability of tonal alignment: the case of Greek prenuclear accents, *Journal of Phonetics* 26, 3-25.

[2] Bruce, G., 1977. *Swedish Word Accents in Sentence Perspective*. Gleerup: Lund.

[3] Kubozono, H., 1987. *The Organization of Japanese Prosody*, Ph.D. Dissertation, University of Edinburgh.

[4] D'Imperio, M., 2002. "Italian intonation: an overview and some questions". *Probus*.

[5] Estebas-Vilaplana, E., 2000. *The Use and Realisation of Accentual Focus in Central Catalan*, Ph.D. Diss., University College London, London.

[6] Farnetani, E.; Kori, S., 1983. Interaction of syntactic structure and rhythmical constraints on the realitzation of word prosody. *Quaderni del Centro di Studio per le Ricerche di Fonetica* (Padova), 2, 288-318.

[7] Liberman, M.; Pierrehumbert, J., 1984. Intonational Invariance under Changes in Pitch Range and Length. In *Language Sound Structure. Studies in Phonology Presented to Morris Halle*, M. Aronoff and R. T. Oehrle (eds.). Cambridge, Massachusetts: MIT Press, 157-233.

[8] Nespor, M.; Vogel, I., 1989. On clashes and lapses. *Phonology* 6, 69-116.

[9] Oliva, S., 1992. *La mètrica i el ritme de la prosa*. Barcelona: Quaderns Crema.

[10] Pierrehumbert, J., 1980. *The Phonetics and Phonology of English Intonation*, Ph.D. Dissertation, Massachusetts Institute of Technology.

[11] Prieto, P.; Shih, C. 1995. Effects of Tonal Clash on Downstepped H* Accents in Spanish. *Proceedings of EUROSPEECH'95. Fourth European Conference on Speech Communication and Technology*, vol. 2, 1307-1310.

[12] Prieto, P.; van Santen, J.; Hirschberg, J., 1995. Tonal Alignment Patterns in Spanish. *Journal of Phonetics* 23, 429-451.

[13] Prieto, P.; Shih, C.; Nibert, H., 1996. Pitch Downtrend in Spanish. *Journal of Phonetics* 24, 445-473.

[14] Prieto, P., 1998. The Scaling of the L Tone Line in Spanish Downstepping Contours. *Journal of Phonetics* 26, 261-282.

[15] Prieto, P.; Oliva, S.; Palmada, B.; Serra, P.; Blecua, B.; Llach, S.; Oliva, V., 2001. Manifestació acústica de la resolució de xocs accentuals en català, *Estudios de Fonética Experimental* 11, 11-38.

[16] Selkirk, L., 1984. *Phonology and Syntax. The Relation between Sound and Structure*. Cambridge, MA: MIT University Press.

[17] Silverman, K.; Pierrehumbert, J., 1990. The timing of prenuclear high accents in English. In *Papers in Laboratory Phonology I. Between the Grammar and Physics of Speech*, J. Kingston and M. E. Beckman (eds). Cambridge: Cambridge University Press, 72-106.

[18] Xu, Y. 1999. Effects of tone and focus on the formation and alignment of fo contours, *Journal of Phonetics* 27, 55-105.