

# Perceptual Evidence for Direct Acoustic Correlates of Stress in Spanish

*Ortega-Llebaria, Marta; Prieto, Pilar<sup>†</sup>; Vanrell, Maria del Mar<sup>‡</sup>*

University of Texas at Austin, ICREA<sup>†</sup> and Universitat Autònoma de Barcelona<sup>‡†</sup>

[mllebaria@yahoo.com](mailto:mllebaria@yahoo.com), [pilar.prieto@uab.es](mailto:pilar.prieto@uab.es), [mariadelmar.vanrell@uab.es](mailto:mariadelmar.vanrell@uab.es)

## ABSTRACT

We provide evidence for the perception of the stress contrast in deaccented 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.

## 1. INTRODUCTION

Stress conveys prominence below the sentence level, while accent signals the most prominent word in a sentence. Thus, understanding how prominence is expressed and perceived in a given language implies disentangling the effects of stress from those of accent. The earliest extensive instrumental study on this topic is that of Huss [5] for English. He examined minimal pairs that differed only in stress, like ‘import’ verb versus ‘import’ noun, embedded in post-focal deaccented contexts. Although he found small duration differences between stressed vowels and their unstressed counterparts, English listeners identified nouns versus verbs only at chance level, showing that in the absence of vowel reduction, English listeners could not perceive stress.

Beckman and colleagues [1,2] obtained results for English that complement those of Huss [5]: vowel reduction is a cue to stress at the foot level, but at the word level, pitch accents differentiate primary from secondary stress. Therefore, Beckman et al. do not interpret stress in terms of an autosegment, but of a structural feature that marks distributional patterns by borrowing its

phonetic content from prominence markers at different levels of the prosodic hierarchy. Thus, there are no “direct” phonetic properties associated with stress and, at lower levels of the prosodic hierarchy, these properties are “parasitic” on vowel reduction differences, at least in languages like English.

Sluijter and colleagues [7,8] examined the stress contrast in deaccented contexts in Dutch, a language that has vowel reduction, but unlike English, reduced vowels appear in inflectional or derivational morphemes, and are therefore not as directly related to stress as they are in English [4]. They found that in the absence of pitch accents, Dutch speakers produced stressed syllables with longer durations, flatter spectral tilts and fuller vowel qualities than their unstressed counterparts, and that duration and spectral tilt, not vowel reduction, were the strongest cues to stress perception in a context of background noise. Thus, they showed that at the lower levels of the prominence hierarchy, stress in Dutch has its own phonetic basis independent of vowel reduction patterns.

New evidence on the issue discussed above can be provided when the languages compared differ maximally in vowel reduction. Ortega-Llebaria and Prieto [10] examined production patterns of stress in Catalan and Spanish. Central Catalan, like English, has a phonological pattern of vowel reduction related to stress while Castilian Spanish does not have vowel reduction. Their results indicate that, in both languages, stress is cued by duration and intensity in the absence of pitch accents, supporting Sluijter and van Heuven’s results [7,8]. However, the duration and intensity cues to stress vary according to vowel type and, in languages like Catalan, they become amplified by vowel reduction.

The key questions we examine in the present perception experiment are: (1) In spite of their small range, are the duration and intensity differences that cue stress in Spanish perceptible, as they are in Dutch? Or in the absence of vowel reduction, do Spanish speakers need a pitch accent

to perceive stress, like English speakers? (2) Is it possible to talk of stress correlates in general, or does the perception of stress depend on vowel type? Previous experiments on the perception of stress in Spanish [3,6,9] have chosen contexts in which the presence of a pitch accent was mandatory, and to our knowledge, their results did not indicate any vowel effect. For our experiment, the use of typically unaccented reporting clauses will allow us to examine the role of non-pitch cues in several vowels.

## 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*, 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* 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 stimulus 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.

In the duration continuum the increments were of 7 milliseconds, while for spectral tilt they were of 4 dBs. These increments were very similar to those employed by Sluijter et al [5]. The main

difference between spectral tilt and overall intensity is that in the former, the increments of loudness are applied only in frequencies higher than 500 Hz, while in the latter, they are applied across all the frequencies of the spectrum. The same three continua were also created for the *mimi-mimí* contrast.

The five levels of the duration continuum were crossed with those of the overall intensity continuum creating a 5\*5 independent stimulus conditions for each vowel, i.e. 25 stimuli for the *mama-mamá* contrast and 25 for the *mimi-mimí*. Similarly, the five levels of duration continuum were also crossed with the five of spectral tilt.

### 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 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 \* 7 repetitions) in 7 blocks of 25 stimuli with an ISI of 500 ms. and 10 seconds break between blocks. After a longer rest, they then listened to the 175 sentences of the ‘duration \* spectral tilt’ condition. Orders of presentation between the two tasks were counterbalanced across subjects.

The remaining 10 speakers performed analogous tasks with the *mimi-mimí* stimuli.

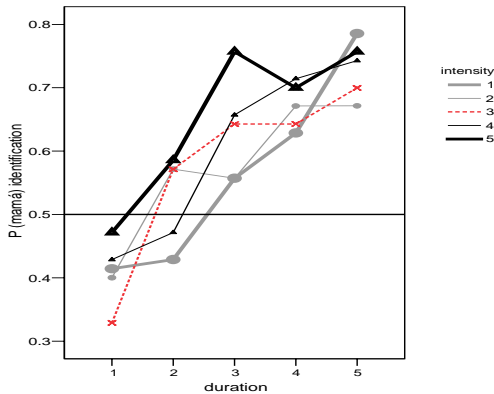
## 3. RESULTS

### 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. Moreover, when intensity is neutral (level 3), speakers rely only in duration to predict stress. However, 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 heavily rely

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.

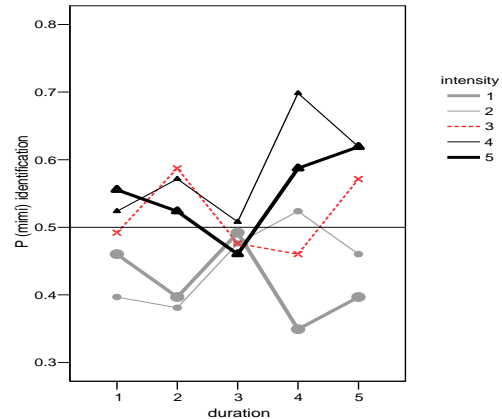


The graph in Figure 2 shows the probabilities of *mimí* judgements. In contrast with the results from *mamá*, subjects rely on intensity, not on duration, to predict stress. Including those stimuli where intensity and duration are in a competing relationship, intensity levels that cue oxytone ratios (levels 4 and 5) obtain higher scores than intensity levels that cue paroxytone stress (levels 1 and 2), indicating that listeners' judgments are based more on intensity cues than duration. Interestingly, when duration is neutral (level 3 in x-axes), all intensity levels scored around chance (.5), showing that speakers stopped using intensity cues when there were no duration differences. Thus, in contrast with [a] listeners rely more on intensity cues for vowel [i], and duration and intensity do not seem to be in an additive relationship. However, duration shapes intensity judgments in that intensity is not used when duration is neutral.

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]:  $F(4,32)=7.933, p<.0001$ ). Partial Eta-Square estimates showed that duration in [a] explained 63% of the variance, while in [i] it accounted for only 11%. This relation inverts for intensity, which explains 23% of variance in [a] and 50% of variance in [i]. Interactions were not significant, showing that duration and intensity are additive in [a]. Multiple comparisons between

intensity levels at neutral duration (stimulus 3) show that without duration cues, intensity has an effect on the prediction of stress only for vowel [a].

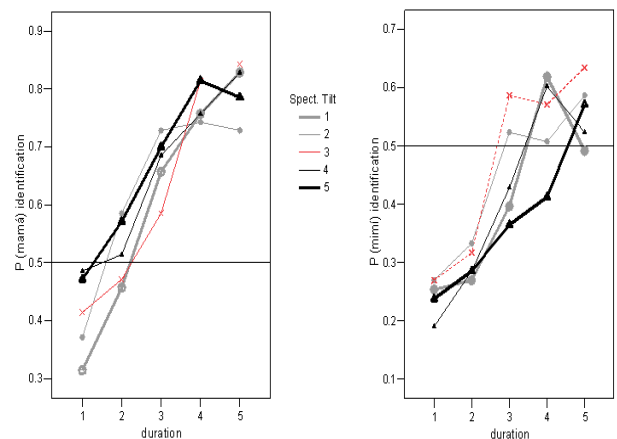
**Figure 2:** Probabilities of *mimí* responses for the 25 stimuli resulting from crossing the 5-levels of duration with the 5-levels of overall intensity in a *mimi-mimí* continuum.



### 3.2. Duration and Spectral Tilt

The two graphs in Figure 3 are similar to that of Figure 1 in that duration has a strong effect on the prediction of stress. Unlike Figure 1, the spread between the curves of the 5 spectral tilt levels is not wide. Thus, speakers rely more on duration than on spectral tilt when perceiving stress.

**Figure 3:** 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*:  $F(4,68)=7.034, p=.003, dur*vowel$ :  $F(4,68)=301, p=.872, intensity$ :  $F(4,68)=1.679, p=.211, int*vowel$ :  $F(4,68)=1.906, p=.165$ ). Partial Eta-Square Estimates indicate that duration explains 60% of the variance in [a] and 55% in [i].

A visual comparison of the four graphs shows that probability scores are consistently higher for [a] than they are for [i]. While 72% of the data scored above .5 for [a], less than 50% scored above .5 for [i]. Mean d-prime scores confirm that it was easier for speakers to detect oxytone words with vowel [a] than with vowel [i] (d-prime for [a]: 1.8, for [i]: 0.72).

#### 4. DISCUSSION

Our results demonstrate that if there is neither pitch accent information neither vowel reduction Spanish speakers detect stress contrasts on the basis of duration and overall intensity differences between adjacent syllables, while ignoring differences in spectral tilt. Thus, Spanish speakers' performance resembles that of Dutch speakers [8] in that they perceive stress through duration and intensity cues. These data show that at the lower levels of the prosodic hierarchy, stress has its own phonetic basis. And that in languages like Spanish and Dutch, cues to stress are independent of vowel reduction patterns.

However, Spanish speakers seem to differ from Dutch speakers in the type of intensity cues they use. While Dutch speakers rely on spectral tilt, Spanish speakers use overall intensity ratios, showing that they can be an effective cue to stress as well.

Above all, our results demonstrates that speakers' perception of stress is conditioned by vowel type: oxytone words can be easier detected when the stressed vowel is [a] than when it is [i]. This result may relate to the fact that speakers use both duration and overall intensity cues in an additive manner to perceive the stress contrast for vowel [a]; while for vowel [i], listeners use duration—and not very successfully—only if overall intensity is not an available cue in the signal. Production studies [10] back up this asymmetry. Duration differences for [i] had a reduced range, in comparison with other vowels, while overall intensity differences were similar across vowels. Moreover, no differences in spectral tilt were observed for any vowel. Thus, listeners' perception patterns reflect their knowledge of production. Spanish speakers take from the signal whatever cue is available, which depends on the vowel type. Thus, we cannot speak of correlates of stress in general because they are conditioned by the vowel.

There is one caveat, however. Even though duration was non-significant for vowel [i] while intensity was—and explained up to 50% of the variation in the data—it seems that listeners made judgments based on intensity only if there were duration differences between the syllables. This pattern needs further investigation.

#### 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 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.

#### 6. ACKNOWLEDGMENTS

Thank you to Eva Estebas and Ana Estrella for their help in the data collection. This research was supported by a grant of Ministerio de Educación y Ciencia, Spain.

#### 7. REFERENCES

- [1] 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.
- [2] Campbell, N., Beckman, M. 1997. Stress, prominence and spectral tilt. In: Botinis, A, Kouroupetoglou, G, Carayiannis, G. (eds), *Intonation: Theory, Models and Applications*. ESCA: 67-70.
- [3] Enríquez, E., Casado, C., Santos, A. 1989. La percepción del acento en español. *Ling. Esp. Actual* 11, 241-269.
- [4] 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.
- [5] Huss, V. 1977. English word stress in post-nuclear position. *Phonetica* 35, 86-105.
- [6] 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.
- [7] Sluijter, A., van Heuven, V. 1996. Spectral balance as an acoustic correlate of linguistic stress. *J. Acoust. Soc. Am.*, 100 (4), 2471-2485.
- [8] 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.
- [9] Solé, M.J. 1985. Experimentos sobre la percepción del acento. *Estudios de Fonética Experimental* 1, 131-242.
- [10] Ortega-Llebaria, M., Prieto, Pilar. Submitted. The phonetics of stress in Catalan and Spanish.