Spanish speakers perceive stress more easily in vowel [a] than in vowel [i]

Marta Ortega-Llebaria*, Pilar Prieto↑±
* University of Texas-Austin (UT)
↑ Institució Catalana de la Recerca i Estudis Avançats (ICREA)
±Universitat Autònoma de Barcelona (UAB), Spain

ABSTRACT

We provide evidence for the perception of the stress contrast in de-accented 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.

Keywords: stress, Spanish, perception.

INTRODUCTION

The goal of this article is to investigate which acoustic cues and cue-interactions Spanish speakers use to perceive primary stress in de-accented 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. 2005 and Solé 1985, among others). Consequently, they found that pitch was the main cue to stress followed by first duration and then intensity. However, as Beckman and Edwards (1994) 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 1958 experiment showing the salience of the F0 contour in cueing stress in pairs such as *pérmit* versus *permít*. 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." (Beckman & Edwards 1994:13).

Nevertheless, several studies have investigated the acoustic cues that speakers from different languages use in the production and perception of stress while controlling for the potential effects of accent. Yet, their results did not reach an agreement about the phonetic nature of stress but in fact lead to opposite views. For example, based on results from English, some researchers claimed that there are no direct acoustic cues related to stress in either production or perception, and consequently, they describe stress as a structural device for marking pitch-accents which is empty of any phonetic content of its own (see for example Huss, 1978, Hayes, 1995 and references therein, Beckman & Edwards, 1994, and Campbell & Beckman, 1997, among others). In contrast, results from Dutch (Sluijter & van Heuven 1996, Sluiter, van Heuven & Pacilly 1997) lead to the opposite view. They claim that primary stress in Dutch has its own phonetic material because variations in duration and intensity correlate with the stress contrast in de-accented contexts. 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).

The earliest extensive instrumental study to examine the acoustic and perceptual correlates of stress without the confounding effects of accent in English was that of Huss (1977). He embedded word minimal pairs that differed in primary stress, i.e. *<u>im</u>port* noun – *im<u>port</u>* verb, within post-focal contexts like those in (1). Note that vowels in *<u>im</u>port* noun – *im<u>port</u>* 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.

(1) 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 can not 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 (1977). Campbell & Beckman (1997) compared full stressed vowels, i.e. vowels with primary stress, to unreduced unstressed vowels, i.e. vowels with secondary stress, in different pitch-accent contexts. Their results revealed that 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 claimed 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" (see also Beckman & Edwards 1994: "Stress is not an autosegment"). The phonetic cues to prominence markers vary according to their level in the prosodic hierarchy. At the lower levels of the hierarchy, patterns of vowel reduction signal prominence in English. At higher levels, the relevant phonetic cues are the pitch excursions linked to pitch accents. Thus Beckman & Edwards' hypothesis predicts that stress differences having no vowel reduction in unaccented contexts will not be perceived in English. As they claim, "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 *digést* should be categorically perceived as different from *dígest* only if the verb has [ə] rather than the full vowel [aɪ] in its first syllable" (Beckman & Edwards 1994:15). As mentioned above, this prediction was borne out in Huss' perception experiment.

Even though English and Dutch are closely related languages, have similar stress placement patterns, and both use vowel reduction, Sluijter and colleagues' (1996, 1997) experiments on stress in Dutch yielded results similar in production but not in perception to those for English. They found that in de-accented 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 used duration and spectral tilt cues, and not vowel reduction patterns, to perceive primary stress. When the same perception task was performed over loudspeakers and there was therefore reverberation noise, Dutch listeners continued to rely on duration cues but increased their reliance on intensity cues, thus

demonstrating the ecological validity of their results. Consequently, Sluijter and colleagues showed that duration and intensity were strong cues to the perception of primary stress in Dutch at the lower levels of the prominence hierarchy, and unlike in English, these cues worked independently of vowel reduction patterns.

Campbell & 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). Even though Dutch and English have vowel reduction, reduced vowels in Dutch tend to appear in inflectional or derivational morphemes, and are therefore not as directly related to stress as they are in English (Gussenhoven, 2003) which makes vowel reduction a reliable cue to stress only in English. Therefore, English speakers can rely on vowel reduction to successfully perceive primary stress while Dutch speakers need to phonologize other cues such as duration and intensity. Thus, from this difference between Dutch and English, it can be inferred that the phonetic material that marks the stress contrast in a given stress-accent language seems to be determined by the strength of the correlation between vowel reduction and stress. If the correlation is strong, vowel reduction is the primary cue to stress. However, if the correlation is weak, other cues like duration and intensity mark the stress contrast.

Ortega-Llebaria and Prieto (submitted) tested the above hypothesis with Central Catalan, a language that like English has phonological vowel reduction consistently related to stress, but only in part of its vowel inventory. Central Catalan allows seven vowel phonemes /i, e, E, a, O, o, u/ to appear in stressed position and three vowels [i, ≅, u] to appear in unstressed position. In unstressed position the vowel /i/ and /u/ surface as [i] and [u] respectively, i.e., *vi* [ˈbi] ‘wine’ > *vinet* [biˈnɛt] ‘wine.dim’, *suc* [ˈsuk] ‘juice’ > *suquet* [suˈkɛt] ‘juice.dim’. The vowels /e, E, a/ merge and surface as schwa [ə], i.e., *peix* [ˈpeʃ] ‘fish’ > *peixet* [pəˈʃɛt] ‘fish.dim’, *sec* [ˈsɛk] ‘dry’ > *sequet* [səˈkɛt] ‘dry.dim’, *sac* [ˈsak] ‘bag’ > *saquet* [səˈkɛt] ‘bag.dim’ . And the vowels /O, o/ merge and surface as [u], i.e., *poc* [ˈpɔk] ‘few’ > *poquet* [puˈkɛt] ‘few.dim’, *boca* [ˈbokə] ‘mouth’ > *boqueta* [buˈkɛtə] ‘mouth.dim’. The following two graphs illustrate the vowel inventory in stressed and unstressed positions in Central Catalan:

(2) Inventory of vowels in Central Catalan

a. Stressed vowels		b. Unstressed vowels	
i	u	i	u
e	o		
ɛ	ɔ	ə	
a			

Thus since Catalan allows only the two shortest vowels /i/ and /u/ to occur in stressed and unstressed syllables, and has a morphophonological alternation between full long /e, ɛ, 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 on vowel quality and other phonetic correlates of the contrast between full and reduced vowels, 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 as well, but vowel reduction also enlarged the duration and intensity differences to the stress contrast. Results from our perception experiment further disproved our working hypothesis (Ortega-Llebaria, Prieto & Vanrell, submitted). They showed that Catalan speakers, like Dutch speakers, did rely on duration and intensity cues to perceive stress in vowel [i] but at the same time they relied on vowel quality, like English speakers, to perceive stress on vowel [a]. However, when vowel reduction patterns for vowel [a] were neutralized in the speech signal, Catalan speakers still perceived the stress contrast by relying on duration and intensity cues, in contrast to Huss' English speakers who were unable to process duration and intensity cues in relation to stress. Moreover, once vowel reduction patterns were made available in the stimuli, Catalan speakers strongly relied on them to perceive stress.

These results do not entirely agree with those obtained for English and Dutch. On the one hand, using vowel reduction as a cue to stress did not prevent Catalan speakers from phonologizing duration and intensity, as Beckman and colleagues showed for English speakers. On the other hand, even though some vowels in Catalan, like in Dutch, do not undergo vowel reduction in unstressed contexts, Catalan listeners did not rely exclusively on duration and intensity cues to perceive primary stress across different vowels, as Sluijter and colleagues proposed for Dutch. Thus, regardless of the reduction patterns of the vowel, Catalan speakers used any cue that was available in the speech signal to perceive stress indicating that the strength of the correlation between stress and vowel quality does not necessarily determine the phonetic content of stress.

In fact, our results for Catalan indicate that the perception of primary stress is more flexible than predicted by this correlation because stress perception is based on a cluster of cues whose weights change according to the intonation of the sentence and vowel type. For example, Catalan speakers use pitch-accents as a cue to primary stress only in accented syllables. Yet, in de-accented contexts, Catalan listeners resort to other cues. 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]. Indeed, the most compelling evidence for this flexibility is that 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 in their absence, on intensity cues.

Results from production experiments in Spanish (Ortega-Llebaria, 2006, Ortega-Llebaria & Prieto, submitted) also support that notion that the perception of primary stress is based on a cluster of acoustic cues that work independently of vowel reduction patterns and whose weights change according to vowel type. In Castilian Spanish there is no phonological vowel reduction, and consequently, the five stressed vowels [a,e,i,o,u] maintain the same quality in unstressed contexts. In spite of that, Spanish speakers produced stressed vowels with longer durations and louder intensities in both declarative sentences, where there is co-variation between stress and accent, and reporting clauses, where all syllables are de-accented, showing that duration and intensity are cues to primary stress in Spanish in the absence of vowel reduction and regardless of the presence or the absence of pitch-accents. However, the lengthening effect of stress was larger in vowel [o] than it was in vowel [i], and intensity cues to stress were larger in reporting clauses, where there are no pitch-accents, than in declarative sentences. We concluded that, although duration and intensity were consistent cues to 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 experiment. Twenty native speakers of Castilian Spanish were asked to listen to and identify oxytone words, which contained different vowels, in de-accented 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 spite of the lack of pitch-accents and vowel reduction patterns, can Spanish speakers perceive a stress contrast?

(2) Despite the fact that production patterns indicate that cues to stress vary according to vowel type, are Spanish speakers able to perceive primary stress across vowels in de-accented contexts?

(3) If they do, how do they use duration cues to stress across vowels? More specifically, when identical duration modifications are applied to the stimuli for vowels [a] and [i], is duration a stronger cue to the perception of primary stress in [a] than it is in [i], reflecting production patterns?

(4) Since results in production indicate that intensity was a stronger cue in de-accented contexts than in declarative sentences, we expect listeners to use intensity in the perception of stress in reporting sentences. Will listeners use overall intensity as a cue to stress as they did in production? Or will they use spectral tilt, like Dutch speakers?

(5) How do intensity cues to stress interact with duration cues?

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 separately manipulating the cues of duration, overall intensity and spectral tilt. 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 became 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 modelled our stimuli after our recorded speaker and she consistently made larger inter-syllablic 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 maintained 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, this difference was 36 ms.

In the spectral tilt continuum, the increments were of 4 dBs in both *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 applied only 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).

	Stimulus 1 Clear paroxytone	Stimulus 2	Stimulus 3 neutral	Stimulus 4	Stimulus 5 Clear oxytone
Syllable ratios	*Syll 1> syll 2*	*Syll 1> syll 2*	*Syll = syll 2*	*Syll 1< syll 2*	*Syll 1< syll 2*
Over.	+3 dBs	+1.5 dBs	same	+1.5 dBs	+3 dBs

Intensity					
Spectral tilt	+ 8 dBs	+ 4 dBs	same	+ 4 dBs	+ 8 dBs
Syllable ratios	*Syll 1> syll 2*	*Syll 1= syll 2*	*Syll 1 < syll 2*	*Syll 1< syll 2*	*Syll 1< syll 2*
Duration	+2 cycles	same	+ 2 cycles	+ 4 cycles	+ 6 cycles

Table 1. Values for the duration, overall intensity and spectral tilt continua. For example, in the overall intensity continuum, syllable 1 of stimulus 1 was 3 dBs louder than syllable 2. In stimulus 2, this difference decreased to 1.5 dBs. In stimulus 3, the two syllables had similar intensities. The second syllable was the louder one in the word for stimuli 4 and 5, the differences being 1.5 and 3 dBs respectively.

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 of spectral tilt, thus yielding grids 3 and 4. A summary of the crossed continua and resulting grids is depicted in Table 2.

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

Table 2. Summary of identification tasks. There were 5 identification tasks, one per grid. Each grid contained the *mama* or *mimi* stimuli resulting from crossing different continua.

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 been born in Madrid, Spain, or had 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 asked to press the space bar in a keyboard as soon as they heard the oxytone word *mamá* in the sentence *Hola – saluda ___ contenta* over headphones. They listened to the randomized 175 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 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 1 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*.

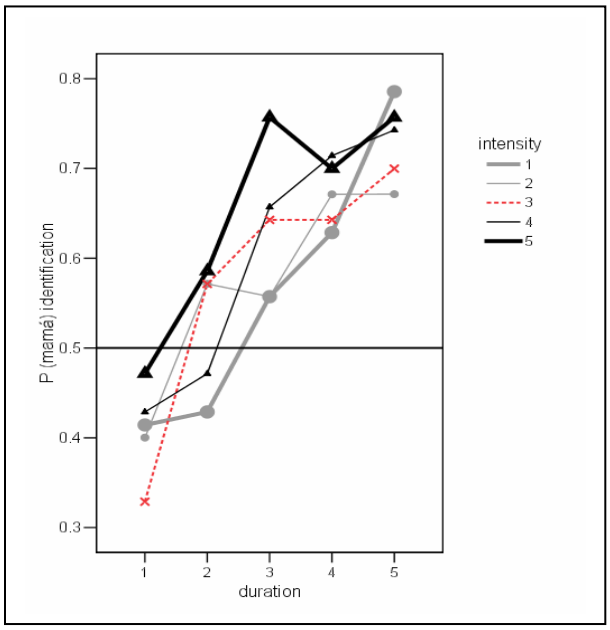


Figure 1: Probabilities of *mamá* responses for the 25 stimuli resulting from crossing the 5-levels of duration with the 5-levels of overall intensity in a *máma-mamá* continuum.

It is easier to observe this effect of intensity in Figure 2, 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 1. Therefore, both duration and intensity contribute to the perception of stress in [a] in an additive manner. Yet, duration seems to be a stronger cue than intensity.

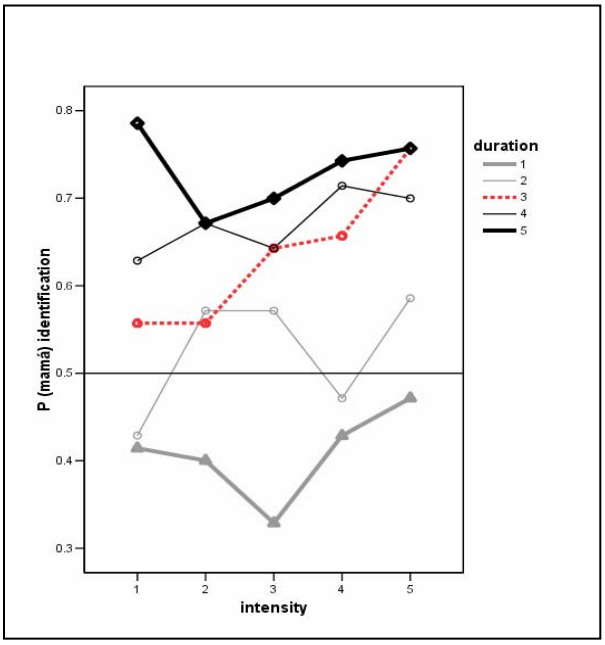


Figure 2**:** Probabilities of *mamá* responses for the 25 stimuli resulting from crossing the 5-levels of duration with the 5-levels of overall intensity in a *máma-mamá* continuum.

The two graphs in Figure 3 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 scored 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.

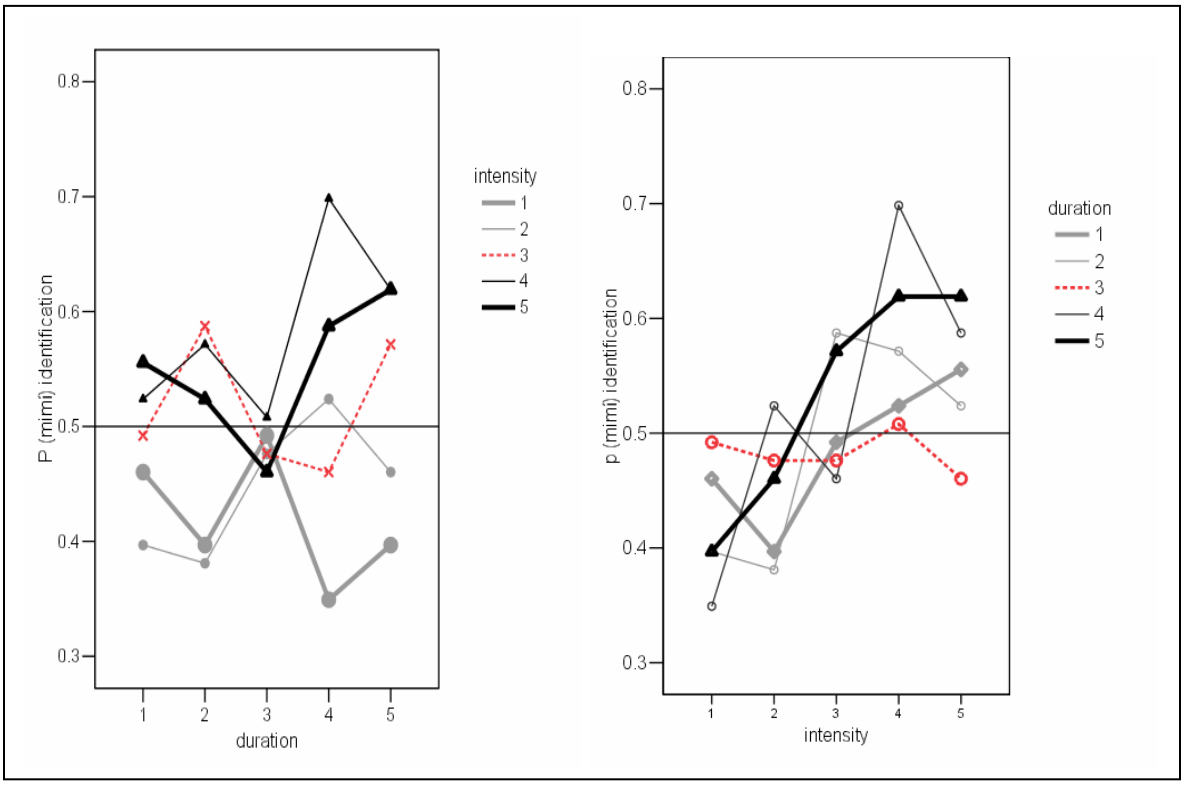


Figure 3: Probabilities of *mimí* responses for the 25 stimuli resulting from crossing the 5-levels of duration with the 5-levels of overall intensity in a *mimi-mimí* continuum.

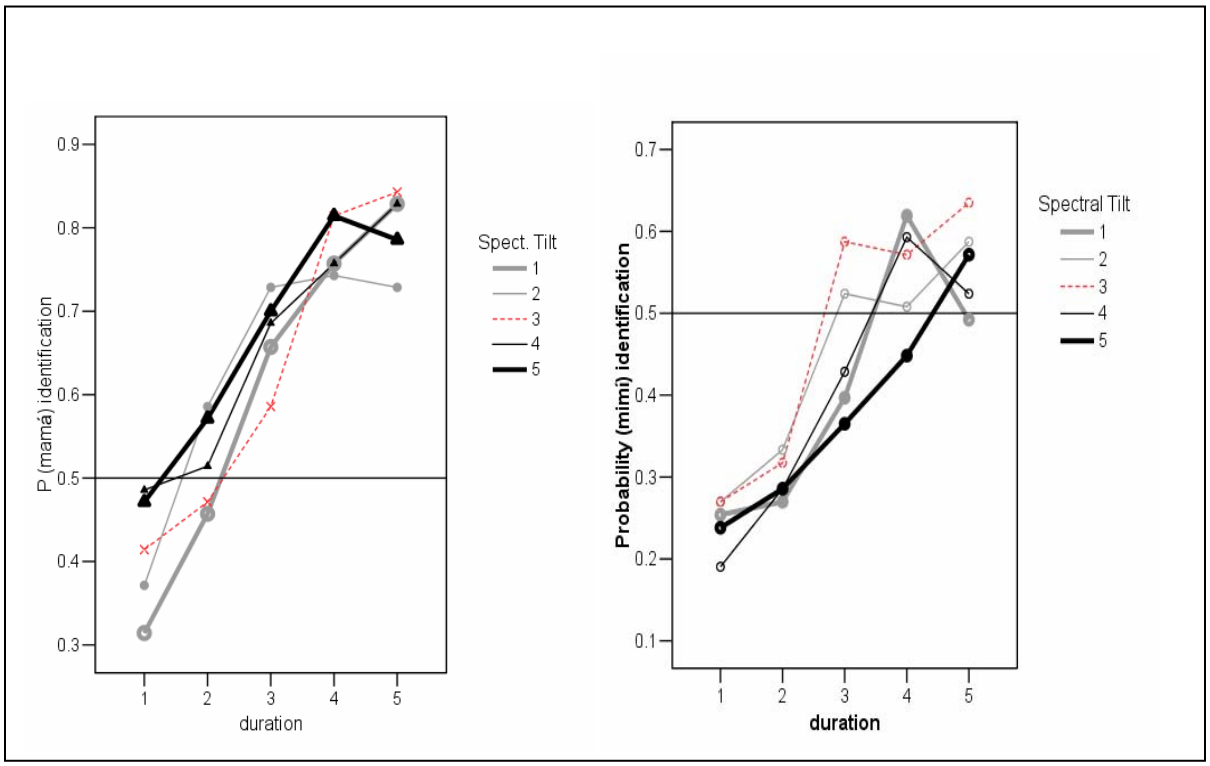
Results from a Repeated Measures ANOVA with the factors of duration and intensity on the probability scores confirmed the above patterns. Duration was significant only in *mamá* ([a]: $F(4,32)=15.753$, $p<.0001$; [i]: $F(4,32)=1.041$, $p=.401$) while intensity was strongly significant in mimí and only 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 to perceive stress in [a], in [i] they use only 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 4 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. Like 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 1. 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 4: 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 the factors of duration and spectral tilt 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)=301, 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 score above .5 for [a], fewer than 50% scored above .5 for [i]. 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 reaches 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 strongly relied 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 use 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 when there are no pitch-accents and no 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 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. These results are very similar to those for Dutch, which provides support for a phonetic basis to the stress contrast. Nevertheless, we can not talk of cues to stress in general because, at least in Spanish, cues to stress vary not only according to the phonology of the language and the requirements of sentence intonation but also according to vowel type.

In Castilian Spanish, the same five vowels that appear in stressed contexts also occur in unstressed syllables. Therefore, the phonology of Castilian Spanish, unlike the phonologies of English, Dutch and Catalan, does not include changes in vowel quality which eliminates one cue to stress in Spanish, i.e. vowel reduction, leaving as possible cues pitch accents, and variations in duration and intensity. Sentence intonation also has an effect on the way pitch, duration and intensity cues are produced in relation to stress. In Spanish, all stressed syllables in declarative sentences bear a pitch accent, causing co-variation between stress and accent. In reporting clauses, however, all syllables are un-

accented (Navarro-Tomás, 1967). Results from our production experiment (Ortega-Llebaria and Prieto, submitted) indicate that duration and intensity were cues to stress in both sentence types, thereby showing that duration and intensity mark the stress contrast independently of pitch accents. However, in reporting clauses, where there are no pitch accents, speakers used intensity cues more than in declarative sentences indicating a possible compensatory relationship between pitch accents and intensity cues.

More interestingly, vowel type also has an effect on the production and perception of stress. In our production experiment, 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). 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 exactly the same manipulations of duration and intensity having been carried out on the *mama* and *mimi* target words. Thus, perception patterns 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 common finding that manipulations of intensity tend to have a weaker effect on stress perception than manipulations of duration. Turk and Sawusch (1996) found that listeners' perception of duration and intensity 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 variation in intensity when listeners were attending to duration. These results indicate that duration is a stronger cue than intensity to the perception of the stress prominence. Therefore, our listeners' higher sensitivity to the stress prominence in [a] than in [i] relates to the fact that following their knowledge of production, they attend more to duration cues in [a] than in [i] when listening to the stress contrast, and since variations in duration have a greater effect on the perception of stress than variations in intensity, they were able to perceive stress prominence better in [a].

The results from Turk and Sawusch (1996) also showed that variations in duration alone, but not variations in intensity alone, constitute a possible cue to the stress prominence. However, our results on the interaction between duration and intensity provide only partial support for these findings. In the stimuli where duration differences constituted ambiguous cues to the stress contrast (stimulus 3 in the duration continuum), and it was consequently difficult to extract information about stress based on duration cues, listeners showed two patterns. For vowel [i], listeners could not extract information about intensity as illustrated by the flat function for 'duration 3' in Figure 3b. In this instance, as predicted by Turk and Sawusch, intensity is not a possible cue to stress in the absence of duration cues. In contrast, listeners were able to make stress judgments based on intensity cues alone for vowel [a]. Probabilities of oxytone responses increased along the intensity continuum when duration cues were neutral (stimulus 3 in the duration continuum of Figure 2). In fact, this is the only stimulus in the duration continuum where

variations in intensity had a significant effect, thus showing that in the absence of duration cues, listeners were able to use intensity cues alone to perceive stress.

In summary, our results indicate that the perception of stress in Spanish is based on a cluster of cues whose weights vary according to the phonology of the language, sentence intonation and vowel type. The phonology of Spanish rules out vowel reduction as a possible cue to stress. In spite of that, stress in Spanish, like in Dutch, is still perceptible. The requirements of sentence intonation determine whether pitch accents, together with duration and intensity, cue the stress contrast. Moreover, our data showed a compensatory relationship between pitch accents and intensity, since in de-accented contexts, listeners relied more on intensity cues than they did in accented contexts. More interestingly, vowel type determines the weights of duration and intensity cues. Spanish speakers produced variations in duration according to stress, but these variations had a more restricted range in vowel [i] than in other vowels. Accordingly, listeners relied less on duration when perceiving stress in [i] than in other vowels, showing their knowledge of production patterns.

Comparing results from Spanish with those from Catalan and Dutch shows that stress is not a reduced type of accent. While in all stress-accent languages, pitch accents aid in 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 in Dutch, and duration, overall intensity and vowel reduction in Catalan constitute the phonetic material that cue the stress prominence in de-accented 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. 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 a deaccented context is based on a cluster of parameters, and it is not restricted to duration and intensity.

ACKNOWLEDGMENTS

Many thanks to Eva Estebas and Ana Estrella for their help with data collection. This research was supported by grants HUM2006-01758/FILO from the Ministerio de Educación y Ciencia-FEDER, Spain and by 2005SGR-00753 from the Generalitat de Catalunya.

REFERENCES

Beckman, M., Edwards, J. 1994. Articulatory evidence for differentiating stress categories. In: Keating, P. A. (ed), *Papers in laboratory phonology III*, Cambridge: Cambridge University Press, 7-33.

Campbell, N., Beckman, M. 1997. Stress, prominence and spectral tilt. In: Botinis, A, Kouroupetroglou, G, Carayiannis, G. (eds), *Intonation: Theory, Models and Applications*. ESCA: 67-70.

Enríquez, E., Casado, C., Santos, A. 1989. La percepción del acento en español. *Lin.. Esp. 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.

Gussenhoven, C. 2003. Vowel duration, syllable quantity and stress in Dutch. In: Hanson K., Inkelas, S. (eds), *The nature of the word: Essays in honor of Paul Kiparsky*. Cambridge, MA: MIT Press.

Hayes, B. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago and London: The University of Chicago 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.

Llisterri, J., Machuca, M. J., de la Mota, C., Riera, M., Ríos, A. 2005. La percepción del acento léxico en español. *Filología y lingüística. Estudios ofrecidos a Antonio Quilis*. Madrid: CSIC, Vol. 1, pp. 271-297.

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*. Cascadilla Press,104-118. http://www.lingref.com/cpp/lasp/2/index.html

Ortega-Llebaria, M., Prieto, Pilar. (Submitted) The phonetics of stress in Catalan and Spanish: cross-linguistic evidence for direct acoustic correlates of stress.

Ortega-Llebaria, Prieto, P. & Vanrell, M. (Submitted) Perception of stress is flexible: Catalan listeners' use of vowel reduction, duration, and intensity cues.

Sluijter, A., van Heuven, V. 1996. Spectral balance as an acoustic correlate of linguistic stress. *J. Acoust. Soc. Am.*, 100 (4), 2471-2485.

Sluijter, A, van Heuven, V., Pacilly, J. 1997. Spectral balance as a cue in the perception of linguistic stress. *J. Acoust. Soc. Am.* 101 (1) 503-513.

Solé, M.J. 1985. Experimentos sobre la percepción del acento. *Estudios de Fonética Experimental* 1, 131-242.

Turk, A, Sawusch, J. 1996. The processing of duration and intensity cues to prominence, *Journal of the Acoustical society of America* 99 (6) 3782-3790.

van Heuven, V.J. 2001. *Boven de klanken / Beyond the Segments*. Leiden University: Inaugural address in acceptance of the Chair in Phonetics. [http://www.let.leidenuniv.nl/ulcl/faculty/vheuven/oratie-eng.htm].