Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Phonetics 33 (2005) 215–242
Journal of Phonetics
www.elsevier.com/locate/phonetics

Stability effects in tonal clash contexts in Catalan ☆

Pilar Prieto

Institució Catalana de Recerca i Estudis Avançats (ICREA) and Universitat Autònoma de Barcelona (UAB), Departament de Filologia Catalana, 08193 Bellaterra (Cerdanyola del Vallès), Barcelona, Spain

Received 14 May 2003; received in revised form 27 July 2004; accepted 27 July 2004

Abstract

Recent work in intonational phonology has shown that L and H points are carefully controlled by the speaker, i.e., they are scaled and synchronized with the segmental material in extremely consistent ways (Arvaniti, 2003; Arvaniti, Ladd, & Mennen, 1988, 2000; Ladd, 2003; Liberman and Pierrehumbert, 1984; Prieto, van Santen, & Hirschberg, 1995; Xu, 1999, 2000, among others). Similarly, it has been shown that pitch pressure environments (namely, proximity to a boundary tone or to an upcoming pitch accent) can drastically affect surface H alignment patterns (Silverman & Pierrehumbert, 1990; Prieto, Shih, & Nibert, 1995). What is less well known are the effects of time and pitch pressure environments on surface scaling and L alignment patterns and whether stability effects are found in these domains. This paper discusses the effects of tonal clash (or strict adjacency between two accents) on the phonetic realization of rising prenuclear accents in Catalan. The analysis of the data shows that the adjacency of two rising accents triggers a drastic temporal reorganization of the f_0 gestures involved, resulting in anticipation of the first gesture and delay of the first L of the second. The data thus reveals that f_0 gestures are roughly timed to accented syllables, keeping a more or less floating relationship with the segmental string. Importantly, no significant differences are found on H scaling in clash vs. nonclash environments, revealing that these points work as real production targets. Moreover, in tonal clash contexts the time of the second tonal rise significantly shortens and the velocity of the rising gesture significantly increases in order to reach a given scaling target. Therefore, the results of the present study reveal that, in pitch pressure environments, a clear contrast is found between the stability effects of the H scaling domain and the adapting behavior of the time and velocity domains.

☆Earlier versions of this work were presented at Speech Prosody 2002 (Aix-en-Provence, April 2002), the International colloquium on tone: phonetic and phonological dimensions (Mexico City, March 2002), and at the ESF Workshop on Experimental Approaches to Tone and Intonation (University of Oxford, April 2003).

E-mail address: pilar.prieto@uab.es.

0095-4470/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2004.07.001

1. Introduction

The quest for phonological and linguistic units in the f_0 continuum has been and still is one of the main research questions in intonation studies. In recent years there has been accumulating evidence that LH points in the tonal space behave as phonological targets. Indeed, H peaks of both nuclear and prenuclear pitch accents are produced with an amazing degree of stability in tonal scaling and alignment across languages (see Arvaniti, Ladd, & Mennen, 2000; Ladd, 2003; Liberman & Pierrehumbert, 1984; Prieto, van Santen, & Hirschberg, 1995; Prieto, Shih, and Nibert, 1996; Silverman & Pierrehumbert, 1990; Xu, 1999, 2002, among others). With regard to LH scaling, previous studies have highlighted that the pitch level of H points is extremely constant across repetitions of the same sentence (Liberman and Pierrehumbert (1984) for English; Prieto et al.(1996) for Spanish; Arvaniti (2003) for Greek) and that the height of a given peak can be successfully predicted by a proportion of the preceding H peak (or "downstep ratio"). On the other hand, recent research on the alignment between f_0 targets and the segmental string exhibit strong "segmental anchoring" effects: for example, crosslinguistically, L points have shown to be synchronized with the syllable onset in Dutch (Caspers & van Heuven, 1993), English (Ladd & Schepman, 2003), Greek (Arvaniti et al., 2000), Italian (D'Imperio, 2000), and Spanish (Prieto et al., 1995).

What are less clear are the effects of time and pitch pressure environments on surface alignment and scaling patterns and whether stability effects are systematically found in these pitch domains. In recent literature, there seems to be contradictory evidence in this respect. On the one hand, previous work on alignment of H and L targets has generally emphasized the influence of the right-hand prosodic context, and specifically, the presence of an upcoming accent, on the alignment of accent peaks (Bruce (1977) for Swedish; Silverman and Pierrehumbert (1990) for English; Prieto et al. (1995); Prieto and Shih (1995); Face (2002) for Spanish; Xu (1999) for Mandarin Chinese). However, in another line of work, evidence suggests that alignment of f_0 targets are consistently governed by segmental anchoring and that strict alignment effects are pervasive under changes of speech rate (Ladd, Faulkner, Faulkner, & Schepman (1999) for English and Xu (1998) for Chinese; see also Ladd (2003)). Specifically, Ladd et al.(1999) find that the f_0 valley and peak of a rising accent in English are anchored to their segmental landmarks as segment durations increase or decrease with speech rate. Similarly, Xu (1998) shows that the end of a rising contour in Mandarin Chinese is synchronized with the end of the syllable, regardless of speech rate and syllable composition.

Similarly, partial contradictory evidence is found in analyses of patterns of f_0 scaling. In tonal clash situations, while Prieto and Shih (1995) report strong stability effects in H points in Spanish, Face (2002) observes a tendency for the first peak to undershoot in this environment. On the other hand, L2 in between peaks is found to be significantly higher in clash environments (Prieto & Shih, 1995; Arvaniti et al., 2000; Face, 2002), indicating that H values as tonal targets take precedence over L.

The article sets out to examine the resolution of tonal clash (or strict adjacency between two acents) on alignment and scaling patterns of prenuclear rising accents (or simple peaks) in Catalan, in order to shed some light on our understanding of the production and coarticulation of f_0 gestures and their coupling with the segmental material. Rising prenuclear accents in Catalan can be phonologically characterized as instances of L* + H accents. As pointed out in Prieto

(2002, Chap. 11) and Estebas-Vilaplana (2000), in unmarked prosodic environments the L valley is always 'anchored' to the onset of the accented syllable and the H target is delayed, being generally located in the posttonic syllable. Estebas-Vilaplana (2003) provides recent evidence that the H target aligns with the word-boundary and thus proposes to characterize Catalan prenuclear accents as sequence of a monotonal L* tone associated with the stressed syllable and a word edge tone associated to the end of the word. We believe that the behavior of this H peak as a word-edge tone or as a trailing tone needs further elucidation (see Estebas-Vilaplana & Prieto, in preparation) and that this matter does not crucially affect the general goal of this paper, namely, to examine the effects of tonal crowding on two phonologically specified targets LH.

As is well-known, languages resolve stress clash situations using different repair strategies, and tonal means are just one of the possible solutions invoked in clashing situations. For example, stress or accent shift is frequently used in English to avoid placing accents on adjacent or near-adjacent syllables (*achromatic lens* > *achromatic lens*; *thirteen men* > *thirteen men;* Selkirk, 1984, among many others). Destressing and deaccenting of the first accent involved in the clash is widely used in Catalan (*deu nens* > *deu nens*; Oliva, 1992; Prieto et al., 2001). Tonal repulsion or overlap are two other options used by languages: for example, in English, the presence of competing pitch movements can be resolved by anticipation of the first gesture or delay of the second (Bruce (1977) for Swedish; Silverman and Pierrehumbert (1990) for English).[1] The two graphs in Fig. 1 illustrate two optional clash repair strategies found in Catalan: (a) deaccenting (or deletion) of the first accent (cf. *nen blanc*, top panel of Fig. 1); (b) repulsion of the two accent gestures (cf. *nen blanc*, bottom panel of Fig. 1).

It is the second type of strategy, i.e., the presence of two adjacent and competing pitch movements, that allows us to investigate the effects of tonal clash on the phonetic realization of accent gestures and their alignment with the segmental content.

2. Speech materials

In our materials, the same clash ((1a) *mo lí net* 'clean mill') and nonclash sequences ((1b) molinet 'little mill' and (1c) *mo lí ne tet* 'clean.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: (1) illustrates the coding scheme used: S–S = stressed–stressed (accent clash); U–S = unstressed–stressed; S–U = stressed–unstressed. Stressed syllables are underlined:

(1)	(a)	El molí net no li agrada 'He/she does not like the clean mill'	(S–S)
	(b)	El molinet no li agrada 'He/she does not like the little mill'	(U–S)
	(c)	El molí netet no li agrada 'He/she does not like the clean.dim mill'	(S–U)

[1]Catalan speakers can also resolve the production of two rising accents by tonal repulsion and also by overlap (i.e., the first accent is produced with a rising tonal gesture and the second with a falling gesture).

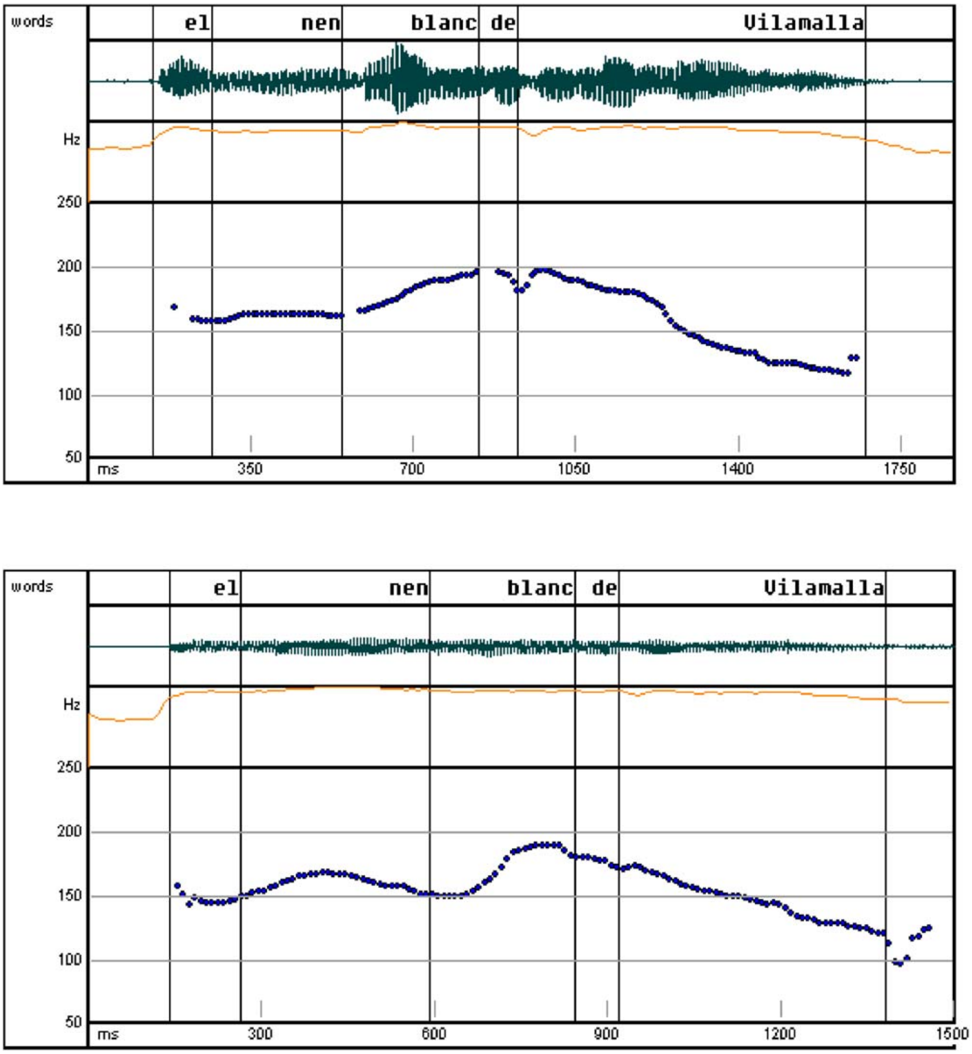

Fig. 1. Waveforms and f_0 contours of the unaccented (top panel) and accented version (bottom panel) of the clash sequence *El nen blanc de Vilamalla* 'The white child from Vilamalla'.

The corpus consisted of the 8 triplet sequences shown in (2). An effort was made to minimize the number of obstruents in the target sequences—note that word-final [t] in *net* was in most of the cases totally assimilated to the following nasal, obtaining an [n]).[2] These triplets were placed in

[2]In Catalan, obstruents in coda position undergo voicing assimilation to the next consonant. Similarly, it is common for plosive consonants [t, d] to undergo manner assimilation to nasals and to laterals. As G. Docherty (p.c.) suggests, the application of the latter process might be well related to fast rate and to a less formal style; yet, no experiments have been carried out to confirm this hypothesis. In our data, manner assimilation was systematically produced by 3 of the speakers (that is, the final final [t] in net turned into [n]), while one of the speakers produced voicing assimilation (that is,

initial position in utterances with the same prosodic structure—cf. Appendix A, which contains the complete list of utterances read by the subjects. The length of the utterances was also kept constant, except for the utterances containing *remo lí*, *colo mí* and *violí* which contain one more syllable.

(2)	molí net	llumí net	camí net	veí net
	molinet	lluminet	caminet	veïnet
	molí netet	llumí netet	camí netet	veí netet
	moló net	remolí net	colomí net	violí net
	melonet	remolinet	colominet	violinet
	meló netet	remolí netet	colomí netet	violí netet

The target utterances were read six times by four speakers of Central Catalan (two females and two males), and five repetitions were used for measurement. Thus the materials under study consist of 480 utterances (24 target sentences × 4 speakers × 5 repetitions). In order to obtain a pronunciation with two adjacent pitch accents and not other optional readings with deaccenting or overlapping, the experimenter recommended a slow speech rate to the informants. Also, tokens (1a; *molí net*) and (1b *molinet*) were listed together and it was indicated that the sentences should be distinguished one from the other. This triggered most of the time a pronunciation with adjacent pitch accents, as shown in Fig. 3. Despite such indications, isolated instances of accent deletion were found in the data (a mean of about 10 tokens per speaker), which were discarded later and replaced by instances from the extra reading materials (recall that 6 repetitions were recorded and only the first 5 were used).

The following measurements were manually placed in each sound file avoiding errors and segmental effects on the pitch contour: utterance-initial f_0 value (In), utterance-final value (Fin), highest f_0 peak for every pitch accent (H1, H2), and lowest f_0 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. Also the onset of the preceding syllable (On0). Fig. 2 illustrates the waveform, f_0 contour and labeling scheme of the utterance *El molí net no li agrada* '(S)he does not like the clean wind mill'

The measurements were made on simultaneous display of waveform, wide-band spectrograms and f_0 tracks. In some cases the identification of peaks and valleys was not trivial. For example, when the L or H points formed a plateau where no clear f_0 value emerged as the lowest or the highest, the midpoint in the plateau was selected (to illustrate this point, see the location of L1 in Fig. 2). Also, microprosodic effects (such as the typical dip produced by nasal segments) were disregarded. With regards to the location of segmental boundaries across vowels and sonorants [m,n,l,r],[3] standard segmentation procedures using spectrograms were followed (see for example

(*footnote continued*)
the final [t] in net turned into [d]). For more information about these phonological processes, see Bonet and Lloret (1998).

[3]Around 90% of the segmentation cases in the database are instances of sonorant + vowel or vowel + sonorant combinations. As mentioned in footnote 2, the plosive coda consonant [t] in the word *net* 'clean' was totally assimilated to the following nasal [n] by 3 of the 4 speakers (see Fig. 2). The fourth speaker produced the voiced counterpart [d].

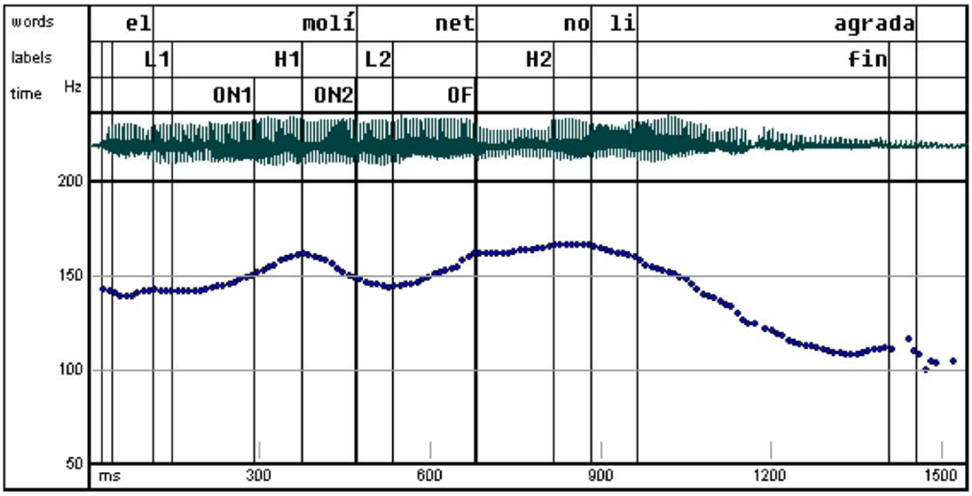

Fig. 2. Waveform display, f_0 contour and labels corresponding to the utterance *El mo <u>lí</u> <u>net</u> no li agrada*.

Peterson and Lehiste, 1960). The beginning or end of a sonorant consonant was identified at the start of the abrupt change from the steady-state period in the spectrogram to the onglide transition movement to the vowel.

3. Results

3.1. Pitch scaling

3.1.1. Utterance-initial and utterance-final values

In order to check the potential variation in the reference scaling points across sentences in different conditions, utterance-initial and utterance-final values were analyzed. In essence, checking utterance-initial and final f_0 values is a way of testing that a given speaker exhibits no variation in the production of the two 'reference' constant points across utterances and thus that sentences are produced in a comparable pitch register. Mean utterance-initial and utterance-final values are shown in Fig. 3 in the three stress conditions (S–S = stressed–stressed; S–U = stressed–unstressed; and U–S = unstressed–stressed) for the four subjects: 120 data points (40 per stress condition) were analyzed per speaker. Mean standard error bars demonstrate the small variability present in the data within groups: sd for starting f_0 values are 4.42 Hz (JP), 5.22 Hz (LR), 5.87 Hz (PG), and 8.56 Hz (PP); sd for ending f_0 values are 3.5 (JP), 3.21 (LR), 6.32 Hz (PG), and 7.90 (PP). By visually inspecting the graph, it is clear that there is no effect of clash on utterance-initial or utterance-final values. In agreement with previous results for English (Liberman & Pierrehumbert, 1984) and Spanish (Prieto et al., 1995, 1996), these values are almost constant across clash/no-clash conditions for the four speakers. The only exception to this pattern is speaker LR, who displays a significantly higher utterance-initial value in clash conditions: this value reflects the fact that LR emphatically pronounced utterances in the clash condition because of the contrastive effect triggered in part by the eliciting situation. This means that this speaker will also display higher H1 and H2 values (cf. Section 3.1.2).

A one-way ANOVA was run with utterance-initial and utterance-final f_0 values as dependent variables and the stress condition as an independent variable. Table 1 shows the ANOVA results for utterance-initial and utterance-final f_0 values. As expected, no clear significant effects of stress condition are found on the scaling of utterance-initial and utterance-final values. In the case of utterance-initial values, speaker PG and speaker LR (who pronounced utterances emphatically in the clash condition) do show an effect.

3.1.2. *H1 and H2 scaling*

Regarding the scaling behavior of H target points, previous investigations of descending intonational contours (Silverman and Pierrehumbert (1990) for English; Prieto et al. (1996) for Spanish; Arvaniti (2003) for Greek) 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 decaying f_0 pattern but a rising trend where H2 is always higher than H1 (cf. Fig. 2). In clash environments, we find contradictory results in the literature: while Prieto and Shih (1995) report clear stability effects on H, Face (2002) observes a small effect of distance between accents on H1 scaling: that is, H1 tends to be lower as the distance to the following accent decreases. Face suggests that this effect is due to the fact that the time allowed for producing the rising gesture is reduced in clash situations.

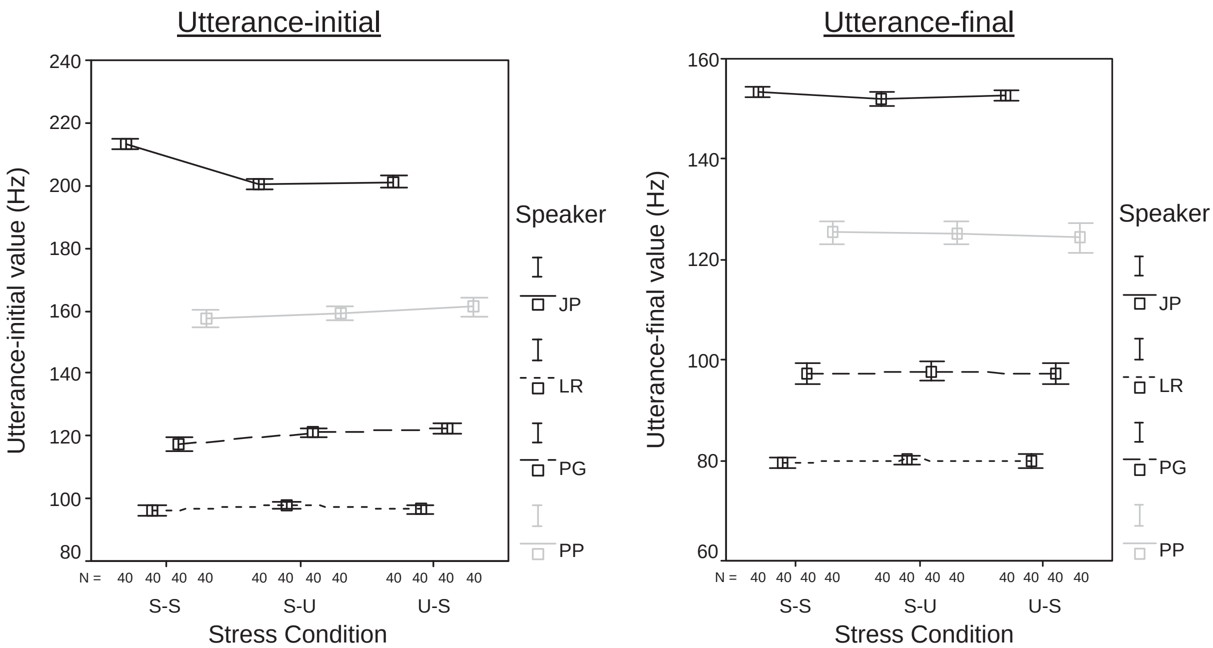

Fig. 3. Mean and standard error values (in Hz) of utterance-initial and utterance-final f_0 values in three different stress conditions (S–S = clash; U–S = unstressed–stressed; S–U = stressed–unstressed) for each speaker.

Table 1
One-way ANOVA results per speaker

	Main effect of clash on f_0-init	Main effect of clash on f_0-fin
Speaker LR	$F_{2,117} = 67.028$; $p < 0.0005$	$F_{2,117} = 1.402$; $p = 0.250$
Speaker JP	$F_{2,117} = 1.931$; $p = 0.150$	$F_{2,117} = 0.217$; $p = 0.805$
Speaker PG	$F_{2,117} = 7.402$; $p = 0.008$	$F_{2,117} = 0.068$; $p = 0.935$
Speaker PP	$F_{2,117} = 2.034$; $p = 0.135$	$F_{2,117} = 0.206$; $p = 0.814$

Utterance-initial and utterance-final values as dependent variables and stress condition as an independent variable

Fig. 4 shows the mean values (in Hz) of H1 and H2 peaks in three different conditions (SS, U–S, S–U) for the four speakers. Indeed, H1 is compared 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 and H2 in the S–U condition. The results show that, in general, and despite the fact that we are dealing with a nondescending contour, H1 and H2 continue to be rather constant within speakers and across different clash conditions. Speakers PG and JP display a small lowering effect on H1 in clash environments (an average of 4 Hz for speaker PG and 6 Hz for speaker JP), even though this difference is only significant for the second speaker. Finally, only LR displays a clear boosting effect on both H1 and H2: as mentioned earlier, this speaker pronounced clash utterances with with a high degree of prominence. By contrast, the rest of the speakers display very small differences in H1 and H2 height in different stress conditions.

The data were submitted to an ANOVA with H1 and H2 as dependent variables and the stress factor as an independent variable (clash vs. no clash). The results are shown in Table 2. For 2 of the 4 speakers no significant effect was found of stress clash on H1. The two significant analyses reflect, on the one hand, the 'boosting' effect on H1 found in LR, and on the other, the H1 lowering pattern found in JP.

In general, H1 and H2 show strong stability effects in scaling across speakers, which can be considered as tonal production targets. The small tendency towards H1 lowering in clash conditions could be explained as an 'undershoot' strategy in situations which exert some articulatory pressure. Yet, this difference is not systematic across speakers and can be optionally overridden. Finally, it is especially significant to note that the height of the second H2 peak can be predicted quite successfully as a proportion of the preceding peak (1.24 for speaker LR, 1.11 for PG, 1.07 for PP and 1.14 for JP). These proportions are quite similar across speakers, indicating that this 'upstep' pattern works in a similar fashion to other downstep patterns described crosslinguistically (see Liberman and Pierrehumbert (1984) for English, Prieto et al. (1996) for Spanish, Arvaniti (2003) for Greek). This behavior can be interpreted as further evidence in favor of the view that the pitch levels of H points are extremely constant and that they can be predicted as a proportion of the preceding H peak, regardless of whether the peaks belong to a downstep or an upstep pattern.

3.1.3. L1 and L2 scaling

What is the effect of time pressure on L scaling? Recent work has reported that the speaker can opt to "undershoot" L points under time pressure conditions. Pierrehumbert, for example,

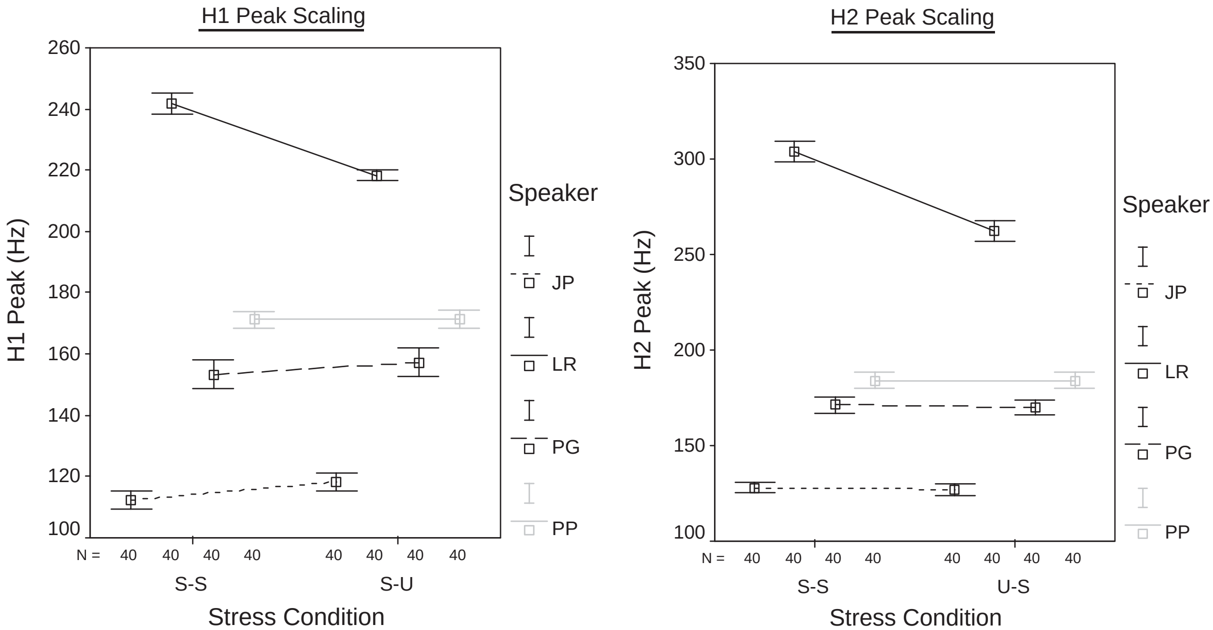

Fig. 4. Mean and standard error values (in Hz) of H1 and H2 peak values in two different conditions (S–S, S–U for H1 and S–S, U–S for H2) for the four speakers.

Table 2
One-way ANOVA results per speaker: H1 and H2 height as dependent variables and stress condition as an independent variable

	Main effect of clash on H1 height	Main effect of clash on H2 height
Speaker LR	$F_{1,78} = 150.925$; $p < 0.0005$	$F_{2,117} = 0.933$; $p = 0.396$
Speaker JP	$F_{1,78} = 8.121$; $p = 0.006$	$F_{2,117} = 1.097$; $p = 0.337$
Speaker PG	$F_{1,78} = 1.453$; $p = 0.232$	$F_{2,117} = 0.933$; $p = 0.396$
Speaker PP	$F_{1,78} = 0.007$; $p = 0.933$	$F_{2,117} = 0.060$; $p = 0.942$

suggests that the degree of dipping between peaks depends on the amount of time available, as f_0 "falls until it is time to start aiming for the next H* level." (Pierrehumbert, 1980, p. 71). Similar patterns have also been found in Japanese (Kubozono, 1988) and Spanish (Prieto, 1998). For Spanish, it has been found that L values decrease substantially in clash contexts (Prieto & Shih, 1995; Prieto, 1998; Face, 2002). Similarly, Arvaniti et al. (2000) have reported that one of the most common strategies of clash resolution in Greek is to undershoot or eliminate L2 and leave H1 and H2 fully realized.[4] Also, Kohler (1983) notes that "pitch peaks [in German] maintained their

[4]Yet, Arvaniti et al. (1998) also report that in Greek "its scaling (of the L values) is not affected by unaccented syllables intervening between accents". Thus, interestingly, in Greek the scaling of L values is not affected by the distance between two peaks but in clash situations L may undershoot.

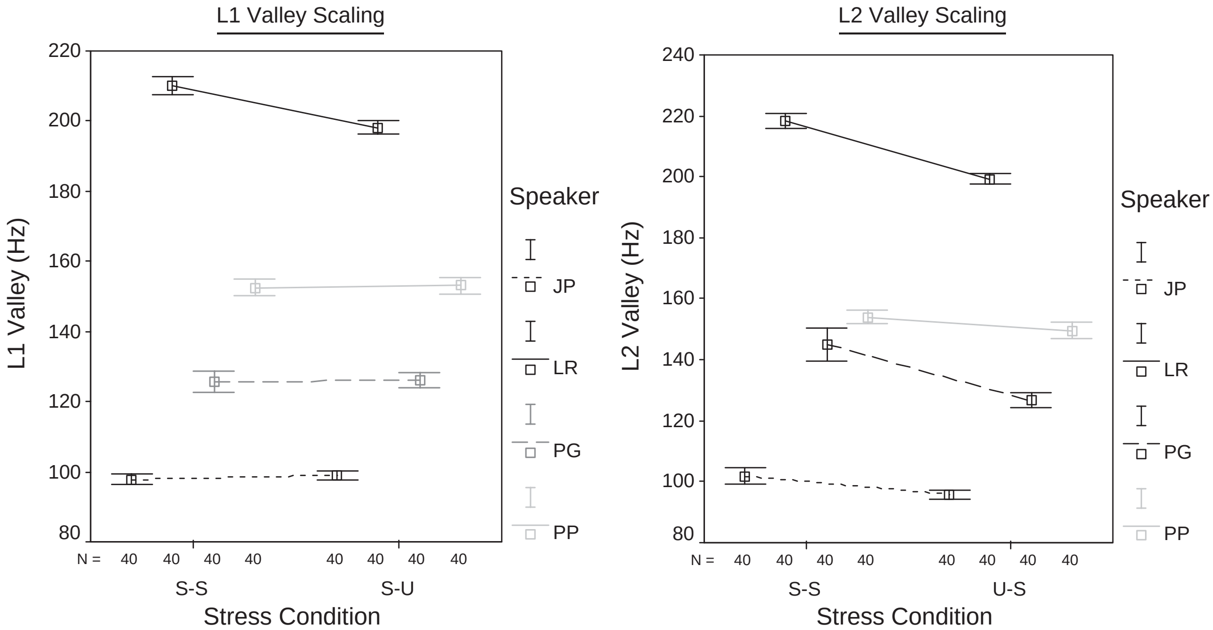

Fig. 5. Mean and standard error values (in Hz) of L1 and L2 in two different conditions (S–S, S–U for L1 and S–S, U–S for L2) for the four speakers.

target values and pitch valleys were raised, meaning that reaching f_0 peaks is more important than preserving the L values."

Fig. 5 displays the mean values (in Hz) of L1 and L2 points in three different conditions (S–S, U–S, S–U) for the four speakers. As in the preceding analysis, L1 is missing in the U–S condition and L2 in the S–U condition. Results clearly show a significant increase in f_0 of the second valley when there is no intervening syllable in between the accents for all 4 speakers, indicating that the falling gesture accommodates to the number of intervening syllables between accents. Two-tailed *t*-tests comparing L2 in these two conditions show that the two populations are distinct (at $p<0.05$) for the four subjects (see Appendix B, Table 1, for complete *t*-test results). By contrast, L1 values are near-constant in clash and nonclash environments (except for speaker LR, as expected).

The data display a strong correlation between L2 and the value of the preceding peak ($R^2 = 0.54$ for speaker PP, $R^2 = 0.63$ for JP and $R^2 = 0.82$ for PG). This tendency has been already reported in languages like English (Pierrehumbert, 1980), Japanese (Kubozono, 1988), and Spanish (Prieto & Shih, 1995; Prieto, 1998; Face, 2002).

The data were submitted to an ANOVA with L1 and L2 as dependent variables and stress clash as an independent variable (clash vs. no clash). The results shown in Table 3 reveal a clear main effect of stress clash on L2 for all subjects and no effect on L1 scaling (except for speaker LR).

In sum, the results of this section reveal a clear contrast between the scaling behavior of H and L values. While H maxima show clear stability effects and behave as production targets in clash conditions, L valleys are clearly undershot. Overall, the data indicate that H values are fully

Table 3
One-way ANOVA results per speaker: L1 and L2 height as dependent variables and stress condition as an independent variable

	Effects of clash on L1 height	Effects of clash on L2 height
Speaker LR	$F_{1,78} = 62.551$; $p < 0.0005$	$F_{2,117} = 103.74$; $p < 0.0005$
Speaker JP	$F_{1,78} = 1.652$; $p = 0.202$	$F_{2,117} = 6.934$; $p = 0.001$
Speaker PG	$F_{1,78} = 0.53$; $p = 0.818$	$F_{2,117} = 18.72$; $p < 0.0005$
Speaker PP	$F_{1,78} = 0.130$; $p = 0.719$	$F_{2,117} = 25.58$;$p < 0.0005$

produced and their surface realization takes precedence over the realization of L values. Still, we believe that the issue of precedence is not enough for claiming that Ls are not targets (i.e., they behave as ''interpolation points'' or ''phonetic transitions'' in between peaks, as Pierrehumbert claimed for English) because they consistently align near the onset of stressed syllables in nonclash situations.

3.2. *Pitch timing*

In this section the temporal behavior of H and L tonal targets is analyzed. Previous investigations of the alignment behavior of LH points in nonclash contexts have noted that, while H positions are more variable and depend on the right-hand prosodic context, L positions are consistently anchored with the onset of the accented syllable (see Caspers and van Heuven (1993) for Dutch; Arvaniti, Ladd, and Mennen (1998) for Greek; Ladd and Schepman (2003) for English; Prieto et al. (1995) for Spanish). This section examines whether or not clash contexts affect L alignment patterns and whether the first pitch accent gesture starts significantly earlier in clash environments. Previous work on pitch timing in tonal clash contexts in Spanish (Prieto & Shih, 1995; Face, 2002) has reported clear temporal effects on L2 (which is substantially moved to the right) and not so clear effects on L1: while Face (2002) reports that the only positions affected are the ones that are directly in contact between the two accents (that is, H1 and L2), Prieto and Shih (1995) found that L1 is also retracted (an average of 60 ms) in clash conditions.

3.2.1. *L1 and L2 alignment*

The two graphs in Fig. 6 plot the mean distance (in ms) from the start of L1 and L2 to the onset of their corresponding accented syllable (which we call *L1 and L2 delay*, respectively) in two stress conditions for the four speakers. The data show that, even though L1 is generally placed before the accented syllable in both clash and nonclash environments, it also tends to be placed significantly earlier in clash contexts (an average of –47.23 ms for speaker LR, –26.91 ms for JP, −2 ms for PG, −49.53 ms for PP). On the other hand, L2 delay values are radically different in the two clash conditions for the four speakers. In clash environments the second rising gesture starts significantly later in clash environments (an average of 28.65 ms later for speaker LR, 57.47 for JP, 68.68 for PG, and 65.65 ms for PP, roughly 1/3 of the syllable duration) than in nonclash contexts.

The data were submitted to an ANOVA with L1 and L2 delay as dependent variables and the stress factor as an independent variable (clash vs. no clash). The results in Table 4 reveal a significant effect of stress clash on L2 delay for the 4 speakers and on L1 for 3 of the 4 speakers:

Is the position of L1 and L2 related to syllable duration? The graphs in Fig. 7 plot L1 delay (or observed distance from the start of the first rise to the onset of the first accented syllable) as a function of the duration of the first syllable in two conditions (S–S and U–S, represented by two different plotting characters). Again, the data show a tendency for L1 points to be placed earlier in clash contexts (cf. the position of the black dots along the *y*-axis). Regarding the duration of S1, two of the speakers display a clear lengthening effect in clash environments (27 ms longer for JP and 26 ms for PG). Finally, the two regression lines (represented by gray dotted and solid black

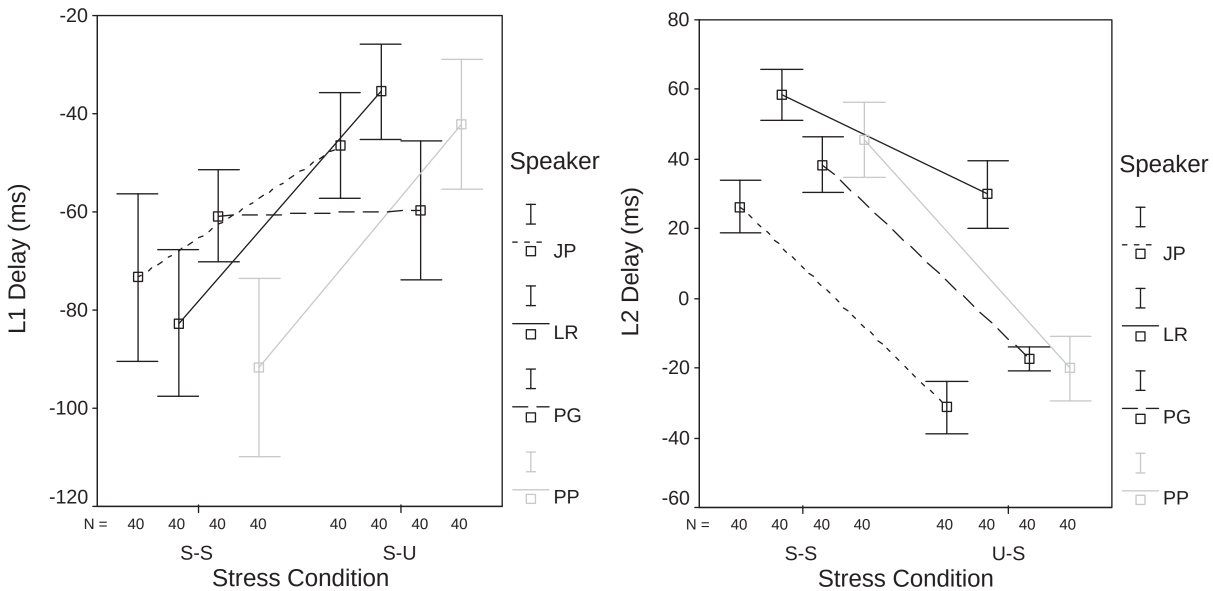

Fig. 6. Mean and standard error values (in ms) of L1 and L2 delay distances in two different conditions (S–S, S–U for L1 Delay and S–S, U–S for L2 Delay) for the four speakers.

Table 4
One-way ANOVA results per speaker

	Effects of clash on L1 delay	Effects of clash on L2 delay
Speaker LR	$F_{1,78} = 28.88$; $p<0.0005$	$F_{2,117} = 123.34$; $p<0.0005$
Speaker JP	$F_{1,78} = 7.231$; $p = 0.009$	$F_{2,117} = 446.36$; $p<0.0005$
Speaker PG	$F_{1,78} = 0.19$; $p = 0.892$	$F_{2,117} = 254.59$; $p<0.0005$
Speaker PP	$F_{1,78} = 19.95$; $p<0.0005$	$F_{2,117} = 368.9$; $p<0.0005$

L1 and L2 delay as dependent variables and stress condition as an independent variable.

lines) summarize the low correlation existing between L1 delay and S1 duration, which can be either positive or negative:

The graphs in Fig. 8 plot L2 delay (or observed distance in ms from the start of the second rise to the onset of the second accented syllable) as a function of the duration of the first syllable in two conditions (S–S and U–S, represented by two different plotting characters) for the four speakers. Clearly, points separated by the clash condition are located in separate regions along the

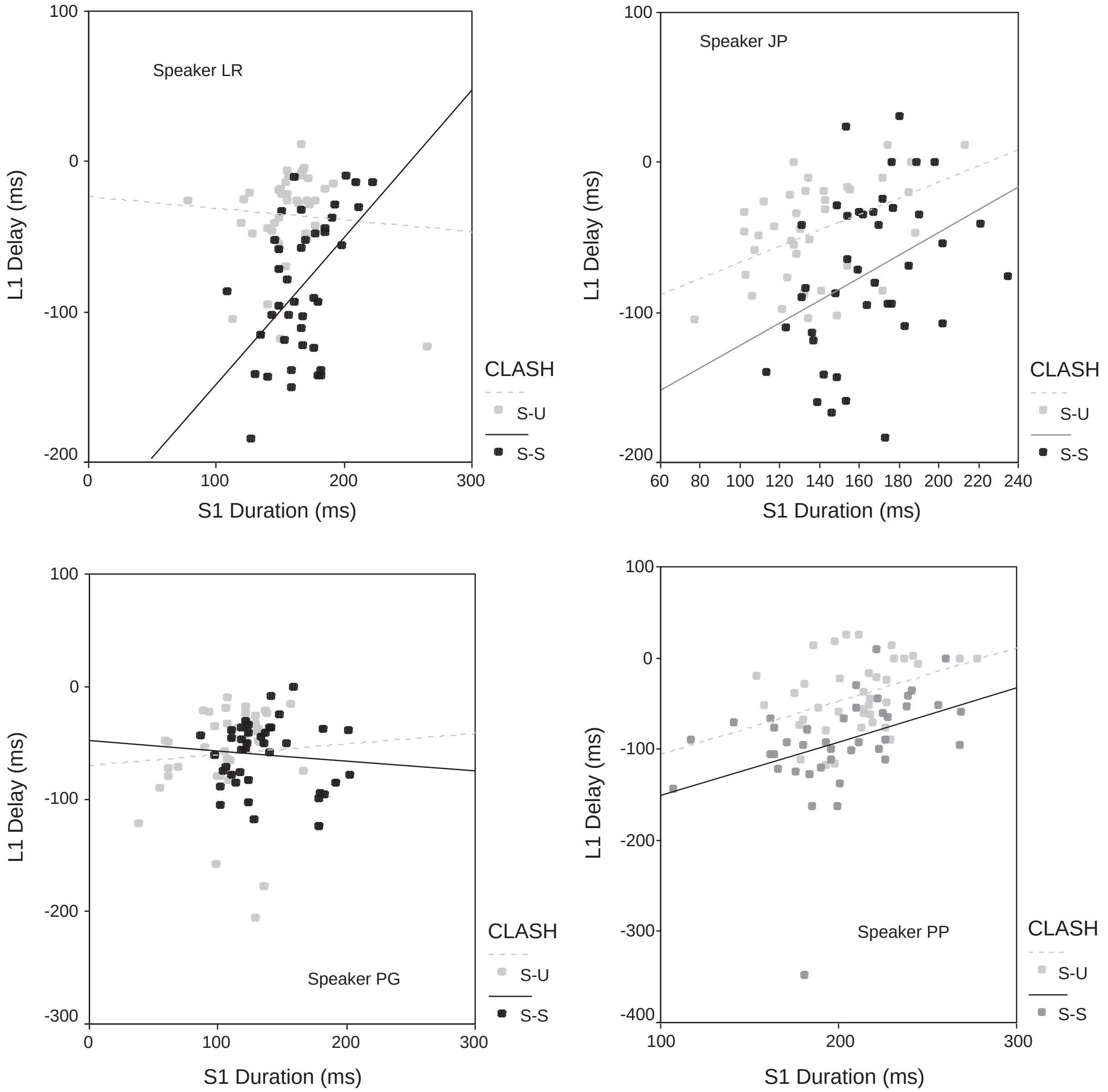

Fig. 7. L1 delay values (in ms) as a function of the duration of the first accented syllable in clash (S–S in black circles) vs. nonclash (S–U in gray circles) environments for speaker LR and JP (top panels), PG and PP (bottom panels).

time axis (*y*-axis). Two-tailed *t*-tests comparing L2 delay values across clash and nonclash environments were highly significant (at $p < 0.0005$) for the four speakers (see Appendix B, Table 2, for complete *t*-test results). Again, the two regression lines (represented by dotted and solid lines) summarize the low correlations between L2 delay and duration of the preceding syllable

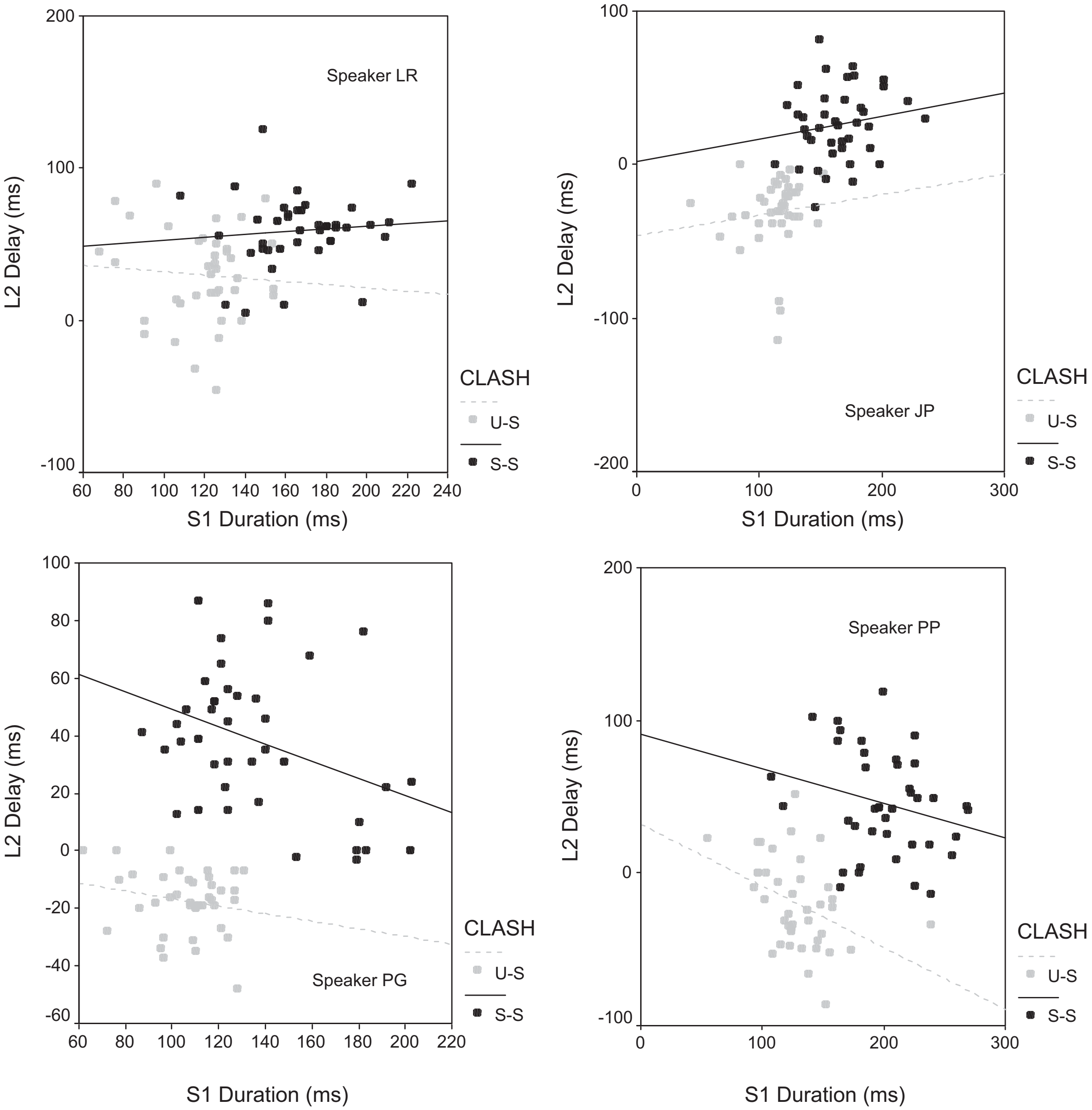

Fig. 8. L2 delay values (in ms) as a function of the duration of the first accented syllable in clash/nonclash environments for speaker LR and JP (top panels), PG and PP (bottom panels).

Table 5
Correlation results per speaker: L1 delay/S1 duration and L2 delay/S1 duration

	Correlation L1 delay/S1 dur	Correlation L2 delay/S1 dur
Speaker LR	$R^2 = 18.1$; $p = 0.109$	$R^2 = 24.4$; $p = 0.07$
Speaker JP	$R^2 = 20.9$; $p = 0.63$	$R^2 = 26.0$; $p = 0.04$
Speaker PG	$R^2 = -3.00$; $p = 0.978$	$R^2 = -0.080$; $p = 0.387$
Speaker PP	$R^2 = 40.1$; $p < 0.0005$	$R^2 = 54.01$; $p < 0.0005$

(which is positive for 2 speakers and negative for the other two): that is, the shorter the preceding syllable, the closer L2 gets to the onset, or the other way around. This is most likely indicating that the speaker can either choose to lengthen the first target syllable (and have more time available) or rather delay the start of the second gesture independently.

To summarize, Table 5 shows the results of the correlation analyses between L1/L2 delay and S1 durations. In the first column, the low Pearson correlation coefficients and the non-significant p values (for 3 of the 4 speakers) reveal a rather weak relationship between the two factors (only significant for speaker PP). On the contrary, results in the second column reveal a significant correlation (for 3 of the 4 speakers) between L2 delay and duration of the first syllable:

Results on the potential effects of S2 duration on L1/L2 delay data revealed no significant correlations or effects between the two measures.

3.2.2. *H1 and H2 alignment*

Do tonal clash contexts affect H alignment patterns? The two graphs in Fig. 9 plot the distance in time from H1 and from H2 to the offset of the accented syllable (see left and right panels respectively) in clash vs. nonclash contexts for the four speakers.[5] As expected, the data display a clear effect of clash on H1 alignment: H1 distance to the offset of the syllable (*H1 Dist to S2*) increases drastically in clash environments. In our data, H1 is generally reached before the end of the syllable in both clash and nonclash environments: the fact that peaks are not delayed in nonclash environments (the typical situation of rising prenuclear peaks in Catalan) is probably due to the fact that such sentences have only one unstressed syllable in between accents. Two-tailed t-tests comparing the values of H1 distance to the offset were significant (at $p < 0.05$) for the four speakers (see Appendix B, Table 3). By contrast, H2 is always roughly aligned with the end of the second syllable (only speaker PP 'displaced' the second peak, which was located an average of 65 ms into the second syllable).

The four graphs in Fig. 10 plot the distance in time from H1 to the offset of the accented syllable (H1 distance to S2) as a function of the syllable duration in clash vs. nonclash contexts.

[5]In this section, I chose to plot the measure H Distance to Offset (or distance between the peak to the *offset* of the accented syllable) because it clearly shows the retraction effects on H1 timing. By contrast, the H Peak Delay measure (or distance between the peak to the *onset* of the accented syllable) does not show this effect as clearly because of compound duration effects.

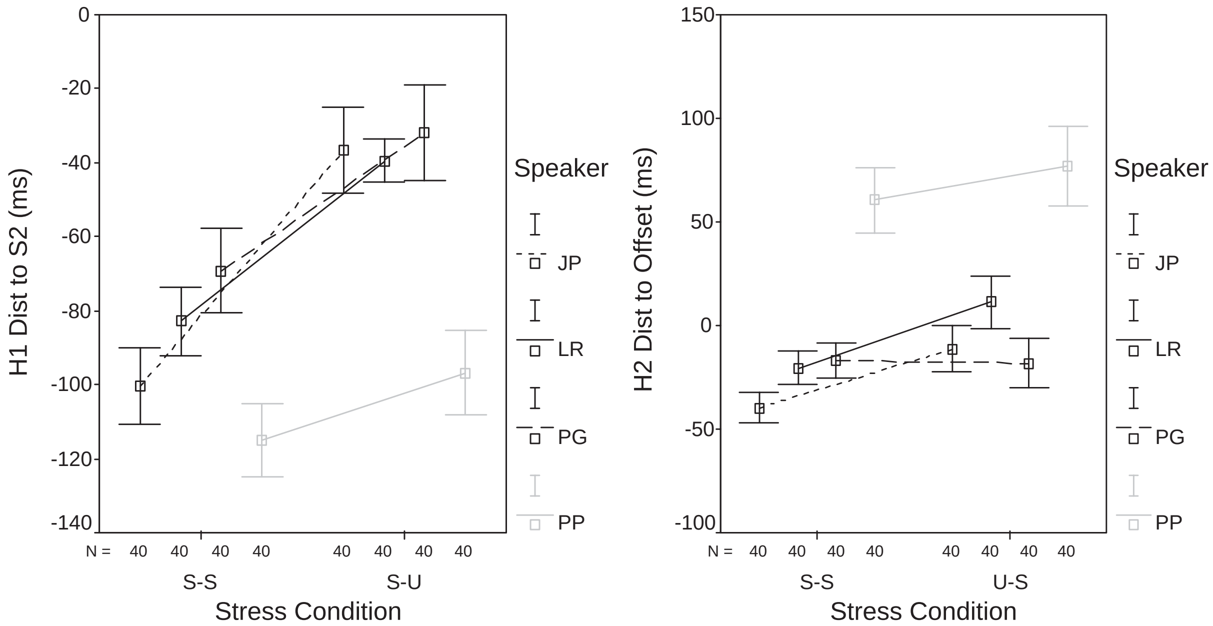

Fig. 9. Mean and standard error values (in ms) of H1 distance to offset (left) and H2 distance to offset (right) in clash and nonclash conditions for the four speakers.

Distance to the offset (*H1Dist2*) increases substantially in clash environments (remember that *t*-tests comparing the values of H1 distance to the offset were significant at $p < 0.001$) for the four speakers. The regression lines in the graphs reveal a negative correlation between the distance between H and the end of the syllable and the duration of the accented syllable. That is, the longer the syllable, the shorter the distance from the peak to the end of the syllable—correlation coefficients are $R^2 = -0.45$ for speaker LR, $R^2 = -0.54$ for JP, $R^2 = -0.68$ for PG, and $R^2 = -0.32$ for PP, statistically significant for all speakers.

Table 6 shows the correlation results per speaker between S1 Duration/H1 delay and S1 duration/H1 distance to the offset. While no significant correlation is found between H1 delay and syllable duration (except for speaker PP), a significant negative correlation is found between S1 duration and distance from H1 to offset for all speakers. That is, the shorter the syllable, the more retracted the peak is. This result seems to disconfirm Silverman and Pierrehumbert's repulsion hypothesis for English: "when a syllable is lengthened by an upcoming word boundary or stress clash, then the peak is moved to the left. At the same time, the syllable is lengthened in such a way that its right-hand edge is moved to the right, relative to the location of the accent peak." (Silverman & Pierrehumbert, 1990, p. 94).

Finally, in contrast with the behavior of H1, no clear effects of pitch clash were found on H2 alignment. In our data, the f_0 peak of the second accent was roughly synchronized with the offset of the syllable, regardless of the clash condition, confirming previous results for Spanish (Prieto & Shih, 1995; Face, 2002).

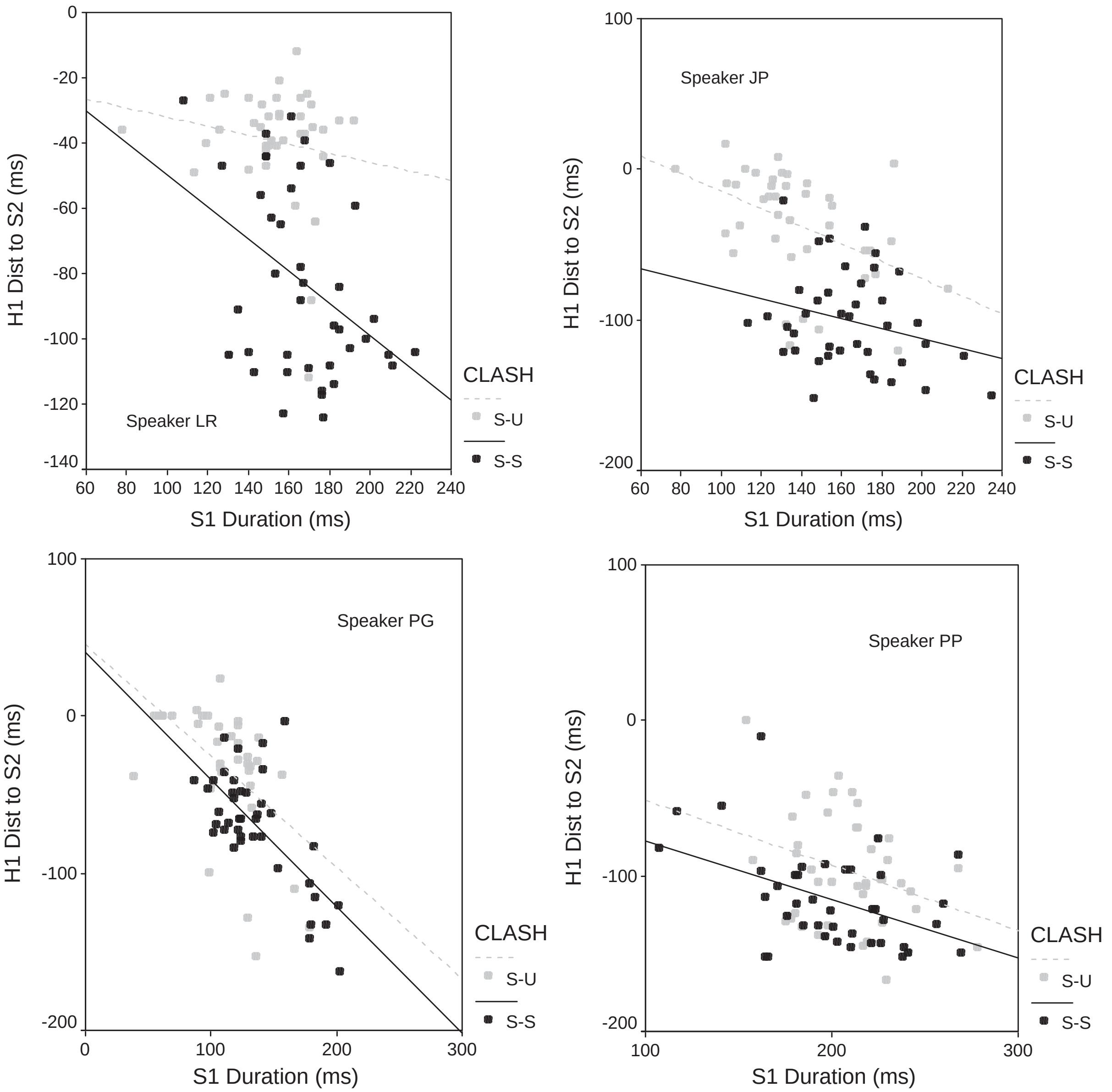

Fig. 10. Mean values (in ms) of the distance between the first peak to the offset of the syllable (bottom panels) as a function of the duration of the syllable in two conditions (S–S, S–U) for all four speakers.

4. Compression strategies

The data in the previous section show that the production of two consecutive pitch accents involves a drastic reorganization of the timing of the accents involved: the first accent gesture is

Table 6
Correlation results per speaker between H1 delay and S1 duration and H1 distance to the offset and S1 duration

	Correlation H1delay/S1duration	Correlation H1Dist-to-Offset/S1duration
Speaker LR	$R^2 = 19.2$; $p = 0.88$	$R^2 = -41.6$; $p < 0.0005$
Speaker JP	$R^2 = 12.2$; $p = 0.280$	$R^2 = -54.4$; $p < 0.0005$
Speaker PG	$R^2 = -15.3$; $p = 0.176$	$R^2 = -68.2$; $p < 0.0005$
Speaker PP	$R^2 = 54.8$; $p < 0.0005$	$R^2 = -32.6$; $p = 0.003$

placed significantly earlier with respect to the first syllable offset and the second valley moves to the right. On the other hand, clear stability effects are found on the scaling of H peaks. This section examines the compression strategies used by speakers to reach such intonational targets. Are the syllables involved consistently lengthened? Is the time of the rise gesture shortened, or is it kept constant? Is the velocity of the rising gestures significantly faster?

4.1. Effects of clash on syllable durations

In the preceding sections it has been noted that clash environments do not have clear effects on syllable duration. The graphs in Fig. 11 plot the mean duration (in ms) of the first syllable (S1) and the second syllable (S2) (see left and right panels, respectively) in clash vs. nonclash contexts for the four speakers. As we see from the graphs, S1 is generally shorter in clash environments, except for speaker LR. By contrast, S2 is longer in clash environments for all speakers. We thus see a difference between the behavior of S1 (which is variable depending on the speaker) and S2, which is consistently lengthened. Even though the results are only significant for 2 speakers, this lengthening can be regarded as a very clear trend in the data. As we will see in the next section, the negative correlation between rise time and duration (statistically significant for three of the four speakers) is revealing that, in clash situations, S2 is lengthened and, at the same time, the rise time of the second accent is shortened in order to be able to fit the accent gesture.

Table 7 shows the one-way ANOVA results performed on the data. The comparisons are only significant for 2 of the 4 speakers in each case, which reveals not such a strong effect of clash on duration patterns.

4.2. Effects of clash on rise time

This section investigates the effect of tonal clash on absolute *rise time* (or temporal distance from the peak to the previous valley) of the rising accents. Previous investigations have reported a tendency to reduce time rises and increase velocity of the rising gestures (cf. Kohler, 1983; Caspers & van Heuven, 1993; Prieto et al., 1995). As Caspers and van Heuven (1993, p. 167) point out in their study of Dutch accent-lending pitch rises and falls, that "when multiple pitch movements have to be executed within a short time span, both speakers shorten and steepen the rise." The graphs in Fig. 12 plot the distance in time from H1 and from H2 to the offset of the accented syllable (see left and right panels, respectively) in clash vs. nonclash contexts for the four speakers.

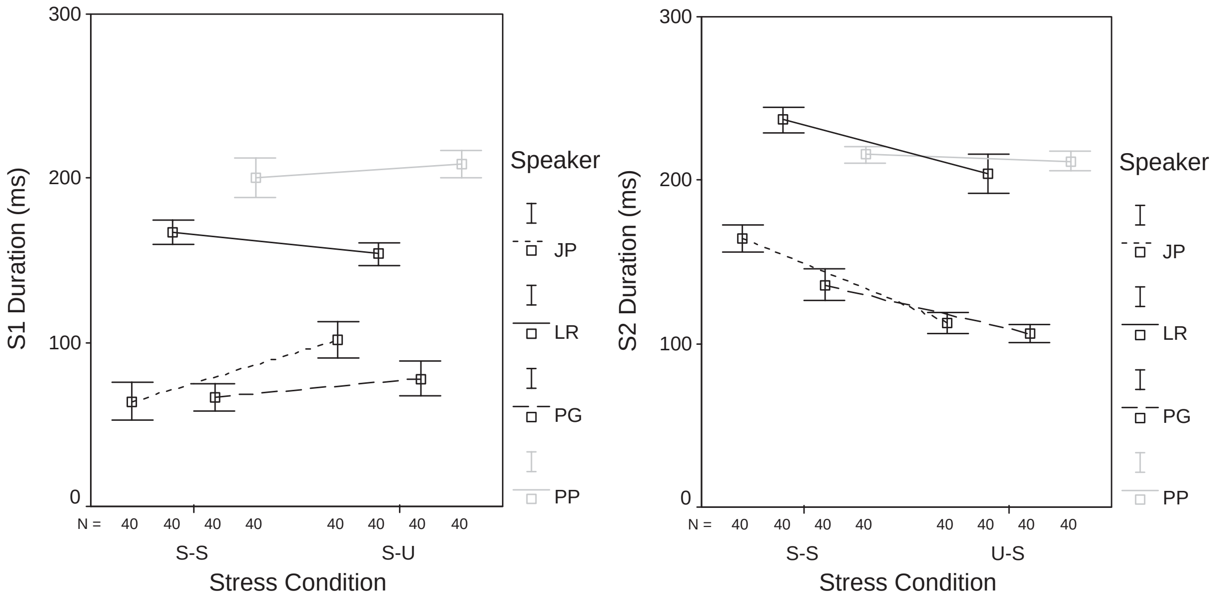

Fig. 11. Mean duration and standard error values (in ms) of the duration of the first syllable (left) and of the second syllable (right) in clash and nonclash conditions for the four speakers.

Table 7
One-way ANOVA results per speaker

	Effects of clash on S1 duration	Effects of clash on S2 duration
Speaker LR	$F_{1,78} = 6.68$; $p = 0.12$	$F_{2,117} = 21.27$; $p < 0.0005$
Speaker JP	$F_{1,78} = 18.48$; $p < 0.0005$	$F_{2,117} = 115.89$; $p < 0.0005$
Speaker PG	$F_{1,78} = 15.035$; $p < 0.0005$	$F_{2,117} = 0.036$; $p = 0.850$
Speaker PP	$F_{1,78} = 1.449$; $p = 0.232$	$F_{2,117} = 1.034$; $p = 0.312$

S1 and S2 duration as dependent variables and stress condition as an independent variable.

Indeed, there is a tendency to reduce the rise time in pitch clash situations while the duration of the syllable is equal or longer (compare the rise times of the second accent, which are shorter, to its duration, which are longer in clash environments). This effect is more drastic for the rise time of the second accent (which is an average of 85 ms shorter in clash environments for speaker LR, 27 ms for speaker JP, 26 ms for speaker PG, and 70 ms for speaker PP). Two-tailed *t*-tests are significant (at $p < 0.007$) for the rise time of the second accent for all speakers (see Appendix B, Table 6). With respect to rise times for the first gesture, it seems that the intonational movement 'adapts' to the time available: that is, they are short in shorter syllables (PG and JP) and long in longer syllables (LR and PP)—cf. also Fig. 11.

The data were submitted to an ANOVA with rise times of the first and second accents as dependent variables and the stress factor as an independent variable (clash vs. no clash). The

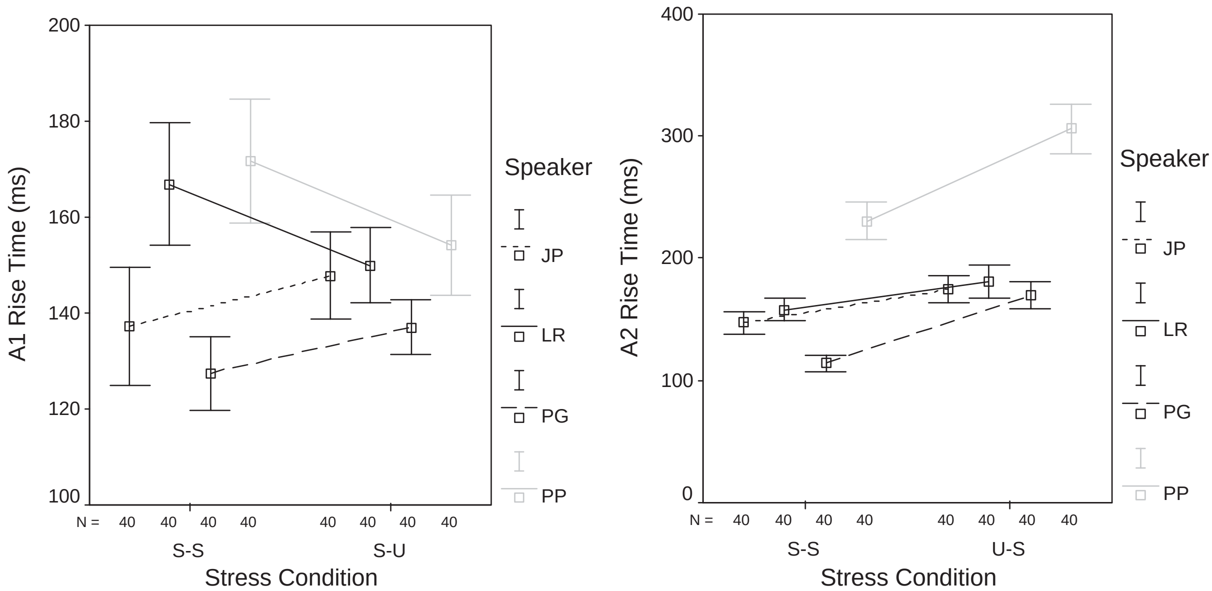

Fig. 12. Mean duration and standard error values (in ms) of the duration of the first syllable (left) and of the second syllable (right) in clash and nonclash conditions for all four speakers.

Table 8
One-way ANOVA results per speaker

	Main effect of clash on rise time 1	Main effect of clash on rise time 2
Speaker LR	$F_{1,78} = 5.276$; $p = 0.24$	$F_{2,117} = 48.251$; $p < 0.0005$
Speaker JP	$F_{2,117} = 1.968$; $p = 0.165$	$F_{2,117} = 29.187$; $p < 0.0005$
Speaker PG	$F_{2,117} = 4.090$; $p = 0.047$	$F_{2,117} = 52.585$; $p < 0.0005$
Speaker PP	$F_{2,117} = 4.521$; $p = 0.037$	$F_{2,117} = 18.994$; $p < 0.0005$

Rise time of the first and second accent as dependent variables and stress condition as an independent variable.

results shown in Table 8 reveal a significant main effect of stress clash on rise time of the second accent for all subjects and on the first accent accent for 2 of the 4 speakers.

The correlation results given in Table 9 demonstrate that the negative correlation between rise time and duration is statistically significant for three of the four speakers.

In sum, the results of our data replicate earlier findings by Kohler (1983) for German, Caspers (1994) for Dutch, Prieto et al. (1995) for Spanish and Arvaniti et al. (2000) for Greek and show that there is no fixed duration for the rise and that tonal crowding can significantly reduce the duration of the rising gestures. As Caspers (1994, p. 104) remarks, ‘‘generally speaking, the duration of the rise is shortened and its f_0 slope is steepened when competing falls are present, reflecting the ascending order of pressure assumed. When multiple pitch movements have to be executed within a short time span, it appears that the same compression strategy is used for the accent-lending pitch rise as in fast speech.’’

Table 9
Correlation results per speaker between rise time and S2 duration

	Correlation rise time2/S2 duration
Speaker LR	$R^2 = -50.7$; $p < 0.0005$
Speaker JP	$R^2 = -21.4$; $p = 0.019$
Speaker PG	$R^2 = -27.6$; $p = 0.02$
Speaker PP	$R^2 = -0.1$; $p = 99.4$

Table 10
One-way ANOVA results per speaker: rise time of the first and second accent as dependent variables and stress condition as an independent variable

	Main effect of clash on rise vel 1	Main effect of clash on rise vel 2
Speaker LR	$F_{1,78} = 32.132$; $p < 0.0005$	$F_{2,117} = 80.195$; $p < 0.0005$
Speaker JP	$F_{1,78} = 4.286$; $p = 0.42$	$F_{2,117} = 5.071$; $p = 0.008$
Speaker PG	$F_{1,78} = 0.566$; $p = 0.454$	$F_{2,117} = 26.977$; $p < 0.0005$
Speaker PP	$F_{1,78} = 1.937$; $p = 0.168$	$F_{2,117} = 9.232$; $p < 0.0005$

4.3. *Effects of clash on the velocity of the rise*

The data in the previous section has shown that the time allocated for the second accent gesture is significantly reduced in clash environments. Following previous work by Caspers (1994) for Dutch and by Prieto et al. (1995) for Spanish, we should therefore expect a significant increase in velocity of this gesture. As Prieto et al. (1995, p. 450) point out, "when peak location is retracted, absolute rise time shortens, and it is the velocity of the accent rise that accommodates to the time constraints imposed. That is, when the speaker is forced to use less time in the rise, f_0 velocity increases". A one-way ANOVA was run on the data, and the results are displayed in Table 10. For all speakers, a statistically significant effect of tonal clash is found on the velocity of the second rise. By contrast, there is no effect of the stress condition on the velocity of the first rise (except for speaker LR, which pronounced accents emphatically in the clash condition). This result agrees with the fact that the effect of tonal clash on the duration of the first rise was not consistent.

Rise velocity was computed as the distance in Hz from valley to the following peak, divided by the temporal distance between the two. The graphs in Fig. 13 plot the velocity of the rise of the second accent as a function of the rise time allocated for that accent in clash vs. nonclash contexts for the four speakers. The data 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 (that is why the clash symbols tend to appear higher in the graphs and more to the left than nonclash symbols). Finally, the regression lines evidence a negative correlation (not totally systematic) between velocity and rise time in the two conditions: that is, in general, the faster the rising movement, the shorter the rise time allocated for that gesture (thus, the special feature

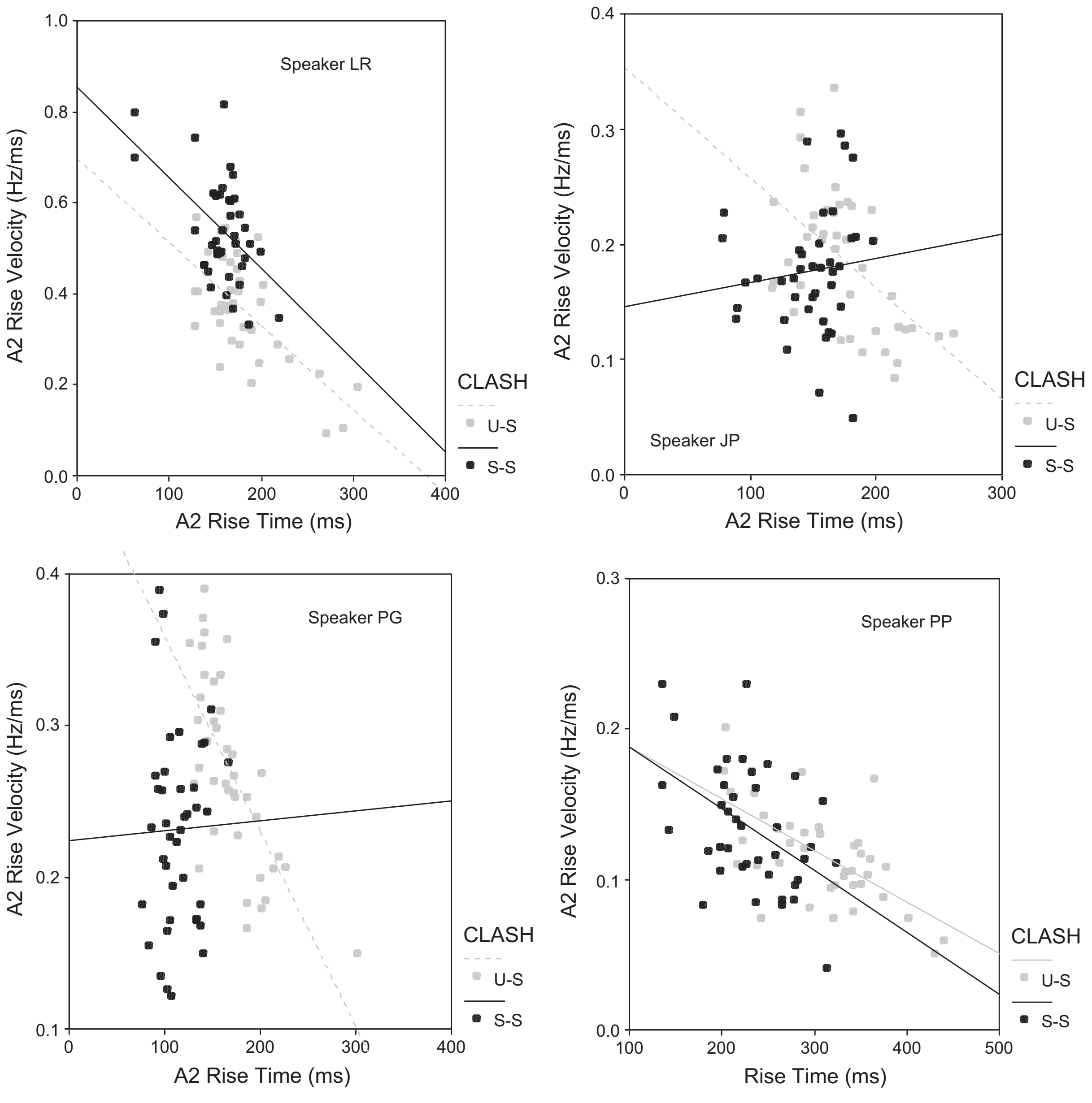

Fig. 13. Mean velocity of the second accent gesture (in Hz/ms) as a function of the rise time of this accent (in ms) in clash/nonclash environments for all four speakers.

characterizing clash environments is the significant compression of the rise time of the second accent gesture).

Table 11 shows correlation results per speaker between rise time and velocity of the rising gesture. In seven out of eight cases, there is a statistically significant negative correlation between the two factors.

Table 11
Correlation results per speaker between rise time and syllable duration

	Correlation rise time1/vel1	Correlation rise time2/vel2
Speaker LR	$R^2 = -28.0$; $p = 0.01$	$R^2 = -71.8$; $p < 0.0005$
Speaker JP	$R^2 = -10.8$; $p = 0.340$	$R^2 = -26$; $p = 0.04$
Speaker PG	$R^2 = -39.5$; $p < 0.0005$	$R^2 = -42.6$; $p < 0.0005$
Speaker PP	$R^2 = -51.9$; $p < 0.0005$	$R^2 = -49.7$; $p < 0.0005$

Thus, speakers make the effort of reaching a certain f_0 target level. In the case of the second accent gesture, they do so both by reducing the rise time allocated for the gesture and by increasing the velocity of the rise. This confirms previous results regarding the relationship between these two tonal features. As Prieto et al. (1995, p. 447) point out, ''Environments that trigger peak retraction also trigger a greater steepness in the rising slope, even when excursions are very low, as in the case of speaker AH. Therefore, the steepness of the rising slope shows a pattern that is inverse to the peak delay pattern for the same speaker. Thus, in the data, velocity of the f_0 rise accommodates to the time allocated for the rising gesture, so that it can reach a similar f_0 peak.''

With regard to the production of the first accent gesture, the data seem to suggest that speakers are not very 'pressed' to produce a compressed movement. Speakers can optionally use different strategies in producing the first rising movement: (a) start the rising gesture earlier (this strategy is used by all speakers but speaker JP); (b) reduce the rise time of the first accent gesture (speakers LP and PG use this strategy and increase the velocity of the movement even though this effect is not statistically significant for any of the speakers).

4.4. *Effects of clash on pitch range*

Summarizing the influence of time pressure on rise time and velocity in our data, the rising movement of the second accent is shortened and steepened. This tendency has also been found in other languages as a general compression strategy in time pressure situations (cf. Kohler (1983) for German; Caspers and van Heuven (1993) for Dutch; Prieto et al. (1995) for Spanish). As Caspers and van Heuven (1993, p. 166) ''the pitch movements become shorter, with a tendency for the f_0 slope to steepen, which suggests time compression (rather than frequency compression) of pitch movements''. Yet, is pitch range also kept constant in the different stress conditions in our data? The graph in Fig. 14 plots the mean excursion size of the second pitch accent (in Hz) in clash vs. nonclash contexts for the four speakers. Generally speaking, there is a reduction of the range in clash environments—brought about by an elevation in pitch of the L2 start of the rise. The clear exception is speaker LR, who produced accents in clash conditions emphatically.

The results in Fig. 14 thus reveal that pitch range is *not* really kept constant in time-pressure situations.

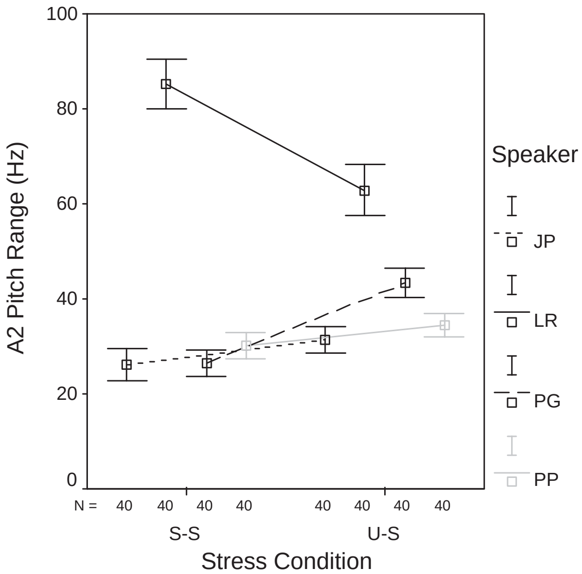

Fig. 14. Mean duration and standard error values (in ms) of the pitch range of the second rising gesture in clash and nonclash conditions for all four speakers.

5. Conclusion

The data shown in this article have revealed consistent differences in the time dimension between accent gestures in clash vs. nonclash contexts. Clash contexts trigger a drastic temporal reorganization of the accents involved, namely, the first valley and peak are placed significantly earlier with respect to the first syllable onset and the second valley significantly later with respect to the second syllable onset. Thus, even though in unmarked prosodic environments the onset of the rise is anchored to the onset of the accented syllable, a fixed synchronization point was not found for L points in clash contexts. This confirms the results of recent experiments on tonal clash (Prieto and Shih (1995), Face (2002) for Spanish, Xu (1999) for Mandarin Chinese) which 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. Furthermore, the data confirm that the segmental anchoring effects found in time-pressure environments, i.e., under increased rate of speech (Ladd et al. (1999) for English and Xu (1998) for Chinese; see also Ladd (2003)), are not found in pitch pressure environments. This provides evidence in favor of distinguishing pitch pressure environments (i.e., proximity to accents and boundary tones) from time-pressure environments (i.e., shorter segment duration from increase in speech rate). In other words, the lack of anchoring effects in pitch clash contexts suggests that different compression strategies are being used in the two environments (among other factors, duration tends to increase in the former and decrease in the latter).

Significantly, no clear effects of pitch clash were found on the scaling of H peaks. Their value was found to be rather invariant across repetitions of utterances by the same speaker, showing clear stability effects in the H scaling dimension and confirming the implicit prediction of the pitch target view of intonational production advocated by the autosegmental-metrical model. In our data, the contrast between the scaling behavior of H and L values seem to indicate that H points function as production real targets: the speaker makes the effort to reach H values despite the time constraints; by contrast, the speaker generally opts to "undershoot" L targets in pitch pressure conditions. This tendency has been already reported in the presence of competing rising accents in languages like English (Pierrehumbert, 1980), Japanese (Kubozono, 1988), Spanish (Prieto & Shih, 1995; Prieto, 1998; Face, 2002) and Greek (Arvaniti et al., 2000).

It has been shown that tonal clash situations are resolved not by adjusting H scaling but by modifying the timing of the accent gestures, both with respect to the segmental content and in terms of velocity of the accent gestures. Speakers make the effort to reach f_0 target peaks, and do so by using the following 'compression' and time reorganization strategies: (a) starting the first rising gesture significantly earlier (in our data, 3 out of 4 speakers aligned L1 earlier in clash conditions); (b) raising L2 points in between the two H's; (c) shortening the duration of the second rise and increasing its velocity. Caspers (1994) also pointed out that what is really crucial from a perception point of view is the constancy of H height and the differences in pitch range between the valley and the peak, and not the velocity of the rise. Indeed, our data also demonstrate the stability of H target peaks and the adapting behavior of rise velocity (which increases when pitch movements have to be executed within a short time span).

Acknowledgements

My thanks go to Mariapaola D'Imperio, Gorka Elordieta, Timothy Face, Sónia Frota, Carlos Gussenhoven, José Ignacio Hualde, Bob Ladd, Eugenio Martínez Celdrán, Daniel Recasens and an anonymous reviewer for helpful comments and fruitful discussion of the materials presented in this paper. I also want to thank the editor of *Journal of Phonetics*, Gerry Docherty, for his accurate comments and help during the revision period. This research was funded by grants 2002XT-00032 and 2001SGR 00150 from the Generalitat de Catalunya and BFF2003-06590 and BFF2003-09453-C02-C02 from the Ministry of Science and Technology of Spain.

Appendix A

El molí net no li agrada
'He/she does not like the clean mill'
El molinet no li agrada
'He/she does not like the little mill'
El molí netet no li agrada
'He/she does not like the clean.dim mill'

El camí net no s'acaba
'The clean path does not end'

El llumí net no s'encén
'The clean match does not light'
El lluminet no s'encén
'The little match does not light'
El llumí netet no s'encén
'The clean.dim match does not light'

El veí net no saluda
'The clean neighbour does not greet'

El caminet no s'acaba
'The small path does not end'
El camí netet no s'acaba
'The clean.dim path does not end'

El meló net no li agrada
'He/she does not like the clean melon'
El melonet no li agrada
'He/she does not like the small melon'
El meló netet no li agrada
'He/she does not like the clean.dim melon'

El colomí net no volava
'The clean pigeon did not fly'
El colominet no volava
'The small pigeon did not fly'
El colomí netet no volava
'The clean.dim pigeon did not fly'

El veïnet no saluda
'The clean neighbour does not greet'
El veí netet no saluda
'The clean.dim neighbour does not greet'

El remolí net no es desfà
'The clean whirlwind does not break up'
El remolinet no es desfà
'The small whirlwind does not break up'
El remolí netet no es desfà
'The clean.dim whirlwind does not break up'

El violí net no sonava
'The clean violin did not play'
El violinet no sonava
'The small violin did not play'
El violí netet no sonava
'The clean.dim violin did not play'

Appendix B

Tables I–IV.

Table I
t-test results per speaker: L2 height as a dependent variable and stress condition (clash vs. no clash) as an independent variable

	Effects of clash on L2 height
Speaker LR	$t = 4.04$, df = 78, $p < 0.0005$
Speaker JP	$t = 12.91$, df = 78, $p < 0.0005$
Speaker PG	$t = 6.35$, df = 78, $p < 0.0005$
Speaker PP	$t = 2.42$, df = 78, $p < 0.05$

Table II
t-test results per speaker: L2 delay as a dependent variable and stress condition (clash vs. no clash) as an independent variable

	Effects of clash on L2 delay
Speaker LR	$t = 4.47$, df = 78, $p < 0.0005$
Speaker JP	$t = 10.86$, df = 78, $p < 0.0005$
Speaker PG	$t = 12.98$, df = 78, $p < 0.0005$
Speaker PP	$t = 9.28$, df = 78, $p < 0.0005$

Table III
t-test results per speaker: H1 distance to the offset of accented syllable as a dependent variable and stress condition (clash vs. no clash) as an independent variable

	Effects of clash on H1 distance to offset
Speaker LR	$t = -8.03$, df = 78, $p < 0.0005$
Speaker JP	$t = -8.31$, df = 78, $p < 0.0005$
Speaker PG	$t = -4.39$, df = 78, $p < 0.0005$
Speaker PP	$t = -2.47$, df = 78, $p = 0.01$

Table IV
t-test results per speaker: Rise tiem of accent 2 as a dependent variable and stress condition (clash vs. no clash) as an independent variable

	Effects of clash on H1 distance to offset
Speaker LR	$t = -2.85$, df = 78, $p = 0.006$
Speaker JP	$t = -3.73$, df = 78, $p < 0.0005$
Speaker PG	$t = -8.79$, df = 78, $p < 0.0005$
Speaker PP	$t = -6.08$, df = 78, $p < 0.0005$

References

Arvaniti, A. (2003). Peak scaling in Greek and the role of declination. In M.J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the XVth international congress of phonetic sciences* (Vol. 2, pp. 2269–2272), Barcelona.

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*, *26*, 3–25.

Arvaniti, A., Ladd, D. R., & Mennen, I. (2000). What is a starred tone? Evidence from Greek. In M. Broe, & J. Pierrehumbert (Eds.), *Papers in laboratory phonology V: acquisition and the lexicon* (pp. 119–131). Cambridge: Cambridge University Press.

Bonet, E., & Lloret, M.-R. (1998). *Fonologia catalana*. Barcelona: Ariel.

Bruce, G. (1977). *Swedish word accents in sentence perspective*. Lund: Gleerup.

Caspers, J. (1994). *Pitch movements under time pressure*. Dordrecht: ICG Printing.

Caspers, J., & van Heuven, V. J. (1993). Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica*, *50*, 161–171.

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

Estebas-Vilaplana, E. (2000). *The use and realisation of accentual focus in Central Catalan*. Ph.D. dissertation, University College, London, 2000.

Estebas-Vilaplana, E. (2003). Catalan pre-nuclear accents: evidence for word-edge tones. In M.J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the XVth international congress of phonetic sciences* (Vol. 2, pp. 1779–1782), Barcelona.

Estebas-Vilaplana, E., & Prieto, P. (in prep). Prenuclear H peaks in Catalan: word boundary tones or trailing tones? In preparation.

Face, T. L. (2002). When push comes to shove: tonal crowding in Madrid Spanish. *The Linguistic Association of Korea Journal*, *10*, 77–100.

Kohler, K. (1983). f_0 in speech timing. *Arbeitsber, Institute of Phonetics* 20, Kiel University (pp. 55–98).

Kubozono, H. (1988). *The organization of Japanese prosody*. Ph.D. dissertation, University of Edinburgh.

Ladd, D. R. (2003). Phonological conditioning of f_0 target alignment. In M.J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the XVth international congress of phonetic sciences* (Vol. 1, pp. 249–252), Barcelona.

Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (1999). Constant 'segmental' anchoring of f_0 movements under changes in speech rate. *Journal of the Acoustical Society of America*, *106*, 1543–1554.

Ladd, D. R., & Schepman, A. (2003). Sagging transitions between high pitch accents in English: experimental evidence. *Journal of Phonetics*, *31*, 81–112.

Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff, & R. T. Oehrle (Eds.), *Language sound structure studies in phonology presented to Morris Halle* (pp. 157–233). Cambridge, MA: MIT Press.

Oliva, S. (1992). *La mètrica i el ritme de la prosa*. Barcelona: Quaderns Crema.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable Nuclei in English. *Journal of the Acoustical Society of America*, *32*(6), 693–703.

Pierrehumbert, J. (1980). *The phonetics and phonology of English intonation*. Ph.D. dissertation, Massachusetts Institute of Technology.

Prieto, P. (1998). The scaling of the L tone line in Spanish downstepping contours. *Journal of Phonetics*, *26*, 261–282.

Prieto, P. (2002). Entonació. In J. Solà, M.-R. Lloret, J. Mascaró, & M. P. Saldanya (Eds.), *Gramàtica del català contemporani* (pp. 393–462). Barcelona: Editorial Empúries.

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.

Prieto, P., van Santen, J., & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, *23*, 429–451.

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, pp. 1307–1310).

Prieto, P., Shih, C., & Nibert, H. (1996). Pitch downtrend in Spanish. *Journal of Phonetics*, *24*, 445–473.

Selkirk, L. (1984). *Phonology and syntax. The relation between sound and structure*. Cambridge, MA: MIT Press.

Silverman, K., & Pierrehumbert, J. (1990). The timing of prenuclear high accents in English. In J. Kingston, & M. E. Beckman (Eds.), *Papers in laboratory phonology I. Between the grammar and physics of speech* (pp. 72–106). Cambridge: Cambridge University Press.

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, *55*, 179–203.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f_0 contours. *Journal of Phonetics*, *27*, 55–105.

Xu, Y. (2002). Articulatory constraints and tonal alignment. In B. Bel, & I. Marlien (Eds.), *Proceedings of the speech prosody 2002 conference* (pp. 91–100). Aix-en-Provence, Laboratoire Parole et Langage: France.