Acoustic Cues of Stress and Accent in Catalan*

Lluïsa Astruc & Pilar Prieto

University of Cambridge & ICREA-UAB
luisa@astruc.info, pilar.prieto@uab.es

Abstract

The goal of this paper is to examine the phonetic correlates of stress and accent in Catalan. We analyzed five acoustic correlates of stress (syllable duration, spectral balance, vowel quality, vowel pitch, and vowel intensity) in two stress conditions and in two accent conditions, which is to say, in stressed and unstressed syllables in both accented and unaccented environments (that is, appositions in sentences such as *Vol la vela, la vella* '(S)he wants the sail, the old sail' vs. right-dislocated subjects in *Vol la vela, la vella* '(S)he wants the sail, the old lady'. In keeping with the findings of Slujter & collaborators on Dutch and on English [1], [2], [3], and Campbell & Beckman on English [4], Catalan reveals systematic differences in the acoustic characterization of the accent and stress dimensions. Specifically, syllable duration, spectral balance, and vowel quality are reliable acoustic correlates of stress differences, while accentual differences are acoustically marked by intensity and pitch cues.

1. Introduction

Earlier work on Catalan stress has examined the correlates of stress in accented contexts, therefore suffering from covariation between the contribution of accent and stress [5], [6]. Work by Slujter & collaborators and Beckman & collaborators was among the first to examine the acoustic and perceptual correlates of stress without the confounding effects of accent [1], [2], [3]. They found that when a Dutch word is unaccented, the stressed syllable can still be distinguished both acoustically [1] and perceptually [3]. Results reveal that in the absence of an accent, stress is cued by duration, spectral balance differences, vowel quality, and intensity. Duration and spectral balance (that is, the difference in intensity between the upper and the lower region of the spectrum) were strong correlates of stress, while overall intensity was a cue of accent rather than of stress. Work on American English [2], [4] confirmed the main results found for Dutch, namely, that stressed syllables are longer than unstressed syllables, regardless of accent, and that overall intensity mainly cues accent, not stress. Recent work on Spanish [7] has shown that syllable duration, spectral balance, and vowel quality are reliable acoustic correlates of stress in this language. Thus, American English, Dutch, and Spanish differ basically in the use of vowel reduction and consonant reduction (flapping, aspiration) to mark stressed positions, but do not differ greatly in the way they use the other acoustic correlates (duration and spectral balance) to signal the presence of stress and accent.

The goal of this article is to examine patterns of duration, spectral balance, vowel quality, vowel pitch, and intensity that cue stress in accented and unaccented Catalan words. The accent contrast is found in minimal pairs of apposition and right-dislocated phrases such as *Vol la vela, la <u>ve</u>lla* '(S)he wants the sail, the old sail' (accented) vs. *Vol la vela, la vella* '(S)he wants the sail, the old lady' (deaccented). The advantage of using such controlled materials is twofold. First, they allow for a strict monitoring of the contribution of stress and accent and how they interact with each other. Second, the fact that both accented and unaccented environments are placed in the same position within the word and the utterance allows for a strict control of position effects such as final lengthening, and declination/downstep effects in F0 and in intensity. It is expected that, as in Dutch, English, and Spanish, we will find in Catalan systematic phonetic differences in the acoustic characterization along the accent and stress dimensions. We are especially interested in analyzing the role of intensity and spectral balance, as there are contradictory reports in the literature. Finally, we also ask whether the fact that Catalan has a phonological use of vowel reduction in unstressed positions might weaken the effect of other phonetic cues in the stressed/unstressed dimension.

2. Methods

2.1. Materials

The corpus contained eight near-minimal pairs of target words with a CVCV structure (*Mimi-Mimí, Lulu-Lulú, mama-mamà, Mila-Milà, Milu-Milú, Vila-Vilà, mula-muler, Mula-Mulà*). The target words were embedded in final position in similar carrier sentences, which were introduced by a disambiguating question. The two sentences had nearly identical segmental composition but different meaning and different intonation. An example of each is given in (1), with the target words underlined and stressed syllables in boldface.

(1)

a. Unaccented

-*Quin personatge vol fer la Mimi?* ('What character does Mimi want to interpret?')

-*Vol ser la protagonista, la* <u>**Mi**mi</u> ('She wants to interpret the main character, Mimi')

b. Accented

-*Quin personatge de l'obra t'agrada més?* ('Which character in the play do you like best?')

-*M'agrada la protagonista, la* <u>**Mi**mi</u> ('I like the protagonist, Mimi')

(1a) is a right-dislocated subject, a type of extra-sentential element which has been described as prosodically detached and deaccented in a variety of languages (for Catalan, and for a review, see [6], [8]). (1b) is an apposition and bears a rising accent (see also [8]).

Figure 1 shows the waveform, spectrogram, F0 trajectory, and segmentation tier of the accented and unaccented versions of the target word *Mimi*.

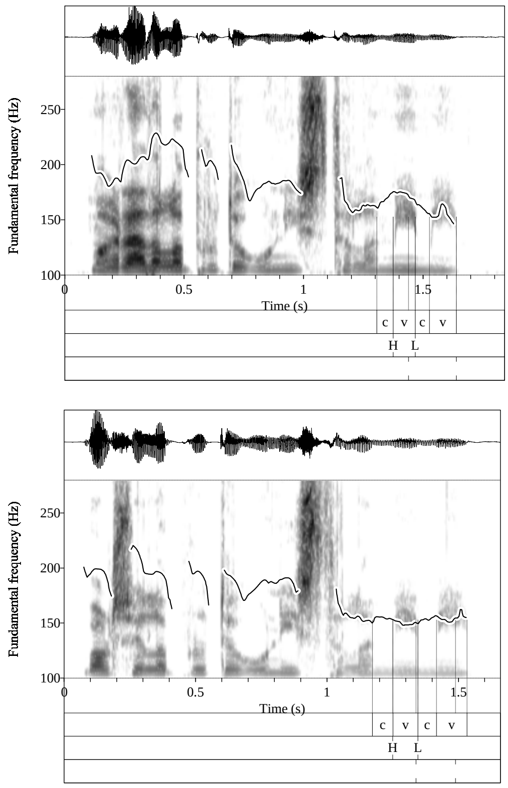

Figure 1. Waveform, spectrogram, F0 track, and segmentation tier of the accented (upper panel) and unaccented versions (lower panel) of the target word *Mimi* within the utterances *M'agrada la protagonista, la Mimi* (accented) and *Vol ser la protagonista, la Mim*i (unaccented)

Thus, the corpus provided materials for the following comparisons to be made.

	Apposition	Right-dislocated subject
[+stress]	[+accent] (a) M'agrada la protagonista, la **Mi**mi	[-accent] (c) Vol ser la protagonista, la **Mi**mi
[-stress]	[+accent] (b) M'agrada la protagonista, la **Mi**mí	[-accent] (d) Vol ser la protagonista, la **Mi**mí

Table 1: Target syllable ***mi*** (in bold) in four accent/stress contexts.

2.2. Subjects and Procedure

The participants were six native speakers of Central Catalan, all women. They were all teachers or researchers at Universitat Autònoma de Barcelona. Their ages ranged from 28 to 45 years. Each participant read the target utterances in the corpus three times, thus yielding 576 examples (8 target syllables x 2 accent conditions x 2 stress conditions x 3 repetitions = 96 examples x 6 speakers). Speakers were recorded at 44.1 kHz directly onto the computer in a sound-attenuated chamber at the Universitat Autònoma de Barcelona. They were instructed to read the target sentences naturally at an average voice level, using a Shure SM10A head-worn microphone to keep constant the distance between mouth and microphone. At the start of each recording, the experimenter monitored the speakers' voice level to ensure consistency between speakers. During the recordings, the experimenter checked that utterances received the right interpretation and the intended prosody. In some sporadic cases, some target utterances had to be repeated.

2.3. Acoustic measures

The sound files were analysed acoustically and instrumentally using Praat (4.3.09), [9], [10]. For the target word in each utterance, segment boundaries (CVCV, upper tier of Figure 1) and the highest and the lowest F0 point in each stressed syllable (H, L, middle tier of Figure 1) were marked by hand, using waveforms and spectrograms to guide the segmentation. In right-dislocated sentences, which were unaccented, H and L marks were placed at the onset and offset of both vowels. Finally, the F1 maximum for each vowel was automatically marked using a Praat script (lower tier of Figure 1). Measures of duration (in ms), pitch (in Hz), frequency of the formants (in Hz), spectral balance, and intensity (both in dB) were extracted automatically for both vowels of the target words. Following Slujter & van Heuven (1996), frequency (F0, F1, F2, F3) and intensity were measured at the peak of F1. Spectral balance was computed by extracting the amplitudes of four frequency bands (B1: 0-500 Hz, B2: 500-1000Hz, B3: 1000-2000Hz, B4: 2000-4000Hz). B1 contains F0 and B2, B3, and B4 contain F1, F2, and F3, the vowel formants. Finally, consonants were also measured to compute syllable duration.

3. Results

3.1. Pitch

In order to check whether appositions were consistently accented, the accentual pitch range of the stressed syllable was measured as the distance between H and L. In right-dislocations, pitch range was measured at the syllabic onset and offset. A one-way ANOVA revealed significant effects of the accent factor on pitch range ($F(1)= 147.534$; $p<.05$), confirming a significant pitch increase in the stressed syllables of appositions but not in those of right-dislocated sentences. The two graphs in Figure 2 plot the confidence intervals for the mean F0 values of the target syllable across the accent and stress conditions for all six speakers for V1 (left panel) and V2 (right panel).

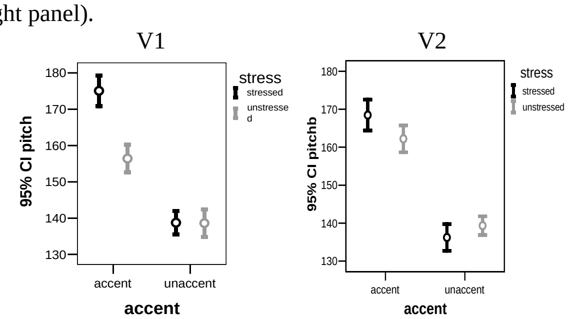

Figure 2. Mean F0 values and confidence intervals (in Hz) of V1 (left panel) and V2 (right panel) in different stress and accent conditions for all speakers.

As expected, accented syllables have a higher mean pitch than unaccented syllables, in both stress conditions. Furthermore, there is no effect of stress in the unaccented condition (multivariate *F*(1,5) = 1.116, *p*<.05, partial Eta squared = .182). Pitch is an acoustic correlate of accent, but not of stress.

3.2. Duration

The two graphs in Figure 3 plot the confidence intervals for the mean duration of target syllables 1 (left panel) and 2 (right panel) in different stress (stressed/unstressed) and intonation (accented/unaccented) conditions for all six speakers.

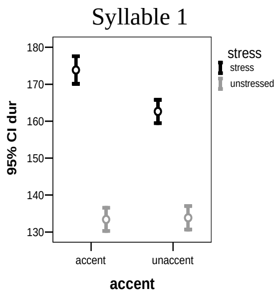

Figure 3. Mean syllable duration and confidence intervals (in ms) of syllable 1 and syllable 2 in different stress and accent conditions for all speakers.

The graphs show that stressed syllables (in black) are systematically longer than unstressed syllables (in grey), with a mean difference of 37.31 ms for syllable 1 and 30 ms for syllable 2, and most importantly, that this difference is maintained in unaccented environments. The duration difference is reduced in syllable 2 due to the effects of phrase-final lengthening. Accent also affects syllable the duration of stressed syllables. Moreover, in sentence-final position, there is a tendency to shorten unstressed syllables in accented words. A repeated measures ANOVA confirms that both factors have a significant effect on syllable duration, and that stress has a much larger and robust effect (multivariate *F*(1,5) = 265.77, *p*<.05, partial Eta squared = .982) than accent (multivariate *F*(1,5) = 11.320, *p*<.05, partial Eta squared = .694) on syllable duration. To further investigate the potential lengthening effects of the presence of an accent in stressed syllables, we undertook separate analyses of each speaker. The data revealed two different patterns depending on the group of speakers: while speakers 1, 2, and 4 have a consistent durational effect of accent, this was not the case for speakers 3, 5, and 6. ANOVA results confirm a significant effect of accent in the first group of speakers (and a non-significant stress-accent interaction) and a non-significant effect of accent in the second group of speakers (and thus a significant stress-accent interaction). Durational cues are consistent indicators of the presence of stress, but they are optional when it comes to indicating the presence of accent. This explains the partially contradictory results in the literature: while Slujter & van Heuven (1996a, 1996b) for Dutch (and English) found an additive effect of accent, Beckman & Edwards (1994:20-25) for English found that this pattern varied across speakers and speech rates.

3.3. Vowel Quality

Vowel quality is computed as the difference between F1-F0, F2-F1, and F3-F2. Vowels were analyzed separately for type ([i], [u], [a]) and for position (V1,V2). As expected, stressed vowels were found to have significantly higher F2-F1 differences than unstressed vowels. An univariate ANOVA confirms significant albeit modest effects of stress on F2-F1 (F(1) = 7.4555, *p*<.05, partial Eta squared = .032). A repeated measures ANOVA shows that stress has significant effects on the F1-F0 dimension for [a], in both V1 and V2 position, across the two accent conditions (multivariate *F*(1,5) = 12.003, *p*<.05 p= .018, partial Eta squared = .706). For the other vowels, neither accent (multivariate *F*(1,5) = .122, *p*>.05, partial Eta squared = .423) nor stress (multivariate *F*(1,5) = .005, *p*>.05 p, partial Eta squared = .001) have significant effects on the F1-F0, F2-F1, and F3-F2 dimensions, in any position. The graph in Figure 4 plots the confidence intervals for the mean F1-F0 difference in Hz for [a] in syllable 2 (where more tokens are available) in different stress (stressed/unstressed) and intonation (accented/unaccented) conditions for all six speakers.

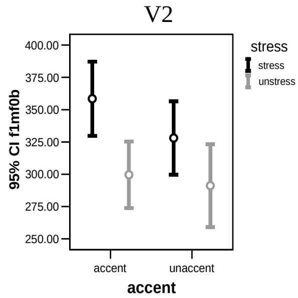

Figure 4. Mean difference between F1 and F0 and confidence intervals, in Hz, of [a] in syllable 2

Unstressed [a] shows a significant drop in F1, both in accented and deaccented environments. The F1-F0 difference thus constitutes a very robust acoustic cue of stress.

3.4. Overall Intensity

The two panels in Figure 5 plot the confidence intervals for the mean overal intensity (in dBs, measured at the peak of intensity) of V1, on the left, and of V2, on the right, across the accent and stress conditions for all six speakers.

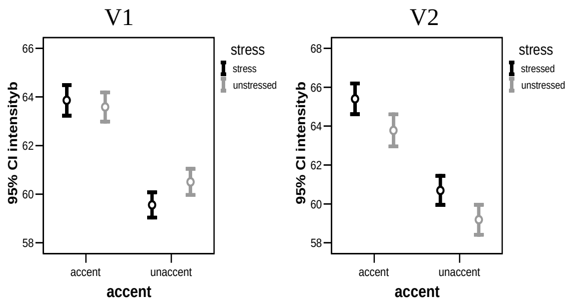

Figure 5. Mean intensity and confidence intervals (in dBs) for V1 (left panel) and V2 (right panel) in different accent and stress conditions for all speakers.

The two graphs show that within conditions stressed syllables are louder, but the presence of an accent is what makes a vowel significantly louder. In V1, stress interacts with accent: it has a significant effect on unaccented vowels (multivariate *F*(1,5) = 15.160, *p*<.05, partial Eta squared = .752), but not on accented ones (notice the overlapping confidence intervals for accented condition in the left-hand panel of Figure 5, which is confirmed by a repeated measures ANOVA (multivariate $F(1,5) = 3.93615$, $p>.05$, partial Eta squared = .440). In V2, the stress-accent interaction is not significant (multivariate $F(1,5) = 4.560$, $p>.05$, partial Eta squared = .477), just as the effects of stress are not significant ($F(1,5) = .221$, $p>.05$, partial Eta squared = .042), while the effects of accent are very robust and significant (multivariate $F(1,5) = 26.521$, $p<.05$, partial Eta squared = .841). Overall intensity is *not* a consistent cue of the presence of stress, neither in V1 nor in V2.

3.5. Spectral Balance

Spectral balance is computed as the differences in energy between the band containing F0 (B1, 0-500 Hz) and each of the bands containing the formants F1, F2, and F3 (B2, 500-1000 Hz; B3, 1000-2000 Hz; B4, 2000-4000 Hz). For the analysis, the vowels were separated according to vowel type ([a], [u], [i]) and position (V1, V2). [u] in V1 position and [a] in V2 position were among the vowels witth the largest number of tokens, and they were also the vowels that showed significant effects of stress on their spectral balance. The other vowels showed no significant effects, perhaps owing to an insufficient number of tokens. The graphs in Figure 6 show [a] in V2 position, the vowel with the largest number of tokens (6).

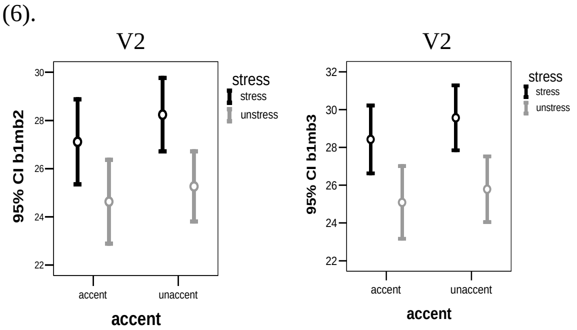

Figure 6. Mean spectral balance and confidence intervals (in Hz) computed as the difference between B1-B2 (left) and B1-B3 (right) for [a] in V2 position in different accent and stress conditions for all speakers.

For [a] (in V2 position) stress has a medium to large effect on the spectral balance between B1and B3 (B1 includes F0, B3 includes F2) (univariate F (1)= 7.756, p=.039, eta squared= .608) and also between B1 and B2 (B2 includes F1) (univariate F (1)= 9.992, $p<.05$, partial Eta squared= .666). For [u] (in V1 position), stress has a robust significant effect on the spectral balance between B1 and B3 (univariate F (1)= 23.039, $p<.05$, partial Eta squared= .882). The B1-B3 difference constitutes a very robust acoustic cue of stress.

4. Conclusions

The results reported in this article confirm a clear asymmetry between the behaviour of the acoustic correlates of stress and accent. While syllable duration, spectral balance, and vowel quality are reliable acoustic correlates of stress differences in Catalan, accentual differences are acoustically marked by cues such as pitch and overall intensity, and also optionally by vowel quality and vowel duration. These results, which are in keeping with previous results on Dutch, English, and Spanish [1], [2], [3], [4], [7], suggest that spectral balance and vowel reduction (cued by the F1-F0 difference) are more robust and systematic cue of stress than pitch and overall intensity. Importantly, our results reveal that, in Catalan, the presence of phonemic vowel reduction in unstressed position does not weaken the use of other acoustic correlates (duration and spectral balance) to signal the presence of stress.

Our results are relevant to current research on speech synthesis. Modern text-to-speech systems use statistically-trained models to predict prosodic targets for synthesis. The performance of these models is heavily dependent upon the linguistic and acoustic features in the data which is used to train them. Studies of the acoustic correlates of prosodic phenomena are valuable indicators of the types of features in the training data that speech synthesis researchers should be paying attention to when developing and training prosodic models.

**Acknowledgments*: We are grateful to the informants who participated in the production experiment (A. Abella, E. Bonet, T. Cabré, A. Gavarró, M. Mata, M. Llinàs, P. Prieto). The second author was recorded for the pilot study, but since her results did not differ from those of the other speakers, they were also included in the analysis. Thanks go to E. Ferragne and M. Ortega-Llebaria for kindly making available to us their Praat scripts, and to D. Green for useful suggestions about the practical application of this research. Grants 2002XT-00032, 2001SGR 00150, and 2001SGR 00425 from the Generalitat de Catalunya and BFF2003-06590 and BFF2003-09453-C02-C02 from the Spanish Ministry of Science and Technology, have funded this research.

5. References

[1] Slujter, A.M.C.; van Heuven, V., 1996a. Spectral balance as an acoustic correlate of linguistic stress. *JASA* 100(4), 2471-2485.

[2] Slujter, A.M.C.; van Heuven, V., 1996b. Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceedings of ICSLP*, 96, 630-633.

[3] Slujter, A.M.C.; van Heuven, V.; Pacilly, J.J.A., 1997. Spectral balance as a cue in the perception of linguistic stress. *JASA* 101 (1), 503-513.

[4] Campbell, N.; Beckman, M., 1997. Stress, prominence and spectral balance. In *Intonation: Theory, Models and Applications*, A. Botinis et al (eds.). ESCA and University of Athens, 67-70.

[5] Barnils, P. 1933. Sobre la quantitat de les vocals tòniques. Escrits de Pere Barnils, *Anuari de l' Oficina Romànica de Llengua i Literatura*, VI, 67-78.

[6] Recasens, D. 1986. *Estudis de fonètica experimental del català oriental central*. Publicacions de l'Abadia de Montserrat: Barcelona.

[7] Ortega-Llebaria, M.; Prieto, P. 2005. Acoustic correlates of stress and accent in Spanish. Paper presented at *PaPI, Phonetics and Phonology in Iberia*, Belllaterra, June 2005.

[8] Astruc, L. 2005. *The intonation of extra-sentential elements in Catalan and English*. Ph.D. Dissertation, University of Cambridge.

[9] Boersma, P., & Weenink, D. (2005). Praat: doing phonetics by computer (Version 4.3.01) [Computer program]. Retrieved from http://www.praat.org/

[10] Wood, S. (2005). Praat for beginners [Manual]. Retrieved: http://www.ling.lu.se/persons/Sidney/praate/