

- Prator, Clifford H. Jr. & Betty Wallace Robinett. 1985. *Manual of American English Pronunciation*. (4th ed.). Orlando, FL: Harcourt Brace.
- Riera, María & Joaquín Romero. 2006. "V+/l/ and V+/r/ Sequences in American English: A preliminary descriptive acoustic study". *Actas del XXIX Congreso Internacional de la Asociación Española de Estudios Anglo-Norteamericanos/Proceedings of the 29th AEDEN International Conference* ed. by Alejandro Alcaraz Sintes, Concepción Soto Palomo & María de la Cinta Zunino Garrido. 529–536. Jaén, Spain: Universidad de Jaén Servicio de Publicaciones.
- . 2007. "Schwa in American English V+/l/ Sequences: Speaking rate effects". *Actes des/ Proceedings of JEL2007: Schwa(s) – 5^{èmes} Journées d'Études Linguistiques de Nantes*. 131–136. Nantes, France: Université de Nantes.
- Roach, Peter, James Hartman & Jane Setter, eds. 2006. *Cambridge English Pronouncing Dictionary*. (17th ed.). Cambridge: Cambridge University Press.
- Rogers, Henry. 2000. *The Sounds of Language: An Introduction to Phonetics*. Harlow, Essex, England: Pearson Education.
- Schmeiser, Benjamin. 2009. "Prosodic and Segmental Effects on Vowel Intrusion Duration in Spanish /r C/ Clusters". This volume.
- Trask, R.L. 1996. *A Dictionary of Phonetics and Phonology*. London: Routledge.
- Van Riper, Charles G. & Dorothy E. Smith. 1992. *An Introduction to General American Phonetics*. (3rd ed.). Prospect Heights, IL: Waveland Press.
- Warner, Natasha, Allard Jongman, Ann Cutler & Doris Mücke. 2001. "The Phonological Status of Dutch Epenthetic Schwa". *Phonology* 18:3. 387–420.
- Wells, J.C. 2000. *Longman Pronunciation Dictionary*. (2nd ed.). Harlow, Essex, England: Longman-Pearson Education.

Perception of word stress in Castilian Spanish

The effects of sentence intonation and vowel type*

Marta Ortega-Llebaria¹ & Pilar Prieto²

¹University of Texas at Austin/²ICREA & Universitat Autònoma de Barcelona

We provide evidence for the perception of the stress contrast in unaccented contexts in Spanish. Twenty participants were asked to identify oxytone words which varied orthogonally in two bi-dimensional paroxytone-oxytone continua: one of duration and spectral tilt, and the other of duration and overall intensity. Results indicate that duration and overall intensity were cues to stress, while spectral tilt was not. Moreover, stress detection depended on vowel type: the stress contrast was perceived more consistently in [a] than in [i]. Thus, in spite of lacking vowel reduction, stress in Spanish has its own phonetic material in the absence of pitch accents. However, we cannot speak of cues to stress in general since they depend on the characteristics of the vowel.

1. Introduction

The goal of this article is to investigate which acoustic cues and cue-interactions Spanish speakers use to perceive primary stress in unaccented contexts. Previous studies in Spanish have examined cues to stress in accented contexts, where there is co-variation between stress and accent (Enríquez et al. 1989; Llisterri et al. 2004 and Solé 1985, among others). For example, Llisterri et al. (2004) examined the perception of stress in one word declarative utterances, which are consistently produced with a pitch accent on the stressed syllable and lengthening on the last syllable. Consequently, they found that pitch was the main cue to stress, followed first by duration, and then by intensity. More specifically, duration and intensity cues were not sufficient to perceive stress unless they were processed together with a pitch accent. However, as Beckman and Edwards (1994:13) explained, this is a source of misunderstanding:

in short utterances, pitch excursions are likely to be interpreted in terms of the sequence at a nuclear accent, as in Fry's 1955 experiment showing the salience

*Many thanks to Eva Estebas and Ana Estrella for their help with data collection. This research was funded by a grant from the Ministerio de Educación y Ciencia, Spain.

of the F0 contour in cueing stress in pairs such as *pérmít* versus *perμίt*. This is probably the major source of the common misunderstanding in the experimental literature that F0 excursion is a direct acoustic correlate of the feature ‘stress’, a misunderstanding that has been incorporated into several standard textbooks.

Nevertheless, several studies have investigated how speakers of different languages use duration, intensity and vowel reduction in the production and perception of stress while controlling for the potential effects of accent. The earliest extensive instrumental study of this sort in English was that of Huss (1978). He embedded word minimal pairs that differed in primary stress, i.e. *ímport* noun – *ímport* verb, within post-focal contexts like those below. Note that vowels in *ímport* noun – *ímport* verb do not become reduced when they do not receive lexical stress, and that consequently, this word minimal pair contrasts vowels with primary and secondary stresses, not vowels with primary stress and unstressed vowels.

The GERMANS’ import sinks. (target noun)

The GERMANS import sinks. (target verb)

Huss’ results showed that although English speakers did produce small duration and intensity differences between vowels with primary and secondary stress, the very same speakers were not able to perceive these differences. Based on these results, Huss concluded that in the absence of pitch-accents and vowel reduction patterns, English speakers cannot perceive primary stress based solely on duration and intensity differences.

Beckman and colleagues (1994, 1997) obtained production results for English that complement those of Huss’ (1978). Campbell and Beckman (1997) compared fully stressed vowels, i.e. vowels with primary stress, with unreduced unstressed vowels, i.e. vowels with secondary stress, in different pitch-accent contexts. Their results revealed that the spectral balance did not differentiate vowels with primary stress from vowels with secondary stress in the absence of a pitch accent. They also observed that their 4 subjects varied in how reliably they used other cues, mainly duration, to mark stress. Based on Huss’ results and their own, they concluded that “there are no direct acoustic correlates for stress, and that, instead, the phonetic properties associated with stress at any level are parasitic on the phonetics of the relevant prominence marker”. More specifically, they agreed with Huss (1978) that at the lower levels of the prominence hierarchy, vowel reduction is a stronger correlate of word stress in English than duration and intensity:

categorical contrasts in ‘primary stress’ will be maintained in post-nuclear position [i.e. in a context where the word does not bear a pitch-accent] only if the words differ also in stress-foot structure. For example *dígést* should be categorically perceived as different from *dígest* only if the verb has [ɪ] rather than the full vowel [aɪ] in its first syllable (Beckman & Edwards 1994:15).

However, studies in other languages have examined lexical stressed syllables in deaccented contexts (Sluijter & van Heuven 1996; Sluijter, van Heuven & Pacilly 1997 for Dutch; Manolescu et al. 2009, this volume for Romanian; Dogil & Williams 1999 for German; Kastrikani 2003 for Greek; Ortega-Llebaria 2006; Ortega-Llebaria & Prieto 2009 for Spanish) and found that stressed syllables were consistently longer than their unstressed counterparts in the absence of pitch accents, revealing that duration was a cross-linguistic correlate of stress at the lower levels of the prominence hierarchy. For example, Manolescu and colleagues (2009, this volume) found that Romanian speakers consistently produced stressed syllables with longer durations than unstressed syllables in both declarative sentences with broad focus, where there is co-variation between stress and accent, and in unaccented post-focal contexts, where both stressed and unstressed syllables have a flat F0 melody.

Duration was also a strong correlate of stress in Dutch (Sluijter & van Heuven 1996; Sluijter, van Heuven & Pacilly 1997), even though Dutch and English are two closely related languages that have similar stress placement and vowel reduction patterns. Sluijter and colleagues found that in unaccented contexts, Dutch speakers, like English speakers, produced vowels with primary stress with longer durations, flatter spectral tilts and fuller vowel qualities than their unstressed counterparts. However, unlike English speakers, Dutch speakers mainly used duration and spectral tilt cues to perceive primary stress. In fact, when the perception task was performed with reverberation noise in the background, Dutch listeners continued to rely on duration cues but increased their reliance on intensity cues, demonstrating the ecological validity of their results. Thus, Sluijter and colleagues showed that at the lower levels of the prominence hierarchy, Dutch differed from English in that duration and intensity were stronger cues to perceiving primary stress than vowel reduction patterns. They conclude that

stress is not just a weaker degree of accent. One would expect to observe lower values along all measured correlates in stressed syllables of unaccented words. However, what we do observe is weakening along only those correlates that are related to the omission of the accent-lending pitch movement. (Sluijter & van Heuven 1996:2483).

Campbell and Beckman (1997) explained these cross-linguistic differences by appealing to the different strategies used by Dutch and American English in making stress perceptible in the absence of a pitch-accent: “Dutch differs from English in having relatively fewer words in which unstressed syllables are reduced” (Campbell & Beckman 1997:70). Consequently, vowel reduction constitutes a sufficient cue to stress only in English, while in Dutch, speakers need to phonologize other cues, such as duration and intensity, as correlates of stress.

Ortega-Llebaria and Prieto (2009) tested the above hypothesis with Central Catalan, a language that, like English, has phonological vowel reduction consistently related to stress, although only in a part of its vowel inventory. Central Catalan allows seven vowel phonemes /i, e, ε, a, ə, o, u/ to appear in stressed position and three vowels [i, ə, u] to appear in unstressed position. In unstressed position the vowel /i/ and /u/ surface as [i] and [u] respectively, i.e. *vi* ['bi] 'wine' > *vinet* [bi'nɛt] 'wine.dim', *suc* ['suk] 'juice' > *suquet* [su'kɛt] 'juice.dim'. The vowels /e, ε, a/ merge and surface as schwa [ə], i.e. *peix* ['peʃ] 'fish' > *peixet* [pə'ʃɛt] 'fish.dim', *sec* ['sek] 'dry' > *sequet* [sə'kɛt] 'dry.dim', *sac* ['sak] 'bag' > *saquet* [sə'kɛt] 'bag.dim'. And the vowels /ɔ, o/ merge and surface as [u], i.e. *poc* ['pɔk] 'few' > *poquet* [pu'kɛt] 'few.dim', *boca* ['bokə] 'mouth' > *boqueta* [bu'kɛtə] 'mouth.dim'. The following graph illustrates the vowel inventory in stressed and unstressed positions in Central Catalan:

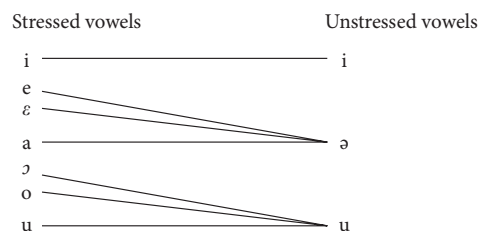


Figure 1. Inventory of stressed and unstressed vowels in Central Catalan. Lines indicate reduction patterns

Thus, since Catalan has a consistent morphophonological alternation between full long /e, ε, a, ɔ, o/ in stressed syllables versus schwa or [u] in unstressed syllables, our working hypothesis was that Catalan should behave more like English in relying mainly on vowel quality and not needing to phonologize duration and loudness as correlates of primary stress in the absence of accent. Results from our production experiment showed that, as expected, Catalan speakers reduced vowel [a] into a schwa when it became unstressed, and that they produced [i] with similar vowel qualities in stressed and unstressed contexts. Crucially, in spite of their different reduction patterns, both vowels were produced with longer durations and louder intensities in stressed contexts and this effect was intensified in vowels that underwent vowel reduction. Thus, having vowel reduction as a cue to stress not only did not prevent speakers from phonologizing duration and intensity cues, but vowel reduction also extended the duration and deepened intensity differences to the stress contrast. Results from our perception experiment further

disproved our working hypothesis (Ortega-Llebaria, Prieto & Vanrell, submitted). They showed that like Dutch speakers, Catalan speakers did rely on duration and intensity cues to perceive stress in vowel [i] but at the same time, like English speakers, they relied on vowel quality to perceive stress on vowel [a]. However, in contrast to Huss' English speakers who were unable to process duration and intensity cues in relation to stress, Catalan speakers still perceived the stress contrast by relying on duration and intensity cues when vowel reduction patterns for vowel [a] were neutralized in the speech signal. Moreover, once vowel reduction patterns were made available in the stimuli, Catalan speakers relied heavily on them to perceive stress.

Our results for Catalan indicate that the perception of primary stress is based on a cluster of cues whose weights change according to vowel type. Our subjects used vowel quality, duration and intensity cues to perceive primary stress in [a], and to duration and intensity cues to perceive primary stress in [i]. However, not a single cue, including vowel reduction, was absolutely necessary for the perception of stress. Even when we eliminated the vowel reduction patterns in [a], thereby creating an unusual context in Catalan, Catalan listeners, unlike Huss' English speakers, still perceived primary stress by relying on duration cues and if they were absent, on intensity cues.

Results from production experiments in Castilian Spanish (Ortega-Llebaria 2006; Ortega-Llebaria & Prieto, submitted) also support that notion that word stress is expressed by a cluster of acoustic correlates, and because in Castilian Spanish there is no phonological vowel reduction (Hualde 2005), these correlates work independently of vowel reduction patterns. While Spanish speakers produced the five vowels, i.e. [a,e,i,o,u], with the identical qualities in stressed and unstressed contexts, they produced stressed vowels with longer durations and louder intensities than unstressed vowels in both declarative sentences, where there is co-variation between stress and accent, and reporting clauses, where all syllables are unaccented. However, the lengthening effect of stress was larger in vowel [o] than it was in vowel [i], and intensity correlates to stress were larger in reporting clauses than in declarative sentences. We concluded that, although duration and intensity were consistent cues in the production of primary stress in Spanish, these cues had different weights according to sentence intonation and vowel type.

In order to examine how the duration and intensity cues found in the previous production experiment are perceived in relation to primary stress, we prepared the present perception experiment. Twenty native speakers of Castilian Spanish were asked to listen to and identify oxytone words which contained different vowels, in unaccented sentences. The target words varied orthogonally in two bi-dimensional

paroxytone-oxytone continua: one of duration and spectral tilt, and the other of duration and overall intensity. The research questions were:

1. In the absence of pitch-accents and vowel reduction patterns, can Spanish speakers still perceive a stress contrast based on duration and intensity cues?
2. How do duration and intensity cues interact?
3. How do Spanish speakers use duration cues to perceive stress across vowels? More specifically, when identical duration modifications are applied to vowels [a] and [i], do speakers rely less on duration cues to perceive stress in vowel [i], reflecting production patterns?
4. Since results in production indicate that overall intensity is a stronger correlate of stress than spectral tilt, do listeners also rely more on overall intensity cues than on spectral tilt to perceive stress?

2. Methodology

2.1 Recordings

A 41 year-old female monolingual speaker of Spanish from Barcelona was recorded saying the sentence *Hola – saluda mama contenta* ‘Hello, greets mama happily’, where the reporting clause *saluda mama contenta* was consistently pronounced with a flat pitch melody. The paroxytone target word *mama* was replaced with oxytone *mamá*, oxytone *mimí* and paroxytone *mimi*, yielding a total of 60 sentences (4 target words * 15 repetitions). Measurements of duration, intensity and spectral tilt were made on all target words and the sentences containing the paroxytone *mama* and *mimi* items with values closest to the average were selected for further manipulation. Vowels from these tokens had the same vowel quality.

2.2 Materials

Three *mama-mamá* (and *mimi-mimí*) continua were created by manipulating the cues of duration, overall intensity and spectral tilt separately. For each continuum, stimulus 1 had the syllable ratio typical of paroxytone words. This ratio decreased in stimulus 2, was close to 0 in stimulus 3, and increased again but in the opposite direction for stimuli 4 and 5, with stimulus 5 replicating the ratio of oxytone words. For example, for the overall intensity continuum, syllable 1 was 3 dBs louder than syllable 2 in stimulus 1. This difference decreased to 1.5 dBs in stimulus 2, and came close to 0 in stimulus 3. In stimulus 4, the second syllable was 1.5 dBs louder than the first one, and this difference increased to 3 dBs in stimulus 5 (see Table 1).

In order to change the duration ratio between the two syllables of *mama* (or *mimi*) while maintaining the same word duration, we cut one cycle from vowel 1 while adding another cycle to vowel 2. Because we modeled our stimuli after our recorded speaker and she consistently made larger inter-syllabic duration differences in oxytone than in paroxytone words to the extent that both syllables of some paroxytone *mama* or *mimi* tokens had roughly the same duration while in oxytone tokens the second syllable was on average 28 ms longer than the first, we kept these ratio differences in our test materials. Therefore, starting from stimulus 1, the number of cycles in each vowel of *mama* was 13–11, 12–12, 11–13, 10–14, 9–15, and in *mimi* 10–8, 9–9, 8–10, 7–11, 6–12. Since each cycle lasted 7 ms, the first vowel in stimulus 1 was 14 ms longer than the second vowel. In stimulus 2, this difference dropped to 0 ms, while in stimulus 3, the second vowel was 14 ms longer than the first. In stimulus 4, the second vowel was 28 ms longer than the first, and in stimulus 5, the difference was 36 ms.

In the spectral tilt continuum, the increments were 4 dBs in both the *mama* and *mimi* tokens. The main difference between the spectral tilt and overall intensity manipulations was that in the former, the increments of loudness were only applied in frequencies ranging from 500 Hz to 3000 Hz, while in the latter, the amplitude increments were applied uniformly across all the frequencies of the spectrum. All increments had values very similar to those employed by Sluijter et al. (1996, 1997).

Table 1. Increments for the overall intensity, spectral tilt and duration continua

	Stimulus 1 Clear paroxytone	Stimulus 2 paroxytone	Stimulus 3 neutral	Stimulus 4 oxytone	Stimulus 5 Clear oxytone
<i>Syllable ratios</i>	<i>Syll 1 > syll 2</i>	<i>Syll 1 > syll 2</i>	<i>Syll 1 = syll 2</i>	<i>Syll 1 < syll 2</i>	<i>Syll 1 < syll 2</i>
Over. Intensity	+3 dBs	+1.5 dBs	same	+1.5 dBs	+3 dBs.
Spectral tilt	+ 8 dBs	+ 4 dBs	same	+ 4 dBs	+ 8 dBs.
<i>Syllable ratios</i>	<i>Syll 1 > syll 2</i>	<i>Syll 1 = syll 2</i>	<i>Syll 1 < syll 2</i>	<i>Syll 1 < syll 2</i>	<i>Syll 1 < syll 2</i>
Duration	+2 cycles	same	+ 2 cycles	+ 4 cycles	+ 6 cycles

Finally, the five levels of the duration continuum were crossed with those of the overall intensity continuum, creating a 5*5 grid for each vowel. For example, grid 1 contained the 25 independent stimuli for the *mama-mamá* contrast and grid 2 included the 25 independent stimuli for *mimi-mimí*. Similarly, the five levels of duration continuum were also crossed with the five levels of spectral tilt, thus yielding grids 3 and 4. A summary of the crossed continua and resulting grids is depicted in Table 2.

Table 2. Summary of the five identification tasks

<i>Crossed continua</i>	<i>Tokens</i>	<i>Grid</i>	<i>Repetitions</i>	<i>Subjects</i>
Duration * overall intensity	25 <i>mama</i>	1	7	Group A = 10 subjects
	25 <i>mimi</i>	2	7	Group B = 10 subjects
Duration * spectral tilt	25 <i>mama</i>	3	7	Group A = 10 subjects
	25 <i>mimi</i>	4	7	Group B = 10 subjects

2.3 Subjects and listening tasks

Twenty native speakers of Spanish participated in the study. Their ages ranged from 21 to 60, and they had either been born in Madrid, Spain, or lived there for most of their lives. None of them reported having any speech or hearing problems.

A group of ten subjects (Group A in Table 2) was told that they would hear the words *mama* (with paroxytone stress) or *mamá* (with oxytone stress) inserted in the sentence *Hola – saluda ___ contenta*. They were asked to press the space bar on a keyboard as soon as they heard the oxytone word *mamá*. They listened to the 175 randomized sentences of the ‘duration * overall intensity’ condition (25 stimuli of grid 1 * 7 repetitions) in 7 blocks of 25 stimuli with an ISI of 1500 ms and a 10-second break between blocks. After a longer rest, they then listened to the 175 sentences of the ‘duration * spectral tilt’ condition (25 stimuli of grid 3 * 7 repetitions). Orders of presentation between the two tasks were counterbalanced across subjects. The remaining 10 speakers (Group B in Table 2) performed analogous tasks with the *mimi-mimí* stimuli.

2.4 Statistics

The effect of duration and overall intensity cues in the perception of primary stress was assessed by performing a Repeated Measures ANOVA with the within-subjects factors of duration and overall intensity and the between-subjects factor of vowel ([a] and [i]) on the probability of identification scores obtained from grids 1 and 2 (see 3.1 below). An analogous analysis was performed on the probability of identification scores obtained from the duration and spectral tilt modifications of grids 3 and 4 (3.2).

3. Results

3.1 Duration and overall intensity

The graph in Figure 2 illustrates the probabilities of *mamá* answers for the 25 stimuli. The slopes of the functions indicate that duration has a strong effect on

the perception of stress: for each intensity level, the oxytone responses increase along the duration continuum regardless of whether they are in a competing or enhancing relationship with intensity. As an example, the probabilities of identification of *mamá* in intensity level 1 (thick grey line) start close to 0 in duration 0.40 increasing progressively to 0.42, 0.55 and 0.63 until they reach 0.80 in duration 5. In addition to the strong effect of duration on stress perception, the spread of the intensity curves indicates that intensity also has an effect because the base rate of *mamá* responses increases with higher intensity levels. Thus, speakers rely heavily on duration, but the addition of intensity as an enhancing cue results in a shift of the response profile toward the alternative *mamá*.

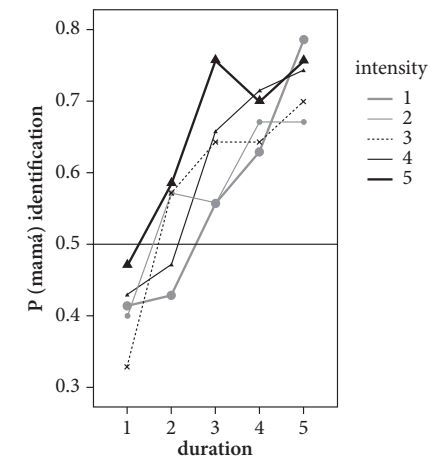


Figure 2. Probabilities of *mamá* responses for the 25 stimuli resulting from crossing the 5-levels of duration with the 5-levels of overall intensity in a *máma-mamá* continuum

It is easier to observe this effect of intensity in Figure 3, which depicts the same probabilities of *mamá* answers as in Figure 1, but with intensity along the x-axis. In general, there are fewer oxytone responses for intensity 1 than there are for intensity 5 which yields a slight ascendant trajectory along the intensity continuum for each duration level indicating that the oxytone responses increase along the intensity continuum. Although this ascendant trajectory becomes most visible at duration 3 showing that intensity has a stronger effect when duration cues are ambiguous, the slopes are by no means as strong as those depicted along the duration continuum in Figure 2. Therefore, both duration and intensity contribute to the perception of stress in [a] in an additive manner. Nonetheless, duration seems to be a stronger cue than intensity.

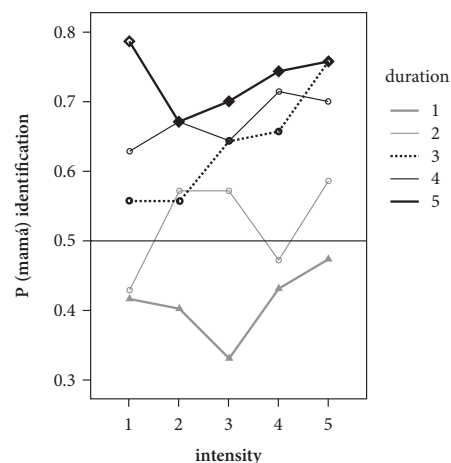


Figure 3. Probabilities of *mamá* responses for the 25 stimuli resulting from crossing the 5-levels of duration with the 5-levels of overall intensity in a *máma-mamá* continuum

The two graphs in Figure 4 show the probabilities of *mimí* judgements. In contrast with the results from *mamá*, subjects do not rely on duration and intensity to predict stress. Rather, they rely exclusively on intensity since oxytone responses increase only along the intensity continuum (as shown in the right-hand graph), but not along the duration continuum (left-hand graph). Even in stimuli with conflicting cues, such as paroxytone duration 1 with oxytone intensity 5, speakers favor intensity over duration and tend to hear oxytone *mimí*. Interestingly, when duration cues become ambiguous (level 3 in the x-axes of the left graph, and the dotted red line on the right graph), all intensity levels score around chance (.5) showing that speakers stop using intensity cues in the absence of clear duration cues to stress. Thus, in contrast with [a], duration and intensity do not seem to be in an additive relationship because listeners rely only on intensity cues to perceive stress in vowel [i]. However, duration shapes intensity judgments in that listeners use intensity solely when duration is not an ambiguous cue.

Results from a Repeated Measures ANOVA with the factors of duration and intensity on the probability scores confirmed the above patterns. Duration was significant only in *mamá* ([a]: $F(4,32)=15.753$, $p<.0001$; [i]: $F(4,32)=1.041$, $p=.401$) while intensity was strongly significant in *mimí* and only marginally significant in *mamá* ([a]: $F(4,32)=3.343$, $p=.042$; [i]: $F(4,32)=7.933$, $p<.0001$) showing that cues to stress depend on vowel type. While listeners use both duration and intensity cues

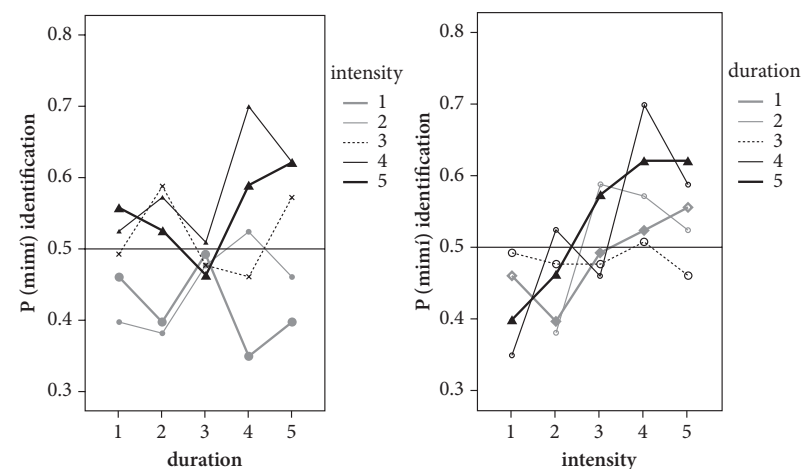


Figure 4. Probabilities of *mimí* responses for the 25 stimuli resulting from crossing the 5-levels of duration with the 5-levels of overall intensity in a *mimi-mimí* continuum

to perceive stress in [a], in [i] they only use intensity. Moreover, since interactions were not significant, they show that duration and intensity are additive in [a]. Partial Eta-Square estimates further confirm these results showing that duration in [a] explains 63% and intensity 23% of the variance, while in vowel [i] intensity alone accounts for 50% of the variance.

Multiple comparisons between intensity levels at neutral duration (stimulus 3) yielded significant results only for vowel [a]. Stimuli 1 and 2 were significantly different from stimuli 4 and 5. Therefore, in the absence of clear duration cues, intensity has an effect on the perception of stress only for vowel [a].

3.2 Duration and spectral tilt

The two graphs in Figure 5 illustrate the probabilities of *mamá* answers (graph on the left) and *mimí* answers (graph on the right) for the 25 *mama* and 25 *mimi* stimuli that resulted from crossing the 5-step duration continuum with the 5-step spectral tilt continuum. As in Figure 1, oxytone *mamá* or *mimí* answers increase along the duration continuum showing that duration has a strong effect on the perception of stress. However, the spread amongst the curves of the 5 spectral tilt levels is not as wide as that in Figure 2. Thus, speakers consistently relied on duration over intensity to perceive stress either when intensity was computed as changes in overall intensity or as variations in spectral tilt. Yet variations in spectral tilt seem to have a weaker effect than those in overall intensity.

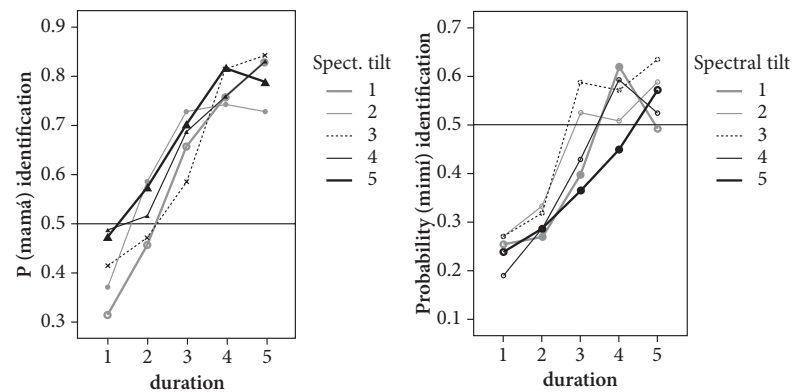


Figure 5. Probabilities of *mamá* (left)/*mimí* (right) responses for the 25 stimuli resulting from crossing the 5-levels of duration with the 5-levels of spectral tilt

Results from a RM ANOVA with duration and spectral tilt factors reveal that duration is the only significant cue in the perception of stress (*duration*: $F(4,68)=7.034$, $p=.003$, *dur*vowel*: $F(4,68)=3.01$, $p=.872$, *intensity*: $F(4,68)=1.679$, $p=.211$, *int*vowel*: $F(4,68)=1.906$, $p=.165$), thus confirming that duration is a strong cue to stress while spectral tilt is not. Partial Eta-Square Estimates further corroborate this result by showing that duration explains most of the variance in the data, i.e. 60% in [a] and 55% in [i].

A visual comparison of all six graphs shows that probability scores are consistently higher for [a] than they are for [i]. While 72% of the data for [a] score above .5, fewer than 50% of the data for [i] score above .5. Mean d-prime scores confirm that it was easier for speakers to detect oxytone words in vowel [a] than in vowel [i], i.e. d-prime for [a] scores 1.8, while for [i] it only reached 0.72.

In summary, results from the perception tasks show that Spanish speakers do perceive stress in reporting sentences, thereby confirming that stress can be perceived in the absence of pitch accents, and since Spanish has no vowel reduction, also in the absence of vowel reduction patterns. Perception of stress was based on a cluster of cues, namely duration and overall intensity, whose weights changed according to vowel type. In [a], speakers relied more on duration than on overall intensity cues. Yet, when duration cues to stress became ambiguous, speakers relied heavily on overall intensity. In contrast with vowel [a], overall intensity, not duration, was the main cue to stress in vowel [i]. Moreover, when duration cues became ambiguous for stress perception, speakers did not rely on overall intensity instead. Finally, spectral tilt did not have any effect on the perception of stress in Spanish.

4. Discussion

Our results demonstrate that even in the absence of pitch-accents and vowel reduction patterns in the speech signal, Spanish speakers still detect the stress contrast on the basis of duration and overall intensity differences between adjacent syllables, while ignoring differences in spectral tilt. Thus, the answer to our first research question is affirmative: speakers of Castilian Spanish perceive stress in unaccented contexts confirming that at the lower levels of the prosodic hierarchy, stress in Spanish has its own phonetic material which works independently of vowel reduction patterns.

However, the use that our Spanish speakers make of duration and intensity cues to perceive word stress differs across vowels. When listening to the *mama* stimuli, our Spanish speakers obtained higher identification scores in the stimuli where both cues were in an enhancing relationship, rather than in the stimuli where these cues were in a competing relationship. This shows that they use duration and overall intensity cues in an additive manner to perceive the stress contrast in vowel [a]. However, for vowel [i], our listeners use duration – and not very successfully – only if overall intensity is not an available cue in the signal. As mentioned in the introduction, the production experiments (Ortega-Llebaria & Prieto 2009) also shows a different use of duration in relation to stress across vowels. The lengthening effect of stress was smaller in vowel [i] than it was in other vowels, possibly because [i] is the vowel used in processes of speaker normalization (Johnson 1990). In contrast, overall intensity differences between stressed and unstressed vowels were similar across vowels. Thus, mirroring production patterns, results from this perception experiment showed that listeners relied more on duration cues when listening to stress in vowel [a] than in [i] in spite of performing the same manipulations of duration and intensity on the *mama* and *mimi* target words. Perception patterns therefore reflect the knowledge that speakers have of production. Since duration cues to stress in production are less salient in [i] than in other vowels, listeners rely less on duration when perceiving stress in vowel [i] than in vowel [a].

Moreover, listeners obtained higher d-prime scores for [a] than for [i] showing that they were more sensitive to the stress prominence in [a]. This asymmetry is related to the finding that manipulations of intensity tend to have a weaker effect on stress perception than manipulations of duration (Llisterri et al. 2004 for Spanish, Turk & Sawusch 1996 for English). For example, Turk and Sawusch (1996) found that English listeners' perception of duration and intensity in relation to stress was non-orthogonal. More specifically, they showed that irrelevant variations in duration when listeners were attending to intensity had a greater effect on the perception of stress than irrelevant variations in intensity when listeners were attending to duration. Thus, since our Spanish listeners attended more to duration cues when perceiving

stress in *mama*, and to intensity cues when perceiving stress in *mimi*, and duration cues have a stronger effect than intensity cues in the perception of stress, listeners were more sensitive to the stress contrast in *mama* than they were in *mimi*.

In short, our results indicate that the perception of stress in Spanish is based on a cluster of cues whose weights and interactions vary according to the phonology of the language, requirements of sentence intonation and vowel type. The phonology of Spanish rules out vowel reduction as a possible cue to stress (Hualde 2005). Moreover, the flat F0 contour of reporting clauses prevents pitch accents from cuing stress. In spite of this, Spanish listeners still perceive the stress contrast based on variations of duration and overall intensity. Moreover, vowel type determines the weights of duration and overall intensity cues. Spanish speakers produced longer vowels in stressed than in unstressed syllables, but these lengthening effects of stress had a more restricted range in vowel [i] than in other vowels. Mirroring production patterns, listeners relied less on duration when perceiving stress in [i] than when perceiving stress in [a]. However, it is necessary to test the perception of stress in the remaining Spanish vowels in order to fully understand the effect of vowel type in the weighting and integration of cues to stress.

Cross-linguistic research provides cumulative evidence that stress is not a reduced type of accent in stress-accent languages like Catalan, Spanish, and Dutch. While in stress-accent languages pitch accents aid to the perception of stress in the sense that only stressed syllables can bear a pitch accent, in the absence of pitch accents, the stress contrast is still perceptible. Duration and overall intensity in Spanish, duration and spectral tilt, and vowel reduction in Dutch, and duration, overall intensity and vowel reduction in Catalan constitute the phonetic material that cue the stress prominence in unaccented contexts showing that stress has its own phonetic material. This material is language specific because it varies according to the phonology of each language, and as shown for Catalan and Spanish, this material also depends on vowel type. In view of these data, it seems equally difficult to define stress as a mere structural device empty of any phonetic content as it is to find cross-linguistic cues to stress. Following Lieberman (1960), it seems reasonable to propose that the perception of stress is based on a cluster of cues that conveys rhythm to an utterance and that the variation we observed in the specific cues and cue weights that make up this cluster could be explained in relation to the perception of the rhythmic patterns of a language.

5. Conclusions

Results from this experiment indicate that in the absence of vowel reduction and pitch accents, Spanish speakers perceive stress by extracting the remaining

information from the signal. This information, in turn, is conditioned by vowel type. Listeners in this study relied mostly on duration to perceive stress in vowel [a] and on overall intensity in vowel [i]. Thus, it seems that perceived prominence in an unaccented context is based on a cluster of parameters, whose weights and interactions are determined by the phonology of the language and the requirements of sentence intonation and vowel type.

References

- Beckman, Mary & John Edwards. 1994. "Articulatory Evidence for Differentiating Stress Categories." *Papers in Laboratory Phonology III* ed. by Patricia A. Keating, 7–33. Cambridge: Cambridge University Press.
- Campbell, Nick & Mary Beckman. 1997. "Stress, Prominence and Spectral Tilt." *Intonation: Theory, Models and Applications* ed. by A. Botinis, G. Kouroupetroglou, G. Carayiannis, vol I, 67–70. GESCA.
- Dogil, G. & B. Williams. 1999. "The Phonetic Manifestation of Word Stress." *Word Prosodic Systems in the Languages of Europe* ed. by H. van der Hulst, 273–311. Berlin & New York: Mouton de Gruyter.
- Enríquez, E., C. Casado & A. Santos. 1989. "La Percepción del Acento en Español." *Linguística Española Actual* 11. 241–269.
- Fry, D.B. 1955. "Duration and Intensity as Physical Correlates of Linguistic Stress." *Journal of the Acoustical Society of America* 27:4. 765–768.
- Hualde, J.I. 2005. *The Sounds of Spanish*. Cambridge: Cambridge University Press.
- Huss, V. 1977. "English Word Stress in Post-Nuclear Position." *Phonetica* 35. 86–105.
- Johnson, K. 1990. "Contrast and Normalization in Vowel Perception." *Journal of Phonetics* 18. 229–254.
- Kastrikani, A. 2003. *The Temporal Correlates of Lexical and Phrasal Stress in Greek, Exploring Rhythmic Stress: Durational Patterns for the Case of Greek Words*. MA dissertation, University of Edinburgh.
- Lieberman, P. 1960. "Some Acoustic Correlates of Word Stress in American English." *Journal of the Acoustical Society of America* 32:4. 451–454.
- Llisterri, J., M.J. Machuca, C. de la Mota, M. Riera & A. Ríos. 2004. "La Percepción del Acento Léxico en Español." *Filología y Lingüística. Estudios Ofrecidos a Antonio Quilis*. vol I, 271–297. Madrid: CSIC.
- Manolescu, Alis, Daniel Olson & Marta Ortega-Llebaria. 2009. *Cues to Contrastive Focus in Romanian*. This volume.
- Navarro-Tomás, Tomás. 1967. *Manual e Pronunciación Española*. New York: Hafner, 6th Edition.
- Ortega-Llebaria, Marta. 2006. "Phonetic Cues to Stress and Accent in Spanish." *Selected Proceedings of the 2nd Spanish Laboratory Phonology* ed. by Manuel Díaz-Campos, 104–118. Somerville, Mass.: Cascadilla Press.
- Ortega-Llebaria, Marta & Pilar Prieto. 2009. *The Phonetics of Stress in Catalan and Spanish: cross-linguistic evidence for direct acoustic correlates of stress*. Submitted.
- Ortega-Llebaria, Marta, Pilar Prieto & M. Vanrell. 2009. *Catalan Speakers' Use of Vowel Reduction, Duration, and Intensity Cues to Perceive Word Stress in Deaccented Contexts*. Submitted.

- Sluijter, A. & van Heuven, V. 1996. "Spectral Balance as an Acoustic Correlate of Linguistic Stress". *Journal of the Acoustical Society of America* 100:4. 2471–2485.
- Sluijter, A, V. van Heuven & J. Pacilly. 1997. "Spectral Balance as a Cue in the Perception of Linguistic Stress." *Journal of the Acoustical Society of America* 101:1. 503–513.
- Solé, M.J. 1985. "Experimentos sobre la Percepción del Acento". *Estudios de Fonética Experimental* 1. 131–242.
- Turk, A. & J. Sawusch. 1996. "The Processing of Duration and Intensity Cues to Prominence". *Journal of the Acoustical Society of America* 99:6. 3782–3790.

Do complex pitch gestures induce syllable lengthening in Catalan and Spanish?*

Pilar Prieto¹ & Marta Ortega-Llebaria²
¹ICREA & Universitat Pompeu Fabra/²University of Texas at Austin

In both Spanish and Catalan, narrow contrastive focus and presentational broad focus in nuclear position have different pitch accent choices, namely a rising or a falling pitch accent, respectively. In words with final stress, narrow contrastive focus displays a rise-fall complex pitch gesture in the last syllable of the utterance. This article investigates the effects of the complexity of such a pitch pattern on the durational properties of the syllables in both languages when compared to the simpler falling pitch movement. The results of the production experiment reveal that, in general, the presence of a complex pitch pattern tends to have a lengthening effect on the target syllable. Yet we also find that some instances of this complex contour can be partially truncated, in which case it does *not* trigger lengthening. In sum, even though truncation and compression have been claimed to be language- and dialect-specific strategies (Ladd 1996; Grabe 1998; Grabe et al. 2000), in our data, truncation can be considered a speaker phonetic realization strategy that interacts with timing in such a way that there is a trade-off relationship between the two factors.

1. Introduction

In languages with a contrast between contour and non-contour tones, there are often restrictions on where those categories can occur. These restrictions vary from language to language, but there are some such patterns that recur independently

*The article further develops materials presented at the *3rd PaPI Conference* (Braga, June 2007). We are grateful to the audience at this conference and especially to John Kingston, G. Elordieta, J.I. Hualde, M. Simonet, and F. Torreira for very useful feedback. We are indebted to our Catalan and Spanish speakers (M. Nadeu, M. Magrans, M. Bosch, M. Albert, A. García, M.P. García). Finally, we would like to thank M. Nadeu for her help in recording and performing the segmentation of some of the speakers. This research was funded by grant 2005SGR-00753 from the Generalitat de Catalunya, and by grants HUM2006-01758/FILO and CONSOLIDER-INGENIO CSD2007-00012 from the Spanish Ministry of Science and Education.

and in unrelated languages. Contour tones tend to manifest themselves in contexts that are longer, while shorter contexts tend to produce neutralization. For example, it has long been noted that contour tones in Chinese languages such as Mandarin, Cantonese, and Fuzhou appear with fewer restrictions in language varieties that have coda contrasts, diphthongs, and longer rhyme durations, that is, in bimoraic syllables (Duanmu 1990, 1994a, b). Similarly, in Tokyo Japanese (unlike Kansai Japanese), a contour tone cannot be realized on a single mora (Maeda & Venditti 1998).¹ Nonetheless, the strictly moraic approach has been challenged by Zhang (2001, 2004), who points out that contour tones appear cross-linguistically in syllables which are long for independent reasons and which do not need to be bimoraic, e.g., syllables at the end of prosodic domains, syllables in shorter words, etc., and argues that the tone-bearing ability is rooted in phonetic behavior. In the realm of perception, experiments by Diehl and Kluender (1989) have revealed that high tones in Chinese may be misperceived as rising tones when syllables are longer.

In pitch-accent languages, it has also been observed that when complex pitch gestures appear as a result of two or three associated tones with a single syllable, either the contour is fully realized and 'compressed', or certain repair strategies appear, such as contour truncation (for examples of these different strategies, see Ladd 1996:132–136; Grabe 1998; and Grabe et al. 2000). Truncation and compression are two distinct phonetic realization strategies: while compression generally involves a speeding up of the pitch realization in order to produce a complete accent shape, truncation involves no pitch velocity change in the contour, which is only partially realized (see Grabe et al. 2000:162). In compressing languages, a few studies have observed that syllables bearing a complex pitch accent are longer than syllables bearing simpler pitch accents. For example, Gili-Fivela (2006) reports that syllables in contrastive focus (e.g., rise-fall gestures) are 7% to 10% longer than syllables in broad focus (e.g., rise gestures), regardless of their position within the word. Similarly, Ortega-Llebaria and Prieto (2006) recently investigated the durational properties of words with broad vs. narrow contrastive focus in Catalan and Spanish. They found that duration was amplified in narrow-focused words although only in final position. Since it is only in final position that we have a complex tonal gesture, the working hypothesis is that complex tonal gestures trigger lengthening. Crucially, durational differences only appeared between narrow vs. broad focus in words with final stress, a context where we find the pressure of realizing a complex rising-falling tonal gesture only in narrow focus. In contrast,

1. On the other hand, in Cantonese, contour tones have a lengthening effect: specifically, there is a statistically significant difference between the duration of a level tone and that of either a morphologically-derived or a sandhi-derived rising tone (Yu 2003).

the lack of duration effect in words with penultimate stress is attributed to the lack of realization of the complex pitch gesture.

The goal of this paper is to systematically investigate the effects of the presence of complex pitch movement on the durational properties of syllables in Catalan and Spanish and to test whether the abovementioned restrictions are rooted in phonetic behavior. Crucially, in both languages, narrow contrastive focus and presentational broad focus have different pitch accent choices (e.g., *She broke her neck, right?* – *No, she broke her LEG*) (see de la Mota 1995; Face 2002 for Spanish; Prieto 2002; Estebas-Vilaplana 2000; Astruc-Aguilera 2006 for Catalan). As Figure 1 shows, in Catalan narrow focus is realized as a rising accent and broad focus as a falling accent. Interestingly, when the accented syllable is in phrase-final position (*mamá*), the rising accent is realized as a complex rise-fall gesture, as follows.²

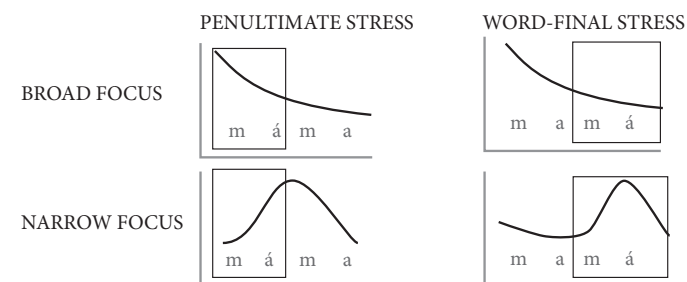


Figure 1. Schematic representation of the realization of the nuclear pitch accent in broad focus and in narrow contrastive focus utterances, in words with penultimate stress and final stress

This contrast will enable us to test the hypothesis that the presence of a complex pitch gesture in words with final stress in narrow focus will trigger an extra amount of lengthening in this syllable. The present article presents an analysis of 1280 utterances produced with the abovementioned contrasts. The analysis revealed the following: (a) in cases where the complex contour is fully realized, a clear difference is revealed between the duration properties of syllables carrying a narrow focus accent in final position vs. penultimate position; this is attributed to the pressure of realizing a complex tonal gesture only in final position (see Figure 1); (b) in cases of truncated (or partially truncated) contours, there is no clear lengthening effect. The potential implications of these results for cross-linguistic work on the constraints and strategies of contour compression and truncation will be discussed.

2. Both in Spanish and Catalan we have two possible words meaning 'Mum', namely, *mamá* and *mama* in Spanish, and *mamá* and *mama* in Catalan.

The article is organized as follows. Section 2 presents the methodology of the production experiment. Section 3 presents the main results of pitch range and duration of the production experiment. Finally, Section 4 discusses the main implications of this work for cross-linguistic studies on the interaction between pitch realization and duration.

2. Experimental investigation

2.1 Method

2.1.1 Materials. In both Catalan and Spanish, broad focus has been described as having a different prosodic realization from narrow contrastive focus (e.g., *She broke her neck, right? – No, she broke her LEG*) (see de la Mota 1995 and Face 2002 for Spanish; Prieto 2002; Estebas-Vilaplana 2000; Astruc-Aguilera 2006 for Catalan; for Romanian, see Manolescu, Olson & Ortega-Llebaria 2009, this volume). In Spanish and Catalan, nuclear broad focus is typically realized with a falling nuclear pitch accent. In contrast, the nuclear narrow contrastive focus is realized with a non-downstepped rising pitch accent, as Figure 2 shows. While the narrow contrastive pitch accent is generally transcribed as L+H*, there is no consensus as to what is the phonological analysis of the falling nuclear pitch accent in both languages. Some analyses such as Astruc-Aguilera (2006) for Catalan and Beckman et al. (2002) for Spanish have proposed the H+L* nuclear pitch accent, identifying this pitch accent with the broad focus nuclear pitch accent of European Portuguese (Frota 2002) and of Italian (D’Imperio 2002). On the other hand, Sosa (1999) transcribes it as having an L* nuclear accent – see the discussion about this topic in Beckman et al. (2002). Yet the main contrast between the two accent types is upheld under any of the analyses that seem tenable.

In order to test the hypothesis that the presence of a complex pitch gesture in words with final stress with narrow-corrective focus will trigger an extra amount of lengthening, we planned a controlled production experiment comparing near-minimal-pair words bearing the two types of pitch accents. A previous study with similar speech materials for Catalan and Spanish (Ortega-Llebaria & Prieto 2006) found that durational differences between stressed syllables in narrow vs. broad focus were only found in words with final stress, and not in words with penultimate stress. This contrast was attributed to pressure on the speaker to realize a complex tonal gesture only in narrow focus. Yet these materials did not control segmental content, which is done in the present experiment. The sentences under study were the following: (1) a broad-focus utterance (e.g., *Catalina me desanimó* ‘Catalina discouraged me’), and (2) a narrow-contrastive focus (e.g., *¿Catalina te animó?*

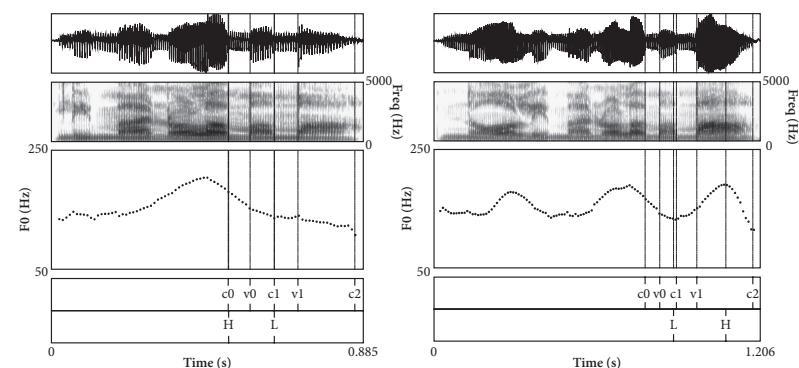


Figure 2. Waveforms, spectrograms, and F0 contours of the Catalan utterances *L'hi mano a la mamà* ‘I am asking this to Mum’ (broad focus, left panel) and *No, l'hi mano a la MAMÀ* ‘No, I am asking this to MUM’ (contrastive corrective focus, right panel)

‘Did Catalina encourage you?’ – *No, Catalina me DESANIMÓ* ‘No, Catalina DISCOURAGED me’).

To enhance the control for vowel and consonantal effects on duration (i.e., segmental effects), the target words selected are four two-syllable nouns that have the same segmental composition and that contrast only in the stress position in both languages: Spanish/Catalan *mamá* ‘Mum’ vs. *mama* ‘Mum’,³ and *Mimi* ‘proper noun’ vs. *mimi* ‘proper noun’. Each target word was placed in: (1) a broad-focus utterance (e.g., Span. *¿Qué pasa? – Se lo dice a mama* ‘(S)he is telling Mum’); and (2) a narrow-contrastive focus, which is indicated in capital letters (e.g., Span. *¿Se lo dice a mamá? – No, se lo dice a MAMA* ‘No, (s)he is telling MUM’), as follows:

Catalan	
/a/	/i/
<i>L'hi dono a la mama</i>	<i>L'hi dono a la Mimi</i>
‘I am giving it to Mum’	‘I am giving it to Mimi’
<i>L'hi dono a la MAMA</i>	<i>L'hi dono a la MIMI</i>
‘(No), I am giving it to MUM’	‘(No), I am giving it to MIMI’
<i>L'hi dono a la mamá</i>	<i>L'hi dono a la Mimi</i>
‘(No), I am giving it to Mum’	‘I am giving it to Mimi’

3. As mentioned above, there are two possible words in Spanish and Catalan for ‘Mum’, each of which have two different stress patterns, namely, *mamá* and *mama* in Spanish, and *mamá* and *mama* in Catalan. Even though both forms are used, *mamá* is more common in Spanish, and *mama* in Catalan.

<i>L'hi dono a la MAMÀ</i> '(No), I am giving it to MUM'	<i>L'hi dono a la MIMÍ</i> '(No), I am giving it to MIMÍ'
Spanish	
<i>Se lo doy a mama</i> 'I am giving it to Mum'	<i>Se lo doy a Mimi</i> 'I am giving it to Mimi'
<i>Se lo doy a MAMA</i> '(No), I am giving it to MUM'	<i>Se lo doy a MIMI</i> '(No), I am giving it to MIMÍ'
<i>Se lo doy a mamá</i> '(No), I am giving it to Mum'	<i>Se lo doy a Mimi</i> 'I am giving it to Mimi'
<i>Se lo doy a MAMÁ</i> '(No), I am giving it to MUM'	<i>Se lo doy a MIMÍ</i> '(No), I am giving it to MIMÍ'

These nouns in two focus contexts were inserted in 20 frame sentences with different verb types (see Appendix) and placed in sentence-final position.

2.1.2 Experimental procedure. Subjects were asked to look at a Power Point presentation comprising 80 slides (20 utterances x 2 words x 2 focus types) which contained suitable contexts for triggering an utterance with either broad or narrow focus, together with the answers, as follows:

BROAD FOCUS

Context: — ¿Què passa?
'What is going on?'
— *L'hi dono a la mamá*
'I am giving it to Mum'

NARROW-CONTRASTIVE FOCUS (with emphasis and assertiveness)

Context: — ¿L'hi dones a la mamá?
'Will you give it to "mamá"?'
— *No, l'hi dono a la MAMA*
'No, I will give it to "mamá"'

First, the experimenter read the context question aloud to the subject. The subject was supposed to read out the answer in the appropriate intonation. If the experimenter thought that the utterance had been mispronounced, the speaker was asked to repeat the sentence. This happened in only a handful of cases, and not for every speaker. The same process was repeated with each one of the slides.

Speakers were recorded individually in a quiet room, using a Sennheiser MKH20P48U3 omnidirectional condenser microphone and a Pioneer PDR609 digital CD-recorder. Speech samples were digitalized at 32000 Hz in 16-bit mono, double-checking that the target utterances were produced with the intended prosody.

2.1.3 Subjects. Four young female speakers of Central Catalan and four young speakers of Peninsular Spanish (between 23 and 40 years of age) participated in the experiment, giving a total of 160 utterances per speaker (20 verb types x 2 focus conditions (broad vs. narrow) * 4 target words stress conditions (*mama*, *mamá*, *mimi*, *mimí*) = 160 utterances per speaker. Thus, we obtained a total of 1280 utterances (160 utterances x 8 speakers = 1280 utterances).

2.2 Data analysis and measurements

The following measurements were made with *Praat* (Boersma & Weekink 2005; Wood 2005) on each of the 1280 target words. Figure 2 shows the labels that we used to segment the target words. In tier 1, we marked the beginning and end of the target segments, for example, 'm', 'a', 'm' and 'a'. In tier 2, we marked the valley (L) and the peak (H) of the pitch accent. In the cases where the pitch line was flat or descending, marks were placed at the beginning and end of the syllable. A Praat script extracted the F0 value in Hz at the marked points and calculated the pitch excursion size by subtracting the F0 values at L from the F0 values at H for each of the 1280 tokens.

2.3 Statistical analysis

In order to ensure that our data did indeed include two types of accents with different intonation properties, a Repeated Measures ANOVA with the within-subject factor of intonation (broad focus vs. narrow focus) and the between-subject factor of language (Catalan and Spanish) was performed on the pitch-range measurements. After this, we performed a Repeated Measures ANOVA on the duration of stressed vowels with three within-subject factors: stress position ([+word-final]/[-word-final]) and accent-type (broad focus vs. narrow focus) and vowel (i, a), and the between-subject factor of language (Catalan and Spanish). Since the vowel was not significant, we collapsed our data across vowels.

3. Results

The main research question posed by the study is whether syllable duration is amplified in contrastive narrow-focused words with respect to broad-focused words. That is, we investigate whether the presence of a complex F0 gesture is accompanied by an increase in duration.

3.1 Pitch excursion size differences

In this section, we check that the two types of focus (broad focus and narrow focus) in nuclear position were indeed realized using two different pitch accents, namely, a falling pitch accent and a rising pitch accent. The boxplots in Figure 3 show the mean pitch distance (in Hz) between LH values (in the case of a rising pitch movement) and between HL values (in the case of a falling pitch movement) of stressed syllables in broad-focused and narrow-focused conditions (striped boxes vs. dotted boxes, respectively) for all four Catalan speakers and all four Spanish speakers. It should be remembered that in cases in which the pitch was descending and no peak and valley could be visually identified, the pitch measures were taken at the beginning and at the end of the target accented syllable. As is clear from the graph, both Catalan and Spanish subjects consistently used a substantial pitch increase in narrow-focused sentences (a mean of 69.20 Hz, s.d. 28.47 in Spanish and 56.82 Hz, s.d. 45.04 in Catalan) and a negative increase in the broad-focus case (i.e., a mean of -33.65 Hz, s.d. 17.16 in Spanish and -28.36 Hz, s.d. 14.01 in Catalan).

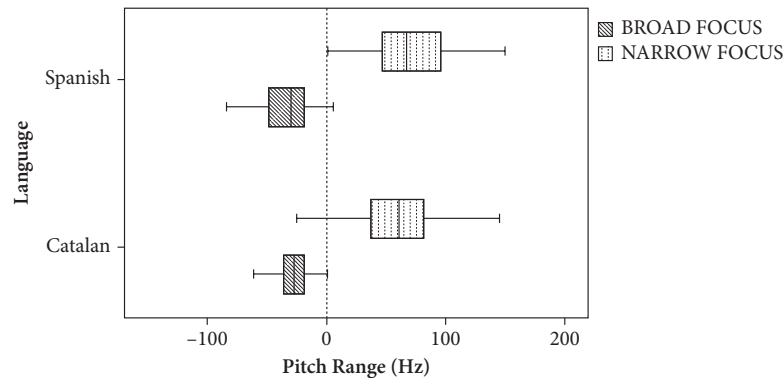


Figure 3. Mean values (in Hz) of the pitch distance between LH values (in the case of a rising pitch movement) and between HL values (in the case of a falling pitch movement) of the stressed syllables in broad focus (striped boxes) and narrow focus (dotted boxes) in Spanish and Catalan

If we plot the data separately by speaker (see Figure 4 below), the same pattern emerges for each one of the subjects, namely, speakers produce pitch accents with larger pitch ranges in narrow focus than in broad focus. While this difference is maintained, speakers also show some variation in the amount of pitch range values, especially in the narrow-focus case. For example, Spanish speaker (MB) has a mean pitch range of 34.93 Hz, while Spanish speaker (MN) has a much larger mean pitch range of 81.60 Hz.

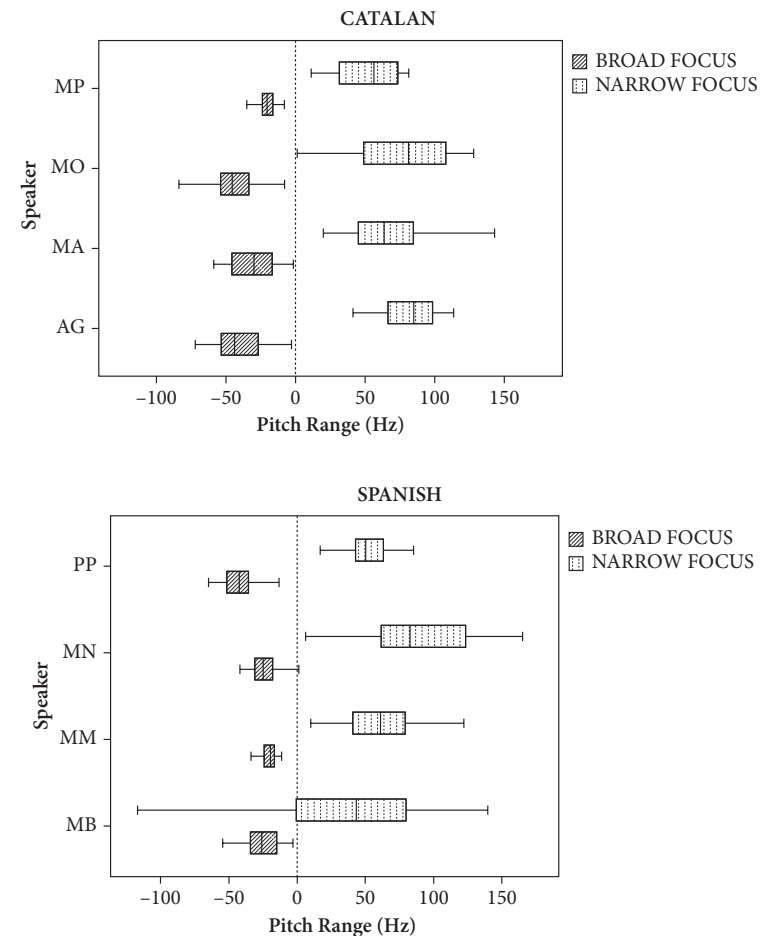


Figure 4. Mean values (in Hz) of the pitch range of the stressed syllables in broad focus (dashed boxes) and in narrow focus (dotted boxes) in Spanish and Catalan, plotted separately for all eight speakers

A Repeated Measures ANOVA on the pitch range of stressed syllables revealed highly significant effects of the accent factor on F0 variation (measured as pitch excursion size), at $F(1,587) = 989.798$; $p < .0001$ and no interaction between the accent*language factor, at $F(1,587) = 1.676$; $p = .196$. Hence, as expected, narrow focus was consistently cued by a rising pitch accent in the two languages. In contrast, broad focus in both languages was cued by a falling nuclear pitch accent.

3.2 Durational differences

Figure 5 below displays the mean duration (in ms) of the target syllable in penultimate (upper graph) and word-final (lower graph) position in Catalan and Spanish across vowels. In all four plots, the duration of the stressed syllables is compared between the narrow-focused (dotted boxes) and broad-focused (striped boxes) words. The graphs reveal that, in general, stressed syllables in narrow-focused words are longer than in broad-focused words in both languages, in both penultimate and word-final position (mean differences in penultimate and final position for Catalan: 39 ms and 55 ms; for Spanish: 69 ms and 207 ms). In general, the data works as expected: we find greater duration values in narrow-focused than

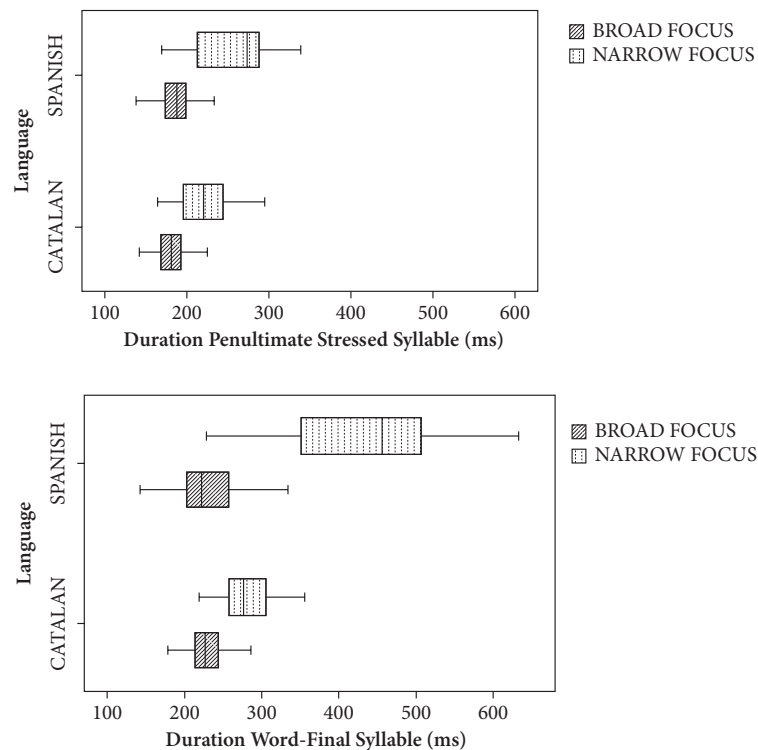


Figure 5. Mean duration values (in ms) of the target syllable in penultimate (upper graph) and word-final (lower graph) position in broad focus (dashed boxes) and narrow focus (dotted boxes), for both Spanish and Catalan

in broad-focused syllables both in final and penultimate positions and in both languages (that is, dotted boxes are always to the right of striped boxes). Yet the graphs also show that there is a contrast between Spanish and Catalan. While Spanish speakers produce an extra amount of lengthening in narrow-focused words in final position (see lower graph), this does not seem to be the case for Catalan speakers (mean and standard deviation for narrow focus in Spanish 437.5 ms (105.5) vs. 281.5 ms (33.7) in Catalan).

The boxplots in Figure 6 depict the same data for each one of the four Spanish subjects. In general, it is very clear that all Spanish speakers display substantial durational differences between narrow- and broad-focused syllables, both in penultimate

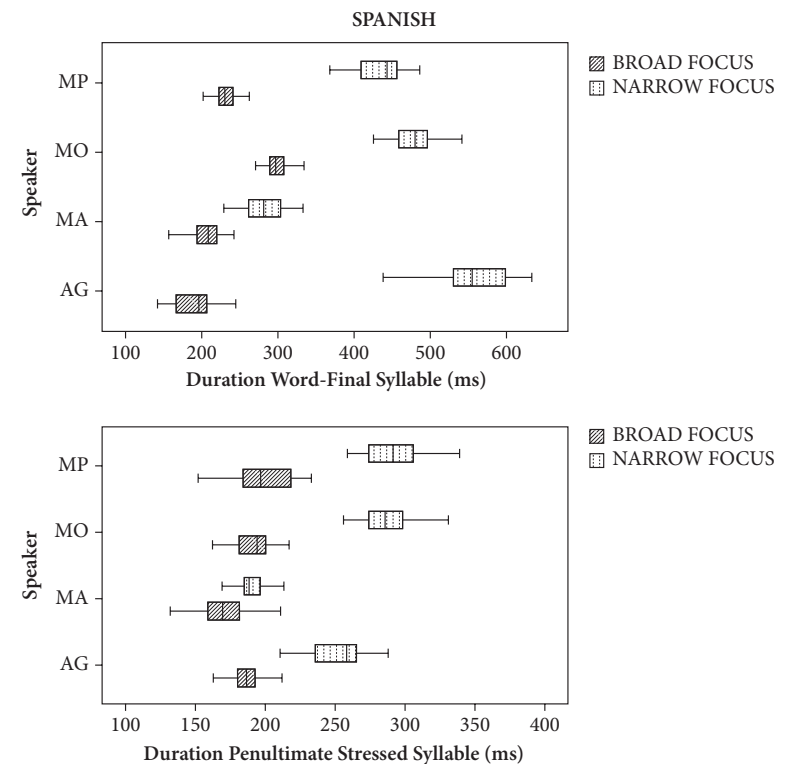


Figure 6. Mean duration values (in ms) of the target stressed syllables in broad focus (dashed boxes) and narrow focus (dotted boxes) in Spanish, plotted separately for all four speakers. Word-final syllables are plotted in the upper graph and penultimate syllables in the lower graph

and word-final position, and, importantly, there is a greater difference between the two when in word-final position, that is, when a complex tonal pattern is realized on the target syllable. For example, for speaker AG, the mean difference in duration between segments in broad focus and segments in narrow focus is 386 ms in words with final stress, while this difference falls to 67 ms in words with penultimate stress, thus showing that the lengthening effect of narrow focus with respect to broad focus is 316 ms longer in words with final stress than in words with penultimate stress. For speaker MO these mean differences are 180 ms and 96 ms respectively, and for speaker MP they are 204 ms and 92 ms, showing that the lengthening effect of narrow focus in words with final stress is 84 ms longer than that of words with penultimate stress for speaker MO, and 112 ms for speaker MP. For speaker MA, the lengthening effect of narrow focus in words with final stress is only 56 ms longer than in words with final stress since mean differences between narrow and broad focus in both conditions are 76 ms and 20 ms respectively.

Yet for Catalan subjects, subject differences may be found (see boxplots in Figure 7): while subjects MN and PP have an extra lengthening effect in narrow-focused syllables only of words with final stress, speakers MB and MM do not display this difference. Thus, similarly to Spanish speakers, for Catalan speakers MN and PP, the lengthening effect of narrow focus in words with final stress is around 40 ms longer than that of narrow focus in words with penultimate stress (mean differences between narrow and broad focus for words with final and penultimate stress for MN: 87 ms and 44 ms, for PP: 40 ms and 2 ms). However, this lengthening effect decreases to less than 10 ms for speaker MM (40 ms in words with final stress and 31 ms in words with penultimate stress) and even shows the opposite direction for speaker MB (52 ms in words with final stress and 83 ms in words with final stress).

We performed a Repeated Measures ANOVA on the duration of stressed vowels with two main within-subject factors: Position ([+word-final]/[-word-final]) and Accent-type (narrow focus vs. broad focus) and vowel (i, a), and the between-subject factor of language (Catalan and Spanish). Both Accent type and Position have a significant effect on the duration of the target syllables (Accent type: $F(1,318) = 910.65, p < .0001$; Position: $F(1,318) = 793.22, p < .0001$) corroborating the hypothesis that narrow focus extends the duration of segments and that this lengthening effect is larger in words with final stress. There was no significant vowel effect, thus indicating that lengthening effects are similar in vowels [a] and [i]. Importantly, there is a significant interaction between Accent type*Language ($F(1,318) = 910.65, p < .0001$), meaning that the lengthening effect of narrow focus is larger in Spanish than in Catalan. In the next section, we investigate possible

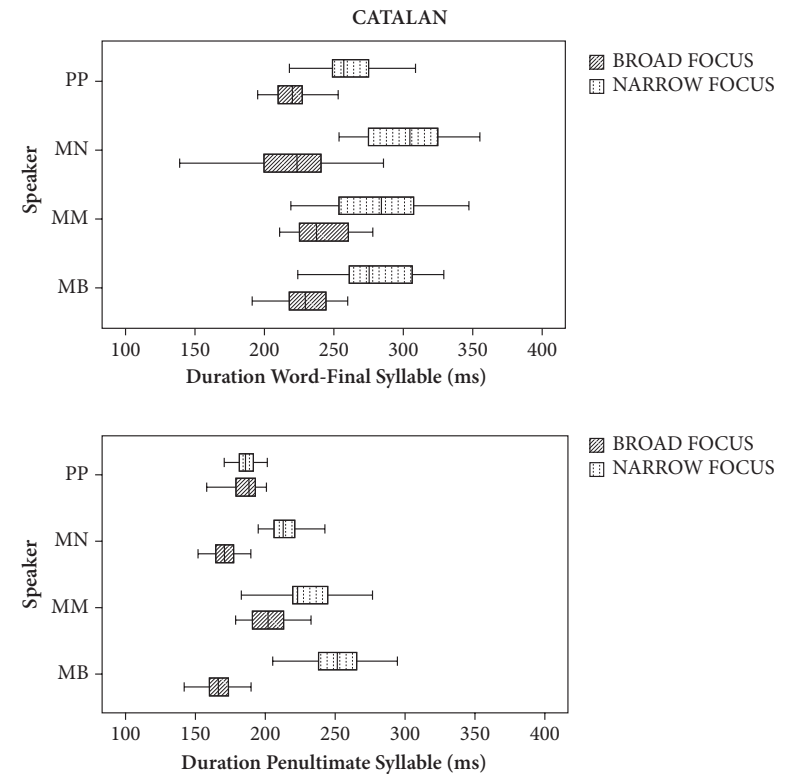


Figure 7. Mean duration values (in ms) of the target stressed syllables in broad focus (dashed boxes) and narrow focus (dotted boxes) in Catalan, plotted separately for all four speakers. Word-final syllables are plotted in the upper graph and penultimate syllables in the lower graph

sources of variation that can explain the distinctive and surprising behavior of these two Catalan speakers.

4. Discussion and conclusion: Sources of variation

When searching for sources of variation in the data, we noticed that several of the complex pitch accents that appear in narrow-focused words in utterance-final

position were truncated. Figure 8 shows the waveforms, spectrograms, and F0 contours of the Catalan utterance *No, l'hi mano a la MAMÀ* 'I am asking this to MUM' as uttered by two different speakers. In the left panel we see a fully compressed contour while the right panel shows a truncated contour.

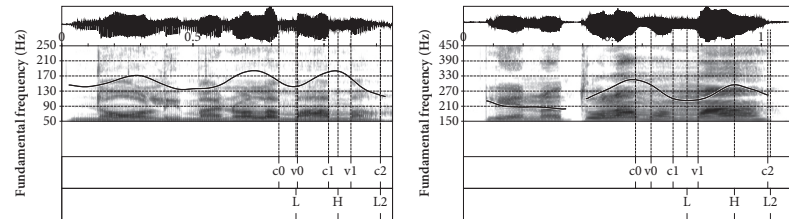


Figure 8. Waveforms, spectrograms and F0 contours of two possible prosodic realizations of the Catalan utterance *No, l'hi mano a la MAMÀ* 'No, I am asking this to MUM'. Left panel shows a fully compressed contour and right panel a truncated contour

We hypothesize that, in our data, unexpected subject differences in durational patterns might be attributed to the presence of truncated contours. Figure 9 shows the mean relative truncation of the final F0 value (in Hz) for the four Catalan speakers (lower panel) and the four Spanish speakers (upper panel). This value was calculated as the distance in Hz from the final F0 value in each of the narrow-focused contours to the reference baseline of the speaker (that is, the bottom of the speaker's pitch range, a value that is obtained at the end of broad focus statements) so that higher values correlate with truncation and lower values with fully realized pitch contours. The graphs show that speakers use different degrees of truncation and further support a view of truncation as a gradient acoustic effect. In general, Spanish speakers produce fewer truncated contours than Catalan speakers. Within the Spanish speakers, MA realizes full contours less often than the other speakers. And, among the Catalan speakers, MB and MM truncate contours more often. Interestingly, if we compare this data with the graphs in Figure 7, we can see that the Catalan subjects with more truncation (MB, MM, and MN) are exactly the ones who display less lengthening in narrow-focused words in final position. Similarly, the Spanish speaker who used full contours less often also showed the shortest lengthening effect of narrow focus in word-final position.

The scatter plot in Figure 10 shows the duration of the final syllable (in ms) in narrow focus contexts as a function of the degree of truncation (in Hz). For each speaker, the 'degree of truncation' was calculated relative to the speaker's baseline.

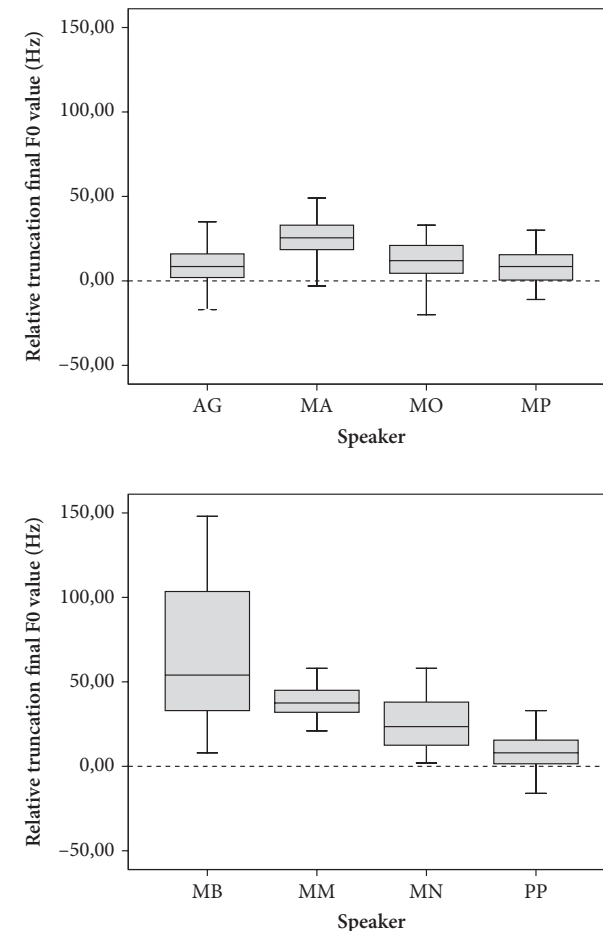


Figure 9. Mean truncation of the final F0 value (in Hz) in narrow-focus utterances relative to the baseline for all Spanish speakers (upper panel) and all Catalan speakers (lower panel)

Results show that there is a slight negative correlation between the two factors: in general, the more truncated the contour is, the shorter the accented syllable. Yet, we should be cautious in interpreting these results because this effect is not very strong and the data seems to be grouped in three different data sets. Linear

regression analyses show that truncation alone explains 15% of the variation in the syllabic duration of oxytone words with complex tones.

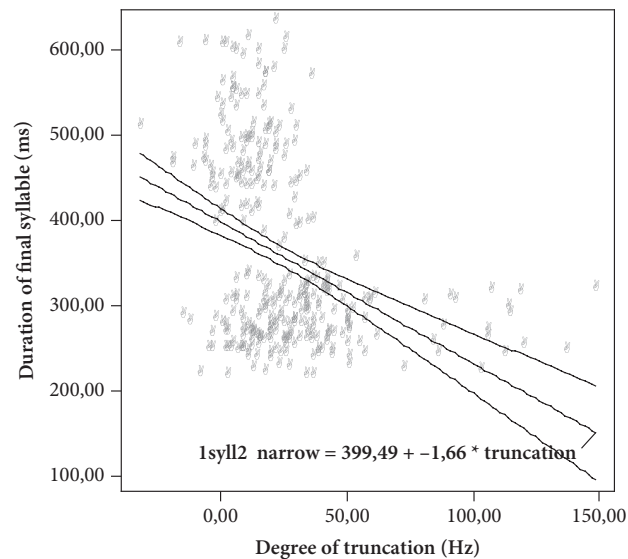


Figure 10. Duration of the final syllable (in ms) in narrow focus contexts as a function of the degree of truncation (in Hz), calculated relative to the baseline, for all speakers

In sum, the results of our production experiment reveal that in both Catalan and Spanish, duration is amplified in narrow-focused words, whether in penultimate or final position. In general, duration is also amplified in syllables with a complex tonal gesture. Figure 11 shows the mean duration of word-final target syllables in different conditions in our data for the two languages: [-stress, -complex pitch gesture], [+stress, -complex pitch gesture], and [+stress, + complex pitch gesture]. It is clear that once we control for final lengthening factors, we have to consider two factors that affect duration and that combine in an additive fashion: (a) stress; and (b) tonal complexity.

In our data, the presence of a complex pitch contour is the strongest lengthening factor. Multiple Comparisons with the Bonferroni adjustment show that the duration differences between these three types of syllables ([-stress, -complex pitch gesture], [+stress, -complex pitch gesture], and [+stress, + complex pitch gesture]) are statistically significant. The mean duration value and standard

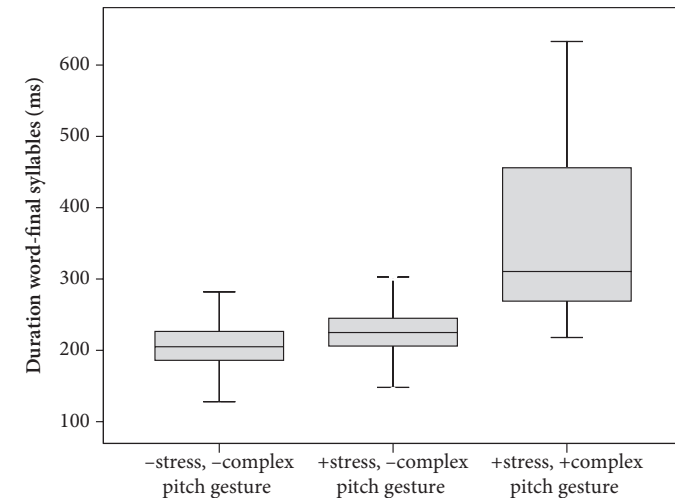


Figure 11. Mean duration of word-final target syllables in different conditions: [-stress, -complex pitch gesture], [+stress, -complex pitch gesture], and [+stress, +complex pitch gesture]

deviation are the following: [-stress, -complex] 206.5 ms (31.1), [+stress -complex] 228.6 ms (37.5), and [+stress, +complex] 359.5 ms (110.5).

In sum, truncation is used by some speakers as a phonetic realization strategy that interacts with timing. As we have just seen, some Catalan speakers truncate (or partially truncate) the complex contour: it is precisely these speakers who do *not* display the expected amount of extra lengthening in these syllables. Thus, in our data, the truncation of complex contours can be regarded as a phonetic realization strategy that interacts with timing in such a way that there is a trade-off relationship between the two factors. The observed phenomenon has consequences for cross-linguistic work on tonal realization strategies, namely, truncation and compression. Different studies have shown that there are cross-linguistic differences in the application of truncation and compression in standard varieties of English and German, and cross-dialectal differences within Swedish and Danish (see Ladd 1996; Grabe 2008; Grabe et al. 2000, among others). For example, while speakers of Cambridge English and Newcastle English compress rising and falling accents, speakers of Leeds English, in identical contexts, perform truncation. The data in this article challenge the view that truncation and compression are language and dialect-specific strategies (Ladd 1996, Grabe et al. 2000), and favor the view that they have to be regarded as phonetic realization strategies that interact in a dynamic way with timing.

References

- Astruc-Aguilera, Lluïsa. 2005. *The Intonation of Extra-Sentential Elements in Catalan and English*. Ph.D. dissertation, University of Cambridge.
- Beckman, Mary, Manuel Díaz-Campos & J.T. McGory and T.A. Morgan. 2002. "Intonation across Spanish, in the Tones and Break Indices Framework". *Probus* 14. 9–36.
- Diehl, Randy L. & Keith R. Kluender. 1989. "On the Objects of Speech Perception". *Ecological Psychology* 1.2. 121–144.
- D'Imperio, Mariapaola. 2002. "Italian Intonation: An overview and some questions". *Probus* 14. 37–69.
- de la Mota, Carme. 1995. *La Representación Gramatical de la Información Nueva en el Discurso*. Ph.D. dissertation, Universitat Autònoma de Barcelona.
- Duanmu, San. 1990. *A Formal Study of Syllable, Tone, Stress and Domain in Chinese Languages*. Ph.D. dissertation, MIT, Cambridge, Mass..
- Duanmu, San. 1994a. "Against Contour Tone Units". *Linguistic Inquiry* 25: 555–608.
- Duanmu, San. 1994b. "Syllabic Weight and Syllabic Duration. A correlation between phonology and phonetics". *Phonology* 11. 1–24.
- Estebas Vilaplana, Eva. 2000. *The Use and Realization of Accentual Focus in Central Catalan with a Comparison to English*. Ph.D. dissertation, University College London.
- Face, Timothy. 2002. "Local Intonational Marking of Spanish Contrastive Focus". *Probus* 14. 71–92.
- Frota, Sónia. 2002. "Nuclear Falls and Rises in European Portuguese: A phonological analysis of declarative and question intonation". *Probus* 14. 113–146.
- Gili-Fivela, Barbara. 2006. "Scaling' e Allineamento dei Bersagli Tonal: l'identificazione di due accenti discendenti", *Atti del Convegno Nazionale AISV (Associazione Italiana di Scienze della Voce), Analisi prosodica: teorie, modelli e sistemi di annotazione*, ed. Renata Savy & Claudia Crocco, 214–232. Torriana: EDK.
- Gordon, Matthew. 2001. A Typology of Contour Tone Restrictions. *Studies in Language* 25. 405–444.
- Grabe, Esther. 1998. *Intonational Phonology: English and German*. Ph.D. dissertation, Max-Planck-Institute for Psycholinguistics and University of Nijmegen.
- Grabe, Esther, Brechtje Post, Francis Nolan & Kim Farrar. 2000. "Pitch Accent Realization in Four Varieties of British English". *Journal of Phonetics* 28. 161–185.
- Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.
- Maeda, Kazuaki & Jennifer Venditti. 1998. "Phonetic Investigation of Boundary Pitch Movements in Japanese". Paper 0800 presented at the 5th International Conference on Spoken Language Processing, held in Sydney, Australia, 30 November – 4 December, 1998.
- Manolescu, Alis, Daniel Olson & Marta Ortega-Llebaria. 2009. "Cues to Contrastive Focus in Romanian". This volume.
- Ortega-Llebaria, Marta & Pilar Prieto. 2006. "Acoustic Effects of Contrastive Focus in Catalan and Spanish". Paper presented at the 3rd Conference on Laboratory Approaches to Spanish Phonology (LASP), held at the University of Toronto, 8–10 September 2006.
- Prieto, Pilar. 2002. "Entonació". *Gramàtica del Català Contemporani*, ed. by Joan Solà, Maria-Rosa Lloret, Joan Mascaró & Manuel Pérez Saldanya, chap. 11 (= pp. 393–462). Barcelona: Editorial Empúries.
- Xu, Yi & Xuejing Sun. 2002. "Maximum Speed of Pitch Change and how It May Relate to Speech". *Journal of the Acoustical Society of America* 111. 1399–1413.

- Yip, Moira. 1989. "Contour tones". *Phonology* 6. 149–174.
- Yu, Alan C.L. 2003. "Contour Tone-Induced Lengthening in Cantonese". *Proceedings of the XVth International Congress of Phonetic Sciences*, Barcelona, 3–9 August 2003 ed. by Maria Josep Solé, Daniel Recasens & Joaquim Romero, 2381–2384. Denver, Col.: Causal Productions.
- Zhang, Jie. 2001. *The Effects of Duration and Sonority on Contour Tone Distribution Typological Survey and Formal Analysis*. Ph.D. dissertation, University of California, Los Angeles.
- Zhang, Jie. 2004. "Contour Tone Licensing and Contour Tone Representation". *Language and Linguistics* 5.4. 925–968.

Appendix

TARGET VERBS (CATALAN)	TARGET VERBS (SPANISH)
L'hi mano a la mama	Se lo mando a mama
L'hi cuso a la mama	Se lo coso a mama
L'hi dono a la mama	Se lo vendo a mama
L'hi porto a la mama	Se lo debo a mama
L'hi torno a la mama	Se lo hago a mama
L'hi bullo a la mama	Se lo ruego a mama
L'hi bato a la mama	Se lo nombro a mama
L'hi poso a la mama	Se lo pongo a mama
L'hi dicto a la mama	Se lo dicto a mama
L'hi envio a la mama	Se lo envío a mama
L'hi pago a la mama	Se lo noto a mama
L'hi deixo a la mama	Se lo pago a mama
L'hi rento a la mama	Se lo pido a mama
L'hi cuino a la mama	Se lo leo a mama
L'hi busco a la mama	Se lo busco a mama
L'hi pinto a la mama	Se lo lavo a mama
L'hi brodo a la mama	Se lo llevo a mama
L'hi ballo a la mama	Se lo dejo a mama
L'hi tallo a la mama	Se lo presto a mama
L'hi baixo a la mama	Se lo guiso a mama