Prosodic manifestations of the Effort Code in Catalan, Italian and Spanish contrastive focus

Maria del Mar Vanrell, Antonio Stella, Barbara Gili Fivela and Pilar Prieto

Journal of the International Phonetic Association / Volume 43 / Issue 02 / August 2013, pp 195 - 220
DOI: 10.1017/S0025100313000066, Published online: 05 July 2013

Link to this article: http://journals.cambridge.org/abstract_S0025100313000066

How to cite this article:

Request Permissions: Click here
This paper investigates the relevance of three prosodic parameters (alignment, duration and scaling) in the conveyance of contrastive focus in Catalan, Italian and Spanish. In particular, we seek to determine how the Effort Code is instantiated in the expression of contrastive focus in both production and perception. According to the Effort Code, putting more effort into speech production will lead to greater articulatory precision (de Jong 1995, Gussenhoven 2004) and this is related to the expression of focus in the sense that wider pitch excursions will be used to signal meanings that are relevant from an informational point of view. A dual production and perception experiment based on an identification task was conducted. Results for the production part show that contrastive focus accents have earlier peaks for all three languages but f0 peaks are systematically lower only in Italian. Syllables bearing the contrastive focus accents are also longer in the three languages. Regarding the results for the perception part, converging evidence is found not only for an active perceptual use of the three prosodic parameters present in production but also for language-specific preferences for particular prosodic parameters.

1 Introduction

According to Gussenhoven (2002), intonational meaning resides in two different components of language, namely the phonetic implementation component and the intonational grammar. In his view, the phonetic implementation component can be directly related to the idea of
universal meanings and, consequently, to the presence of the so-called ‘biological codes’, which have been defined as variations in pitch that are ethological in origin (Ohala 1983, 1984). In the paralinguistic domain, three different biological codes have been proposed. The Frequency Code was proposed by Ohala (Ohala 1983, 1984, 1994) and has to do with the correlation existing between the size of the larynx and the rate of vibration of the vocal cords in the expression of dominance relationship. Thus, a higher pitch tends to be associated with a smaller organ of production while a lower pitch signals that the larynx is larger. Some of the affective meanings related to the Frequency Code include submissiveness versus dominance, vulnerability versus confidence, aggressiveness versus friendliness, etc. As for the informational interpretation of the Frequency Code, higher pitch reflects uncertainty whereas low pitch is related to certainty (and, in turn, questioning versus asserting). In the Production Code (Lieberman 1967, Gussenhoven 2002), the subglottal air pressure necessary for speech production is identified with the exhalation phase of the breathing process. Speakers are more energetic at the beginning of the utterances, and a decrease in subglottal pressure will correlate with a gradual drop in intensity and a lowering of the fundamental frequency. Hence, high pitch will be associated with the beginning of utterances and low pitch with utterance endings. The informational interpretation of the Production Code is related to initiation and finality. According to Gussenhoven (2002, 2004), high beginnings signal new topics and low beginnings continuations of topics. With respect to endings, the reverse pattern is found: ‘high endings signal continuation, low endings finality and end of turn’ (Gussenhoven 2002: 51). Finally, in accordance with the Effort Code (de Jong 1995, Gussenhoven 2002), a wider excursion of the pitch movement will lead to an increase in the importance of the message from an informational point of view. In other words, the interpretation is that the speaker is being forceful because s/he believes the contents of her/his message to be important. Thus, the grammaticalisation of the informational interpretation of the Effort Code is typically found in the expression of focus. In these cases, focalised constituents will tend to appear with wide pitch excursions (Gussenhoven 2002, 2004). Affective interpretations of the Effort Code include surprise and helpfulness. In addition, according to Baumann et al. (2007) the manifestation of the Effort Code indicating increased emphasis can be expressed at levels other than pitch range such as peak timing (as also pointed out by Gussenhoven 2002), duration of segmental domains like syllables or feet, and articulatory effort in vowels. Thus, the Effort Code can be realised by speakers to express contrastive focus through pitch range and duration, but they can also resort to either peak height or peak delay as an alternative to pitch range.

Although the proponents of the biological codes (Ohala 1983, 1984, 1994; Gussenhoven 2002, 2004) have assumed that paralinguistic meaning is universal since it is biologically-based and that it is generally true that languages tend to follow the form–function relations found in the paralinguistic domain, they also point out that it is important to keep in mind that grammaticalisation may involve unnatural and arbitrary forms (Gussenhoven 2002). Evidence for these mismatches is found in languages which mark statements through final rises, as in Belfast English (Jarman & Cruttenden 1976) or Chickasaw (Gordon 1999), or in languages which mark questions by means of falls, as in Roermond Dutch (Gussenhoven 2000). Yet another piece of evidence that some languages do not always follow the general biological patterns is the fact that not all languages express prominence through an increase in pitch. For example, languages such as Akan (Kügler & Genzel 2012) express contrastive focus through lower realisations of both high and low tones. In this study, we will address the question of how the informational interpretation of the Effort Code is found in the expression of contrastive focus in closely related languages (Catalan, Italian and Spanish).

1 Related research on the expression and recognition of emotions (van Bezooijen 1984, Scherer 2000, Chen, Gussenhoven & Rietveld 2004) has demonstrated that although there are universal cues for emotion, it is also possible to find language-specific patterns.
It is well known that languages differ in the ways that they express narrow contrastive or corrective focus, that is, in the way they mark a constituent that is the direct rejection of an alternative (i.e. ‘it is B, not A’). According to Elordieta (2007), a cross-linguistic typological distinction can be established depending on whether contrastive focus is marked (a) by means of intonation, (b) via syntactic movement optionally combined with prosodic mechanisms to ensure prominence, or (c) by using specific morphemes also optionally combined with syntactic displacement, prosodic marking or a combination of both. In languages such as English or German the intonational prominence may be shifted to different positions in the clause while the syntactic structure remains the same. On the other hand, languages such as Catalan, Italian or Spanish use syntactic operations for focus marking (Vallduví 1991, Ladd 1996), although recent studies show that in these languages intonational strategies are as acceptable as syntactic ones (Estebas Vilaplana 2009 for Catalan; Face & D’Imperio 2005 for Italian and Spanish), especially when word order is kept invariant. With respect to prosody, the following prosodic cues have been shown to distinguish contrastive accents from non-contrastive accents (henceforth C and NC accents, respectively) in a number of languages: (a) alignment of the peak of the C/NC accent with respect to segmental landmarks such as the onset or offset of the accented syllable; (b) pitch range (also pitch scaling of the peak) of the C/NC accent; and (c) duration of the syllable bearing the C/NC accent. As far as alignment is concerned, crosslinguistic evidence has shown that C accents are characterised by the use of retracted pitch peaks (de la Mota 1995 and Beckman et al. 2002 for Spanish; Estebas Vilaplana 2009 for Central Catalan; D’Imperio 2002 and Gili Fivela 2002 for Neapolitan and Pisa Italian, respectively; Smiljanic 2004 for Serbo-Croatian; Manolescu, Olson & Ortega Llebaria 2009 for Romanian). On the other hand, pitch range (or tonal scaling) as a prosodic marker of contrastive focus across languages is a controversial issue. In languages such as Romanian (Manolescu, Olson & Ortega Llebaria 2009) and Serbo-Croatian (Smiljanic 2004), contrastive focus leads to greater pitch excursion on the accented syllable, while for Italian and Spanish it is a matter of debate among researchers whether f0 pitch range increase is or is not an acoustic correlate of contrastive focus (D’Imperio 2002 and Gili Fivela 2005, 2006 for Italian; de la Mota 1995 and Face 2001, 2002 for Spanish). The importance of duration in the conveyance of contrastive focus has been documented in studies such as those by Eady et al. (1986) for English, Jun & Lee (1998) for Korean, and Baumann et al. (2007) and Kügler (2008) for German. Although it is apparent from the above-mentioned studies that languages employ different prosodic strategies to convey the contrast between C and NC accents, very few studies have analysed the specific contribution of each parameter as well as the interplay between them in a systematic way.

The present study aims to explore the prosodic manifestations of the Effort Code in the expression of contrastive focus (peak alignment, duration and pitch range) as well as the interplay between these prosodic cues in Catalan, Italian and Spanish. Previous research conducted on the contrast between initial C and NC accents for Catalan (Prieto, D’Imperio & Gili Fivela 2005) and Castilian Spanish (de la Mota 1995; Nibert 2000; Face 2001; Hualde 2002).
2002, 2003) concluded that the main difference is based on a different alignment of the peak (NC accents have late peaks, L+>H*, while C accents have earlier f0 peaks, L+H*, see Figure 1, left panels). Likewise, in the varieties of Italian spoken in both Pisa (Gili Fivela 2006, 2008) and Lecce (Stella & Gili Fivela 2009), NC initial accents have late peaks (L+H*), while C accents have earlier f0 peaks (H*+L) (see Figure 1, right panels). Although previous works on Catalan, Italian and Spanish have shown that syllables bearing C accents tend to be longer than syllables bearing NC accents (for Catalan, see Estebas Vilaplana 2009, for Italian, Gili Fivela 2005, 2006; for Spanish, de la Mota 1995), it remains unclear what the specific contribution is of peak height of the syllable bearing the contrastive accent as well as how duration and peak height interact with peak timing in both production and perception data. According to the Effort Code and based on previous research conducted on this topic in the three languages under study, the hypothesis would be that since the three languages seem to be characterised as having retracted peaks for contrastive focus, one effect of the Effort Code will be the use of higher peaks or accents with a wider pitch span. Longer duration of the syllables bearing the contrastive accents might also be used as an enhancement of peak height.

In this study on the prosodic manifestations of the Effort Code, we will also consider the interface between production and perception in intonation. Most of the speech perception theories can be grouped into two categories according to the listener’s cognitive intervention: there are, on the one hand, those in which the intervention of the listener in the decoding process is crucial (active models) and, on the other hand, those for which the cognitive intervention of the listener is little or nil (passive models) (Tatham & Morton 2006). In other words, what distinguishes these theories is the role of the listener in the decoding process. However, they all assume that (a) the perception of speech consists of labelling the acoustic signal with an appropriate symbolic representation based on the phonology of the specific language, and (b) this process of transformation between the acoustic signal and the symbolic representations is not so straightforward since the acoustic signal tends to present some deformations which must be repaired by auditory perception (Ru, Chi & Shamma 2003, Tatham & Morton 2006). Thus, we would expect to find a correspondence between production and perception in the decoding of contrastive focus in speakers of different languages. Our hypothesis is that the correspondence between production and perception would go beyond a mere sum of the individual properties of source and receiver and would point to a synergy, a maximisation of

---

3 The symbol ‘>’ indicates that the peak is delayed with respect to the end of the accented syllable, that is, it is aligned with the postaccentual syllable (Prieto et al. 2005).

4 Production and perception experiments of the contrastive pattern across varieties of Italian have yielded different phonological analyses, such as H*+L vs. L+H* (or H*) L-. For instance, this is true for Pisa and Florence Italian, where the low target following the peak appears to behave differently (in particular, it is stable with respect to the peak, independently of the postaccentual material available, or not; see Avesani 2003, Gili Fivela 2002). However, as discussed in Gili Fivela et al. (in press), the two options may be found also in relation to other varieties, depending on the phonetic shape of the pitch patterns, albeit no explicit experimental investigation has thus far been carried out (for discussion see Gili Fivela et al. in press).
Effort Code in Catalan, Italian and Spanish contrastive focus

In order to test these hypotheses, we conducted a production experiment in which the data obtained for each of the two focal conditions and for the three languages respectively were analysed acoustically. In the second experiment, we tested the specific contribution of tonal alignment, duration and tonal scaling through a perception experiment based on an identification task. This study is innovative because it seeks to examine not only the specific prosodic cues that convey contrastive focus in the three languages, but also their respective perceptual relevance. From an empirical point of view, it will also add to the current body of research on specific varieties of Catalan (Majorcan) and Italian (Lecce), languages that have not been studied previously in the field of focus, while allowing us to compare different languages following a uniform controlled methodology.

The paper is organised as follows. Section 2 presents the methodology (participants, materials and procedure) and the results for the production experiment. Section 3 presents the methodology and results for the perception experiment, that is, the identification task and the Curve Estimation analysis. Finally, Section 4 discusses the major findings of this research and states its overall conclusions.

2 Experiment 1: Production experiment

The aim of this experiment was to investigate the prosodic features of alignment, pitch range and pitch scaling, and duration of the initial accents in C and NC utterances in Catalan, Italian and Spanish. To this end a corpus was designed which contained declarative sentences with C and NC initial accents with two different stress patterns (paroxytones and proparoxytones). Five native speakers of each language participated in this production experiment, reading a total of 300 sentences per language.

2.1 Methodology

2.1.1 Participants

Five native speakers of Majorcan Catalan (three women, two men) aged between 20 and 36 years (mean = 25.8 years), five native speakers of Lecce Italian (three women, two men) aged between 22 and 29 (mean = 26.4) and five native speakers of Madrid Spanish (three women, two men) aged between 35 and 45 (mean = 38.4) participated in this experiment. All participants had been born and raised in their respective geographic regions. They reported having little or no prior training in phonology/phonetics and no history of hearing disability.

2.1.2 Experimental materials and elicitation method

The corpus used in this study contains declarative sentences with NC or C accents in initial position. Sentences bearing initial NC accents are carriers of new information in which there is no particular constituent that is focalised, which is equivalent to saying that the entire expression is the focus constituent. On the other hand, contrastive focus ‘marks a constituent
that is a direct rejection of an alternative’, which implies the exclusion of an alternative (Gussenhoven 2007: 91). For instance, when the sentence *Mary is coming* is pronounced as an answer to the question *What’s happening?*, the entire sentence is new information with no specific element emphasised. However, when the same sentence is an answer to the question *Is Peter coming?*, *Mary* is highlighted and is in contrastive focus (Manolescu et al. 2009).

To elicit the target sentences in our production experiments, the experimenters initially posed a *wh*-question to the speaker which asked for completely new information (*What did they tell you?*). The speaker was then asked to read a sentence bearing the NC accent (i.e. *That Marina is coming tomorrow*). At that point the researcher used a question challenging the proposition which rhymed or was similar to the NC target word (i.e. *Tina?*) so that the C sentence (i.e. *No, MARINA is coming tomorrow*) was more naturally produced by the speaker although it was presented in writing too. Importantly, the contrastive word was presented in capital letters which is a sort of orthographic convention indicating emphasis. Sentences bearing the target words had a simple syntactic structure, namely a subject NP, an attribute AdjP or an adverbial complement AdvP or PP, and a verbal VP, each consisting of just one prosodic word (see Table 1 for one example of the target sentences used for each language grouped according to stress pattern). Importantly, the experimenters were native speakers of the language varieties investigated here.

Ten different target words per language were used (five paroxytones and five proparoxytones). Whenever possible, the target words were composed of voiced consonants (to avoid segmentally-induced effects on the f0 curve). The target words always contained the stress on the penultimate or antepenultimate syllable to provide more room for tonal accent realisation and all consisted of three syllables (see Table 2 for two of the target words used for each language and the corresponding stress pattern). The choice of the specific words was determined by the following two criteria applied to the extent possible: (a) at least the vowel of the accented syllable as well as the syllable structure of the syllable bearing the focus accent should be the same across the three languages, and (b) target words should be high frequency words so as to limit the likelihood of pronunciation problems.

Five speakers per language read three repetitions of the two blocks (paroxytone and proparoxytone target words) for a total of 300 sentences per language (5 distinct words-sentences × 2 focal conditions × 2 stress patterns × 3 repetitions × 5 speakers). The series of question–answer pairs were presented through a PowerPoint presentation at random irrespective of the stress pattern. Speakers were asked to read the sentences at a normal speech rate and avoid marked readings, pauses or speech disfluencies. When a break or irregularity occurred, the question–answer pair was presented a second time at the end of the full experimental session and the subject’s response recorded again. The contrastive prosody did not present problems of inducement.

Catalan speakers were recorded in a quiet room at the Universitat Pompeu Fabra and Spanish speakers were recorded under similar conditions at the Universidad Autónoma de Madrid using a Marantz Professional PMD660 digital recorder and Rode NTG-2 microphone. Italian speakers were recorded in a sound-proofed room at the Centro di Ricerca Interdisciplinare sul Linguaggio in Lecce using a Shure SM-86 condenser microphone and Goldwave 5.25 software installed on a desktop computer equipped with an Edirol UA-5 external sound card. The sentences were digitised at a 44100 Hz sample rate and 16 bit amplitude resolution.

### 2.1.3 Analysis

The recordings were processed using Praat (Boersma & Weenink 2009). The intonational patterns of the utterances were analysed manually by the authors by means of auditory analysis and inspection of the f0 traces. Spanish speaker 4 had a low f0 with creaky voice and this made it difficult for Praat to collect L points. As a consequence, five utterances by this subject were eliminated from the database. The final database for Spanish thus contained a total of 295 utterances. The other two languages contained 300 examples each.
Table 1  One example for each language of the series of question-answer pairs used in this experiment grouped according to stress pattern (paroxytones and proparoxytones). Target words are marked in boldface.

<table>
<thead>
<tr>
<th>Sentences bearing NC accents</th>
<th>Sentences bearing C accents</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Paroxytone target words</strong></td>
<td><strong>Paroxytone target words</strong></td>
</tr>
<tr>
<td>Catalan</td>
<td>Catalan</td>
</tr>
<tr>
<td>Q: Què t’veu dit?</td>
<td>Q: Na Tina?</td>
</tr>
<tr>
<td>‘What did they tell you?’</td>
<td>‘PERS ARTICLE Tina [proper name]?’</td>
</tr>
<tr>
<td>A: Que no Marina vendrà demà.</td>
<td>A: Na Marina vendrà demà.</td>
</tr>
<tr>
<td>Italian</td>
<td>Italian</td>
</tr>
<tr>
<td>Q: Hoi saputo qualcosa?</td>
<td>Q: La Sabina?</td>
</tr>
<tr>
<td>‘Have you found out anything?’</td>
<td>‘PERS ARTICLE Sabina [proper name]?’</td>
</tr>
<tr>
<td>A: Si, la Marina verrà domani.</td>
<td>A: No, la Marina verrà domani.</td>
</tr>
<tr>
<td>Spanish</td>
<td>Spanish</td>
</tr>
<tr>
<td>Q: ¿Qué te han dicho?</td>
<td>Q: ¿Tina?</td>
</tr>
<tr>
<td>‘What did they tell you?’</td>
<td>‘¿Tina [proper name]?’</td>
</tr>
<tr>
<td>A: Que Marina vendrá mañana.</td>
<td>A: Marina vendrá mañana.</td>
</tr>
<tr>
<td><strong>Proparoxytone target words</strong></td>
<td><strong>Proparoxytone target words</strong></td>
</tr>
<tr>
<td>Catalan</td>
<td>Catalan</td>
</tr>
<tr>
<td>Q: Què t’veu dit?</td>
<td>Q: Sa màquina?</td>
</tr>
<tr>
<td>‘What did they tell you?’</td>
<td>‘The machine?’</td>
</tr>
<tr>
<td>A: Que la l’amina no està acabada.</td>
<td>A: Sa l’amina no està acabada.</td>
</tr>
<tr>
<td>Italian</td>
<td>Italian</td>
</tr>
<tr>
<td>Q: Hoi saputo qualcosa?</td>
<td>Q: La pagina?</td>
</tr>
<tr>
<td>‘Have you found out anything?’</td>
<td>‘The page?’</td>
</tr>
<tr>
<td>A: Si, la lamina è piegata.</td>
<td>A: No, la lamina è piegata.</td>
</tr>
<tr>
<td>Spanish</td>
<td>Spanish</td>
</tr>
<tr>
<td>Q: ¿Qué te han dicho?</td>
<td>Q: ¿La máquina?</td>
</tr>
<tr>
<td>‘What did they tell you?’</td>
<td>‘The machine?’</td>
</tr>
<tr>
<td>A: Que la l’amina no està acabada.</td>
<td>A: La l’amina no està acabada.</td>
</tr>
</tbody>
</table>

Table 2  Examples of two of the target words used for each language grouped according to stress pattern (paroxytones and proparoxytones).

<table>
<thead>
<tr>
<th>Paroxytone target words</th>
<th>Italian</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalan</td>
<td>Marina [maˈrina]</td>
<td>Marina [maˈrina]</td>
</tr>
<tr>
<td></td>
<td>(proper name, person)</td>
<td>(proper name, person)</td>
</tr>
<tr>
<td>bombona [bomˈbona]</td>
<td>remone [reˈmone]</td>
<td>bombona [bomˈbona]</td>
</tr>
<tr>
<td>‘gas cylinder’</td>
<td>‘big ear’</td>
<td>‘gas cylinder’</td>
</tr>
<tr>
<td>Proparoxytone target words</td>
<td>lamina [ˈlamina]</td>
<td>lamina [ˈlamina]</td>
</tr>
<tr>
<td></td>
<td>‘sheet’</td>
<td>‘foil’</td>
</tr>
<tr>
<td>(proper name, city)</td>
<td>‘monk’</td>
<td>(proper name, city)</td>
</tr>
</tbody>
</table>
Six segmental and two tonal targets were labelled by hand by the first and second author, using the following segmental and tonal labels:

**Segmental labels**
- Onset of the target utterance (o)
- Onset of the first syllable of the target word (osy1)
- Onset of the syllable with which the C/NC accent is associated (oasy)
- Offset of the syllable with which the C/NC accent is associated (ofasy)
- Offset of the target word (ofw)
- End of the utterance (e)

**Tonal labels**
- Initial f0 valley value associated with the C/NC accent (L)
- f0 peak value associated with the C/NC accent (H)

Measurements corresponding to these labels were collected automatically into a file in .txt format through a Praat script. With respect to f0 timing, two measures of H and L alignment were used for purposes of subsequent statistical exploration: (i) temporal distance (in ms) from the L point to the onset of the syllable bearing the C/NC accent (see Figure 2, left panel), and (ii) temporal distance (in ms) from the H peak to the end of the syllable bearing the C/NC accent (see Figure 2, central panel). Pitch range was measured as the distance in semitones between the peak (H) and the preceding low (L) target (see Figure 2, right panel).

Figure 3 illustrates the orthographic, segmental and prosodic transcription of the utterance ‘Marina will come tomorrow’ produced with NC and C meaning for each of the three languages under study.7 The first horizontal tier contains the orthographic transcription. The

---

7 As suggested by one of our reviewers, the very high peak on NC accents for Italian could be due to a putative reading style effect, leading to a possible realization of a L+H* followed by an H- phrase accent. In order to discard this possibility, an additional perception experiment for the three languages was performed. Six native speakers of each variety participated in the experiment. The materials consisted of four sentences (two of the sentences had paroxytones as target words and the other two sentences had proparoxytones) produced by three different speakers with a contrastive or a non-contrastive intonation. Thus, we had a total of 24 sentences for each language and listener (2 sentences × 2 focus types × 2 stress patterns × 3 speakers). Listeners were asked to identify the sentences either as a neutral statement, a corrective statement or a statement with another function. In addition, participants were asked to express the degree of naturalness of the sentences on a seven-point scale. The results showed that the listeners were able to identify both meanings accurately. Thus, Catalan listeners identified NC statements correctly 91% of the time and C statements 97% of the time. For Italian listeners, the percentages were a little bit lower but still well above chance level: 88% for NC accents and 91% for C accents. The results for
Figure 3 Waveform, spectrogram, f0 contour and orthographic, segmental and prosodic transcription for the NC utterance (left) and the C utterance (right) for Catalan (upper panel), Italian (center panel) and Spanish (lower panel).
second tier marks the segmental labels explained above, whose number can vary depending on the stress pattern of the target word. Finally, the third tier contains the tonal information regarding the C/NC accent.

These are the criteria that were followed in order to determine the tonal landmarks:

- **Initial f0 valley value in the C/NC accent.** In the minima, since it corresponds to the onset of the rise, the L was located at the point where the f0 rising velocity increases. This roughly corresponded in Catalan and Spanish with the onset of the syllable bearing the focal/non-focal accent while in the case of Italian the L target could occur before the accented syllable.

- **f0 peak value in the C/NC accent.** The peak was generally easy to pinpoint and was situated just before the fall. Sometimes a plateau was found. When this happened, the H was placed at the end of the plateau, immediately before the contour started to fall.

### 2.1.4 Statistical analyses

In this experiment repeated measures ANOVA tests were performed because we had a situation in which the same people took part in different experimental conditions (as opposed to different people taking part in different experimental conditions). Various measures were used as dependent variables for statistical exploration, namely (a) H alignment (i.e. distance in ms from the H tone to the end of the accented syllable), (b) L alignment (i.e. distance in ms from the L tone to the start of the accented syllable), (c) pitch range of the C/NC accent (distance in st between the peak and the preceding low target), (d) H scaling of the C/NC accent (i.e. value in Hz of the H tonal target), (e) L scaling of the C/NC accent (i.e. value in Hz of the L tonal target) and (f) duration of the accented syllable. For each one of the dependent variables, we ran a repeated measures ANOVA with focal condition (two levels) as the within-subject factor and language (three levels) as the between-subject factor across subjects. The results from the repeated measures ANOVA tests include the $F$-ratio, the degrees of freedom from which it was calculated and the significance value.

### 2.2 Results

#### 2.2.1 Alignment results

Figure 4 shows the box plots of the distance from H to the end of the accented syllable in milliseconds for NC (dark grey boxes) and C (light grey boxes) accents for all speakers of Catalan (top left), Italian (top right) and Spanish (bottom). Observe that there is a substantial difference between median points of NC and C accents for Catalan and Spanish on the one hand and Italian on the other. Thus, in the case of Catalan and Spanish the peak for C accents is realised at the end of the accented syllable (with the exception of Spanish speaker 5) while for NC accents the peak is systematically aligned after the end of the accented syllable. For Spanish revealed a correct identification percentage of 91% for NC sentences and 95% for C sentences. As for the naturalness rating, listeners agreed that C statements were more natural than NC statements but still tended to give high scores (5 or 6 on a seven-point scale) for NC utterances. These results clearly rule out the existence of an L+H* H- pattern since even though the listeners could base their decisions on three possible responses (neutral statement, corrective statement or a statement with another function), they primarily opted for the neutral statement or the corrective statement answers. They also assigned a high naturalness score, showing that they judged stimuli as instances of a spontaneous rather than a read speech style.

8 In the specific case of H Alignment as dependent variable, stress pattern (two levels) was also used as the within-subject factor since graphs with the data split into paroxytone and proparoxytone words showed important effects of H tonal alignment for Catalan and Spanish NC utterances. Moreover, previous research on H tonal alignment for Catalan and Spanish (Prieto et al. 1995, 2010; de la Mota 2005; Simonet & Torreira 2005; Simonet 2006) showed that H peaks are affected by the position of the accented syllable with respect to the end of the word.
Italian, the situation is a little bit different since both NC and C accent peaks are aligned before the end of the accented syllable. However, in the three languages the C peaks are aligned always earlier than NC peaks. The repeated measures ANOVA analysis with H alignment as the dependent variable revealed statistically significant effects of focal condition ($F(1,4) = 154.248, p < .0001$) and language ($F(2,8) = 80.380, p < .0001$). Post-hoc tests with the Bonferroni Adjustment comparing means of H alignment for NC and C accents across languages showed that each of the languages differs from the others but that Italian differs far more than the other two. Thus, the differences between Italian and Catalan are statistically significant at $p < .01$ and the differences between Italian and Spanish are significant at $p = .001$.

As for the effect of the position of the stress on the peak timing, it is observed that in Catalan and Spanish NC sentences the peak aligns later in proparoxytones, that is, the distance between the peak and the beginning of the accented syllable is greater in words with the stress on the antepenultimate syllable than in words with penultimate stress. This effect was not seen in C sentences for any of the three languages or in Italian NC sentences. Such a
finding is not new in the literature since it has been attested by different scholars for Catalan and Spanish (Prieto, van Santen & Hirschberg 1995, de la Mota 2005, Simonet & Torreira 2005, Simonet 2006, Prieto, Estebas Vilaplana & Vanrell 2010). In order to verify whether this effect of stress pattern on H alignment was statistically significant, a repeated measures ANOVA test was performed with H alignment as the dependent variable and focal condition (two levels) and stress pattern (two levels) as the within-subject factor and language (two levels, only Catalan and Spanish) as the between-subject factor across subjects. The results confirmed statistically significant effects of focal condition \( (F(1,4) = 165.625, p < .0001) \), stress pattern \( (F(1,4) = 72.223, p = .001) \), language \( (F(1,4) = 18.032, p < .05) \) and a stress pattern by focus interaction \( (F(1,4) = 44.048, p < .01) \).

With respect to the distance from L to the start of the accented syllable, the L point is located for the three languages near the start of the accented syllable for both NC and C accents. These results confirm previous studies reporting that the L point is affected little or not at all by focal conditions (Estebas Vilaplana 2009 for Central Catalan; Gili Fivela 2005, 2006 for Pisa Italian; de la Mota 1995 for Spanish). The repeated measures ANOVA analysis with L alignment as the dependent variable revealed no statistically significant effects of focal condition \( (F(1,4) = 1.725, p > .05) \) but significant effects of language \( (F(2,8) = 8.689, p = .01) \). The interaction focal condition \( \times \) language was not significant \( (F(2,8) = .664, p > .05) \) and the post-hoc tests with the Bonferroni adjustment comparing means of L alignment for NC and C accents across languages were not significant either.

### 2.2.2 Pitch range and pitch scaling results

Figure 5 shows the box plots of pitch range of the initial NC (dark grey boxes) and C (light grey boxes) accents in semitones for all speakers of Catalan (top left), Italian (top right) and Spanish (bottom). As can be observed, the pitch range of the C/NC accent seems to be a very stable cue for Italian speakers since all the speakers with the exception of speaker 1 mark C through narrower pitch excursions. By contrast, we find no consistent tendency for either Catalan or Spanish. Thus, Catalan speakers 3, 4 and 5 and Spanish speakers 3 and 5 realise the C pitch accents with wider pitch excursions but there are speakers such as speaker 2 for Catalan, and speakers 2 and 4 for Spanish that present the opposite tendency, that is, they realise the C pitch accent with a narrower pitch excursion. Furthermore, some Catalan and Spanish speakers seem not to use pitch range to mark the difference between NC and C accents (speaker 1 for both Catalan and Spanish). The repeated measures ANOVA tests with pitch range of the C/NC accent as the dependent variable revealed no significant effects of focal condition \( (F(1,4) = .194, p > .05) \) or language \( (F(2,8) = 3.131, p > .05) \), but a focal condition \( \times \) language interaction \( (F(2,8) = 5.170, p < .05) \) was found. This significant interaction could indicate the very consistent behaviour exhibited by Italian compared to the tendency shown by Catalan and Spanish.

Additionally, we looked at the tonal scaling of L and H tones of the C/NC accent. We observed that the behaviour of the H tones is very similar to the results presented with respect to pitch range of the C/NC accent. Thus, lower C peaks are very systematic in Italian. However, in Catalan and Spanish no specific tendency is found. Some speakers tend to realise higher C peaks but other speakers produce lower C peaks. In addition, other Catalan and Spanish speakers do not use H scaling at all to mark the difference between NC and C accents. Regarding the L scaling, generally speaking no differences depending on focal condition were found. However, some speakers (speaker 3 for Catalan, and speakers 3 and 4 for Italian) tend to realise C accents by means of lower f0 values but at first sight it seems to be a very isolated tendency since the other speakers tend not to realise the focal distinction on the basis of L targets. All in all, these results show that the difference between C and NC initial accents is marked through a difference in tonal scaling of the H tone and not a difference in pitch range, since H tones of C accents are raised in Catalan and Spanish and lowered in Italian, while L tones are not affected by focal condition.
2.2.3 Duration results

Figure 6 shows the box plots of duration of the initial accented syllable in milliseconds for NC (dark grey boxes) and C (light grey boxes) accents for all speakers of Catalan (top left), Italian (top right) and Spanish (bottom). The graphs reveal that there is a significant and systematic effect of duration only for Italian, where we see that syllables bearing C accents become longer. For Catalan and Spanish, there are also some speakers (speakers 1, 4 and 5 for Catalan, and speaker 2 for Spanish) with a slight tendency to exhibit the same pattern found in Italian. The repeated measures ANOVA tests performed on the data with duration of the accented syllable as the dependent variable revealed significant effects of focal condition \( (F(1,4) = 66.359, p = .001) \) and language \( (F(2,8) = 49.207, p < .0001) \). The interaction focal condition \( \times \) language was also significant \( (F(2,8) = 8.774, p = .01) \). Post-hoc tests with the Bonferroni Adjustment comparing means of duration for NC and C accents across languages showed that Italian is significantly different from both Catalan \( (p = .01) \) and Spanish \( (p = .001) \).
2.3 Discussion

This production experiment shows interesting results related to the prosodic realisation of NC and C accents in initial position in Catalan, Italian and Spanish. In order to compare the two different accents, measures of alignment of the H and L targets, pitch range and f0 values of the peaks and valleys and duration of the syllables bearing the C/NC accent were taken. The results reveal an important effect of focal condition on the alignment of the H target for the three languages under study. These results thus confirm previous observations that both Catalan and Spanish speakers align the C peak near the end of the accented syllable while the peak for NC accents is aligned much later (de la Mota 1995 and Beckman et al. 2002 for Spanish; Estebas Vilaplana 2009 for Central Catalan). In Italian the difference between C and NC accents is also conveyed by differences in alignment with the syllable. C accents are always retracted in comparison with NC accents and are aligned roughly halfway through the syllable. By contrast, we find later peaks for NC accents, which tend to be aligned near the end of the syllable bearing the focal accent. Interestingly, the alignment realisation of the NC accents in Italian coincides with the realisation of the C accents in Catalan and Spanish, which can cause Catalan/Spanish speakers to perceive Italian speakers as focalising excessively.
Differences were found with respect to pitch range of the focus accent depending on the focal conditions across languages, this difference being caused by a different pitch height solely of the H tone. Importantly, for Catalan and Spanish, tonal scaling of the H target seems not to play a key role in marking the different focal conditions (as suggested by Face 2001, 2002 for Spanish) but, by contrast, it is a very stable cue in Italian. Our results support previous findings (Gili Fivela 2005, 2006, 2008) in characterising C accents as having lower peaks (see the Introduction for an overview about scaling results found for Pisa Italian). However, one could hypothesise that since the Italian C accent is the only one of those analysed here which is not just rising (since it is rising-falling), the difference in pitch range between both focal conditions could have to do with the pitch range of the pitch fall following the peak rather than that of the pitch rise preceding the peak. Nonetheless, when we inspected the behaviour of the L2, we observed that it also went lower in C accents. As a consequence, we discard any wider range for either the pitch rise or the pitch fall.

The results for Italian regarding tonal scaling are particularly interesting because they imply that Lecce Italian does not conform to the general prediction of the Effort Code. In an attempt to explain this behavior, our proposal is that the complexity of C accents in Italian, that is, the presence of a rising-falling movement, can act as a substitute for a salience marker in this language. A complex movement (a rising-falling movement) is more effortful than a simple movement (just falling or rising). Thus, the speaker is producing a greater effort to mark this constituent as ‘strong’ from an informational point of view. Interestingly, other languages such as Akan have been reported to not express prominence through an increase in pitch (Kügler & Genzel 2012).

Our results reveal substantial differences in duration only for Italian, in the sense that syllables bearing C accents are longer. Also in the case of Catalan and Spanish, there were some speakers that displayed a similar behavior to what was seen in Italian. This allows us to conclude that for Catalan and Spanish duration is an optional cue while for Italian it seems to be used systematically.

Finally, it is important to acknowledge that although we limited our study to the use of three prosodic correlates (namely peak timing, peak height or pitch excursion, and duration as markers of contrast), other prosodic parameters such as intensity as well as hyperarticulation in vowels/syllables can also be important in the marking of focus structures (see Jun 1993 on Korean; Krahmer & Swerts 2001 on Dutch; Heldner 2003 on Swedish; Avesani, Vayra & Zmarich 2007 on Italian; Baumann et al. 2007 on German; and Koreman et al. 2009 on German and Norwegian, among others).

3 Experiment 2: Perception experiment, identification task

As we saw in the previous experiment, speakers use different cues to different extents depending on the language to distinguish NC from C initial accents. Our next goal was therefore to test whether there is a correspondence between the cues used in production for each specific language and those used in perception, that is, whether the specific use of the

---

9 Though we found an effect of focus type on H tonal scaling whose magnitude depended on the language, the L point was less affected by it. However, there were three speakers (one Catalan speaker and two Italian speakers) that tended to realise lower L points for C. Interestingly, the Catalan speaker in question did not produce different tonal scaling of H at all. Thus, in some way, s/he was using a strategy based on scaling of L target to mark the difference. With respect to Italian speakers, they both marked the difference between the two accents through lower peaks for C accents. It seems that their strategy was to use a lower register for C.

10 Akan is a tone language that belongs to the Kwa branch of the Niger-Congo family spoken in Ghana (Kügler & Genzel 2012).
Table 3 Specific values of the steps of alignment, duration and scaling for each language.

<table>
<thead>
<tr>
<th>Language</th>
<th>Alignment</th>
<th>Duration</th>
<th>Scaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalan</td>
<td>22 ms</td>
<td>10 ms</td>
<td>8 Hz (from 334 to 286 Hz)</td>
</tr>
<tr>
<td>Italian</td>
<td>15 ms</td>
<td>12 ms</td>
<td>12 Hz (from 239 to 311 Hz)</td>
</tr>
<tr>
<td>Spanish</td>
<td>22 ms</td>
<td>10 ms</td>
<td>8 Hz (from 322 to 274 Hz)</td>
</tr>
</tbody>
</table>

Effort Code by each language in production is also reflected at the perceptual level. Our hypothesis was that irrespective of whether languages follow the Effort Code prediction or not in terms of production, perception will reflect the specific production patterns found for each language.

3.1 Methodology

3.1.1 Participants

Twenty native speakers of Majorcan Catalan (5 women, 15 men) aged between 16 and 45 years (mean = 28.65 years), twenty native speakers of Lecce Italian (17 women, 3 men) aged between 19 and 45 (mean = 27.75) and twenty native speakers of Madrid Spanish (14 women, men) aged between 23 and 53 (mean = 31.75) participated in this experiment. They reported having little or no prior training in phonology/phonetics and no history of hearing disability.

3.1.2 Materials

The prompt stimuli used in this perception experiment were created from a set of recordings of the sentence Na Marina vendrà (Cat.)/La Melania verrà (It.)/Marina vendrá (Sp.) ‘Marina/Melania [proper names] will come’ with C meaning/prosody, as spoken by one female native speaker of each language who had participated in the production experiment (but did not participate otherwise in the perception experiment). Catalan and Spanish materials were recorded in a quiet room at the Universitat Pompeu Fabra in Barcelona using a Marantz Professional PMD660 digital recorder and Rode NTG-2 microphone. Italian materials were recorded in a sound-proofed room at the Centro di Ricerca Interdisciplinare sul Linguaggio in Lecce using a Shure SM-86 condenser microphone and Goldwave 5.25 software installed on a desktop computer equipped with an Edirol UA-5 external sound card. All sentences were digitised at a 44100 Hz sample rate and 16 bit amplitude resolution. The criteria for selecting the materials were based on quality and clarity; we also guaranteed that the original alignment, duration and scaling values roughly corresponded to the mean values obtained for that speaker in the production experiment.

Four seven-step continua per language were created in which alignment, duration and scaling were manipulated from a C to a NC interpretation. The manipulation was done as follows: in the first continuum only alignment was manipulated while duration and scaling were neutralised; for the second continuum we manipulated alignment and duration but scaling was maintained; in the third continuum only alignment and scaling were manipulated while duration was kept constant and, finally, in the fourth continuum all three prosodic features were manipulated (see Table 3 for the step values used for each parameter and language). Note that only alignment was manipulated in isolation, in contrast with duration or scaling. This decision was made after observing that alignment was the only cue having a constant effect for the five speakers in the three languages. It also reflected a concern to keep the duration of the experiment short enough to keep the participants from experiencing monotony and boredom.

The manipulations were done using Praat (Boersma & Weenink 2009). Keeping in mind that the manipulation of duration might interact with the manipulation of alignment (especially
in the case of Italian, where the alignment differences take place on the same syllable),
the following precautions were taken: (a) the order of manipulation was always duration,
alignment and then scaling; (b) once a new step for duration was created, we carefully
checked that the peak occupied the same position (in terms of ratio) that it had occupied
before we manipulated duration; and (c) given that the manipulations in duration could cause
a little irregularity in the f0 rising movement, this irregularity was repaired by means of a
linear stylisation of the rising-falling movement. Figure 7 illustrates schematically the creation
of the continua.

3.1.3 Procedure
Subjects were seated in front of a laptop in a quiet room and the stimuli were played back
through headphones. Listeners received written instructions about how they were supposed
to respond after each stimulus. In addition, the prompt stimuli appeared in a dialogue so
that listeners would clearly understand the difference in meaning between the two stimuli.
Thus, if Catalan listeners perceived the NC Na Marina vendrà ‘Marina [proper name] will
come’, they were to press the ‘N’ key on the keyboard (for Normal, indicating a ‘normal’
declarative), whereas if they perceived the C Na MARINA vendrà, they were to press the ‘C’
key (for Correcció ‘Correction’, since the speaker is correcting the information that s/he has
just received from the interlocutor).

Stimuli were presented using E-prime version 1.2 (Psychology Software Tools Inc., 2009)
software running either on a MSI U100 Wind Notebook laptop equipped with a Realtek HD
sound card and Sennheiser HD 202 headphones (Catalan and Spanish) or on an HP desktop
computer equipped with a Creative Sound Blaster sound card and Sennheiser HMD 280
headphones (Italian). The final identification experiment consisted of 140 tokens per language
(7 steps \times 4 continua \times 5 blocks). The experiment lasted approximately 15 minutes and
there was a break of 10 seconds between each of the five blocks.

3.1.4 Statistical analyses
Since our data were categorical and not normally distributed, a Generalised Linear Mixed
Model (GLMM) analysis was performed. Mixed-effects models offer important advantages
over more traditional analyses such as repeated measures ANOVA. Some of these advantages
are that they do not need to assume sphericity, they can deal with missing data and they work
with both fixed and random factors. Identification responses were used as a dependent variable
for statistical exploration with stimulus (seven levels), parameter (four levels) and language
(three levels) as fixed factors, and listener and block as random factors. The results from the
GLMM tests include the F-ratio, the degrees of freedom from which it was calculated and
the significance value.

3.2 Results

3.2.1 Identification results
The four graphs in Figure 8 show the identification rate for the alignment continuum (top
left), alignment + duration continuum (top right), alignment + scaling continuum (bottom
left) and alignment + duration + scaling (bottom right). The ‘identification rate’ is defined as
the number of ‘normal’ responses over the total number of responses. We observe that for all
three languages speakers are able to change from a C to a NC interpretation merely on the
basis of alignment (Figure 8, top left); however, for Catalan and Spanish the identification
rate as ‘normal’ for stimulus 1 is higher than expected. This would indicate that the parameter
‘alignment’ is more helpful for Italian speakers than for Catalan or Spanish speakers. The
identification results are better when ‘duration’ is added to ‘alignment’ (Figure 8, top right). Here
the three languages have a very similar behaviour. However, when ‘scaling’ is added to
Figure 7  Schematic representation of the creation of the alignment continuum (top row, left), alignment + duration continuum (top row, right), alignment + scaling continuum (second row, left) and alignment + duration + scaling continuum (second row, right) for Catalan (and also for Spanish). Schematic representation of the creation of the alignment continuum (third row, left), alignment + duration continuum (third row, right), alignment + scaling continuum (bottom row, left) and alignment + duration + scaling continuum (bottom row, right) for Italian. White boxes represent consonants and shaded boxes vowels. Boxes representing the syllable bearing the nuclear accent are marked with thicker borders.
‘alignment’ a different picture emerges. It seems that while Catalan and Spanish speakers do well with this combination, Italian speakers are more sensitive to duration than to scaling, judging by the relatively high and low identification rates of ‘normal’ responses obtained for the Italian stimuli 1 and 7 respectively (Figure 8, bottom left). For Catalan and Spanish, the results improve substantially when all three prosodic features are at work (Figure 8, bottom right) but for Italian they are a little poorer than when only alignment and duration are exerting an effect. The GLMM analysis reveal a significant main effect of stimulus ($F(6, 18.276) = 330.750, p < .0001$), parameter $F(3, 8.276) = 33.771, p < .0001$) and language $F(2, 8.276) = 3.087, p < .05$). Further pairwise comparisons between the pairs of parameters for each language indicate that for Catalan the combination of parameters in which scaling is present differ statistically from the others ($p < .001$). Interestingly, in Italian no differences between any of the pairs are found, meaning that alignment plays a major role by itself already (otherwise the adding of a parameter would have caused a greater increase in listeners’ performance). For Spanish the results are similar to those for Catalan, with the difference that it is the presence of scaling or/and duration that makes the difference ($p < .0001$).

In summary, the results from the identification task confirm that the patterns found in production are also found in perception, meaning that irrespective of whether in production Catalan, Italian and Spanish do or do not follow the prediction made by the Effort Code with regard to the contrastive focus, the same patterns are found at the perceptual level. We therefore observe that while Catalan and Spanish speakers need the peak to go higher in pitch so that the C meaning is perceived, for Italian speakers the C peak has to be lower compared to the NC peak. We also detect some preferences for a specific parameter or combination of parameters depending on the language. Thus, the results from pairwise comparisons suggest that while Catalan and Spanish listeners need the coalescence of alignment with another prosodic cue, Italian listeners do not show any preference for any specific parameter or combination of parameters. However, in the next section we compare the four curves mathematically so that the specific contribution of each parameter for each language can be evaluated.

### 3.2.2 Curve Estimation results

In order to compare mathematically the four identification curves obtained according to the parameters that were manipulated (alignment, alignment + duration, alignment + scaling, and alignment + duration + scaling) for each language, and, consequently, to be able to identify which parameter or combination of parameters is the most efficient, the set of data points making up each of the four curves was fitted to a logistic function through the Curve Estimation procedure (Keating 2004) in SPSS (SPSS for Windows). From the SPSS fitted logistic curves, we obtained the ‘b1’ value, which is related to the slope of the curve. Table 4 shows the slope values for the identification curve obtained for each continuum and language. The slope value is not returned directly by the function, but since the term ‘b1’ is related to the slope, we will use this value as a measure of slope. Lower values reflect steeper curves (shaded boxes in Table 4), i.e. conditions under which listeners are more successful at distinguishing between the two different meanings. The results confirm that Catalan and Spanish listeners need the three prosodic parameters to be at work in order to clearly distinguish the two different focal conditions, while Italian listeners have an important preference for the combination of alignment and duration. Alternatively, Catalan listeners prefer the combination of alignment and scaling, Italian listeners consider alignment alone sufficient and, finally, Spanish speakers opt for the combination of alignment and duration. In sum, these results are consistent with the results obtained from the GLMM analysis.

---

11 Recall that in the specific case of Italian, stimulus number 1 of the scaling continuum corresponds to the lowest peak while in Catalan and Spanish it was stimulus number 7 that showed the lowest pitch height.
Figure 8  Identification rate for the alignment continuum (top left), alignment + duration continuum (top right), alignment + scaling continuum (bottom left) and alignment + duration + scaling continuum (bottom right) for Catalan (black bars), Spanish (dotted line) and Italian (light grey bars). Error bars represent the standard error of the mean.

Table 4  Slope values for each continuum and language calculated by means of a Curve Estimation procedure. Shaded cells indicate the lowest values (dark shading: the lowest value, light shading: the second lowest value) for each language.

<table>
<thead>
<tr>
<th></th>
<th>Alignment</th>
<th>Alignment+duration</th>
<th>Alignment+scaling</th>
<th>Alignment+duration+scaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalan</td>
<td>.445</td>
<td>.320</td>
<td>.237</td>
<td>.208</td>
</tr>
<tr>
<td>Italian</td>
<td>.400</td>
<td>.364</td>
<td>.557</td>
<td>.460</td>
</tr>
<tr>
<td>Spanish</td>
<td>.570</td>
<td>.394</td>
<td>.439</td>
<td>.373</td>
</tr>
</tbody>
</table>
4 Discussion and conclusions

In this study, we explored the specific contribution of tonal alignment, duration and tonal scaling to the production and perception of contrastive focus in Catalan, Italian and Spanish. In the production experiment, measures of alignment of the H and L targets, pitch range, f0 peaks and valleys of the focal accent, and duration of the syllables bearing them were taken. An important effect of focal condition on the alignment of the H target for the three languages was found. Consistent with previous studies, C peaks occurred earlier than NC peaks in all three languages (see Estebas Vilaplana 2009 for Central Catalan; D’Imperio 2002 and Gili Fivela 2002, 2005 for Neapolitan and Pisa Italian, respectively; de la Mota 1995 and Beckman et al. 2002 for Spanish). By contrast, little or no focal effect on the alignment of the L target was found for any of the three languages (Estebas Vilaplana 2009 for Central Catalan; Gili Fivela 2005, 2006 for Pisa Italian; de la Mota 1995 for Spanish). With respect to scaling features, Italian speakers are the only ones to use the tonal scaling of the H target in a systematic fashion. Our results agree with previous studies of Pisa Italian (Gili Fivela 2005, 2006) which report that C peaks are lower than NC peaks. For Catalan and Spanish, though results were not so clear cut, there was a tendency for some speakers to mark the C accents through higher peaks. Finally, syllables bearing C accents were systematically longer only for Italian (similar to Gili Fivela 2005, 2006) but a tendency in the same direction was found for Catalan and Spanish (consistent with Estebas Vilaplana 2009 for Central Catalan; and de la Mota 1995 for Spanish).

The Italian scaling results are interesting in terms of the Biological Codes (Gussenhoven 2002, 2004), specifically the Effort Code, since they contradict the general predictions described by this code. Interestingly, another study has shown that the Akan language does not follow the general prediction of the Effort Code either (Kügler & Genzel 2012). This study undertook a production experiment to investigate the tonal and durational means of encoding focus and givenness in Akan. To this end the researchers recorded eleven speakers uttering sentences with target words occurring in sentence-medial and sentence-initial position in different information structure constructions. The sentences were elicited through context questions that placed target words either in wide, narrow or contrastive focus in pre-focal or post-focal position in order to study givenness. Prosodic parameters such as duration and f0 were measured on the target words. Contrary to the predictions of the Effort Code (which predicts a correlation between the importance associated with the element and the effort, resulting in higher pitch targets), they found significantly lower realisations of both high and low tones in contrastive focus. Hence, they propose an alternative view on the prosodic expression of focus in the sense that what seems to be important is not the direction of change to mark the focus but rather the deviation from the neutral register. These language-specific strategies to mark contrastive focus are hinted at by Gussenhoven (2002) when he states that ‘an important aspect of the present conception of intonational meaning is that while the nature of the meanings is related to the way our speech organs produce pitch variation, there is no implication that the physical condition that lies at the basis of these meanings need to be present in order to create the forms’. Turning to our own results for Italian, our interpretation is that the substitute variable in pitch height in the specific case of Lecce Italian is the complexity of the tonal movement, which is to say that it is the rising-falling movement aligned with the accented syllable that is being used as a C marker. We speculate that the Italian impossibility of going higher in pitch in the C accent is due in origin to mechanical reasons, that is, since the f0 has to go up and down within the minimum temporal space of one syllable, speakers have found it easiest to simply avoid pitch range expansion. Once speakers became aware of this connection between mechanical procedure and meaning they applied it to conveying the contrastive focus pragmatic meaning. In other words, what started out being an articulatory consequence of the complexity of the pitch movement has now grammaticalised into a prosodic device to mark contrastive focal meaning. We could say,
then, that in Italian the scaling strategy remains subordinate to the alignment strategy due to articulatory or mechanical factors.

One could also argue that the other biological codes proposed by Gussenhoven are relevant for contrastive focus marking in Italian. The complex tonal movement in Lecce Italian is analysed as an $H^*+L$ pitch accent (Stella & Gili Fivela 2009). As shown in the graphs here (see Figure 1 above, right panels) and systematically discussed for other varieties, e.g., Pisa Italian, the pitch accent involves both a rise to the peak starting from a phonetic low target and a fall to a low target that is found in stable position with respect to the peak. Thus, according to a phonetically transparent transcription the accent could be represented as $L+H^*+L$ as opposed to the rising accent $L+H^*$. On the basis of the Frequency and Production Codes respectively (Gussenhoven 2002, 2004), the trailing low component in the pitch accent may be interpreted as conveying self-confidence and finality, in addition to forcing the related realisation of a rising-falling pattern with the possible mechanical consequences described above. In short, the Frequency and Production Codes suggest that the presence of the low trailing tone may play a role in distinguishing the pitch accent for contrastive focus (peremptorily and conclusively opposing an element in the background) from the $L+H^*$ accent used in non-contrastive utterances for introducing a topic and expressing continuation (Gili Fivela 2008: 191). In this respect the differentiation of the two pitch accents would be due to the interpretation of the pitch accent as a whole together with the duration characteristics and with reference to other biological codes rather than just the Effort Code. This could explain why scaling information related to the peak is not lent much importance: it is the pattern as a whole, involving the falling movement (i.e. the low trailing tone target) that is primarily taken into account.

Another issue is why Catalan, Italian and Spanish speakers realise contrastive peaks earlier (and not necessarily higher for Catalan and Spanish) than non-contrastive ones. A possible explanation would be that a retracted peak would allow more available room to lower the pitch after the accented syllable, which would make the accent more prominent. A lower $L_2$ would be related to the Frequency and Production Codes expressing peremptoriness and conclusiveness, which are very much related to the idea of correcting or denying a previous assertion. This hypothesis has been tested only for Italian, in which we do find a consistently lower $L_2$ in contrastive sentences. However, further phonetic analyses for Catalan and Spanish are needed to be able to fully verify this hypothesis. The presence of a lower $L_2$ in contrastive sentences is closely related to the post-focal compression (Xu 1999, Xu, Chen & Wang 2012) that seems to be common in the languages studied here. However, it is important to highlight that although post-focal reduction is a very important mechanism to mark contrast in these languages (as stated by various scholars), it is not an indispensable one. Thus, in a previous study conducted by the same authors of the present work on the same languages and using similar materials (see Vanrell et al. 2011) and the gating paradigm, it was shown that listeners have no need to hear the post-focal region to distinguish between the two focal conditions. Interestingly, even when only alignment was intact while duration and scaling were neutralised, listeners were still able to recognise fairly well the accent type by the second or third gates (in which listeners were able to hear just the portion going from the beginning of the sentence to half or three quarters of the way through the first accented syllable).

The different sets of results yielded by the identification task (i.e. the identification responses and the Curve Estimation analysis) present converging evidence for the following two assertions: (a) there is a direct correspondence between production and perception regardless of whether in terms of production the languages under study agree with the assumption derived from the Effort Code (it is particularly interesting to notice that Italian speakers perform in the expected way in perception since they rate lower peaks as being C in meaning), and (b) the three languages have their own specific preferences for particular prosodic parameters. Thus, though only alignment was a stable cue in production for Catalan and Spanish speakers while all three parameters were robust for Italian speakers, the three languages used a combination of the three prosodic features whenever they were available in perception, showing specific preferences. The fact that speakers can have a rather unsystematic
use of the different cues in production and then become more efficient in perception may be striking at first. We believe that our results provide evidence in favour of a synergy between production and perception in intonation. Production can be more subject to variation but these deformations are repaired by the auditory system with the aim of obtaining an optimal communicative transaction.

As for the language-specific preferences for particular parameters, we first observe that when possible, Catalan listeners select the three cues together, but if they are not all available, they have a special predilection for the combination of alignment and scaling. Second, Italian speakers do have a preference for the combination of alignment and duration or, in lieu of that, for alignment alone. Finally, Spanish speakers, like Catalan speakers, give priority to the stimuli combining the three prosodic cues: yet when they are not available, alignment and duration can also be sufficient, as they are for Italians. One might ask why while Catalan and Spanish speakers primarily select the three cues together, Italian speakers are satisfied with only two parameters. Our interpretation is that Italian speakers select the salience cue par excellence, that is, the rising-falling movement (expressed by means of alignment) and then a possible extra consequence of this complex movement, an increase in the duration of the syllable bearing the focus accent. It is understandable, then, that these cues are preferred over scaling, though this preference contradicts universal assumptions about language. Another possibility is that there was a methodological problem in the manipulation of our materials. Since we wanted our experimental materials to be comparable between the three languages, we only manipulated the scaling of the peak of the contrastive accent despite being aware that the scaling of the L2 had an important role in the conveyance of the contrastive/non-contrastive meaning. This methodological decision might have blurred the effect of scaling on the perception of the contrastive/non-contrastive meaning in Italian.

In sum, the results reported in this paper represent further evidence that typologically related languages such as Catalan, Italian and Spanish can use different strategies as prosodic contrastive markers. These results contribute to a better comprehension of the principles involved in both production and perception of contrastive focus across Romance languages. Finally, we would also like to highlight the importance of testing cross-linguistically the perceptual salience of the cues found in production in order to fully understand the interplay that exists between production and perception mechanisms.

Acknowledgements

A preliminary version of this work was presented at the workshop Sobre entonació del català i Cat_ToBI (July 2011, Barcelona, Spain), the conference Tone and Intonation in Europe 4 (September 2011, Stockholm, Sweden) and the 7th Conference of the Associazione Italiana di Scienze della Voce (January 2011, Lecce, Italy). We are grateful to the audience at these talks, especially to Cinzia Avesani, Francesco Cangemi, Mariapaola D’Imperio, Sonia Frota, Martine Grice, Carlos Gussenhoven and Lisa Selkirk, for their helpful comments and discussion. An early draft of this paper was part of the Ph.D. dissertation of the first author, so we would also like to thank the Ph.D. Committee members (Mariapaola D’Imperio, Gorka Elordieta and Gemma Rigau) as well as two external reviewers (Aoju Chen and Frank Kügler) for the comments and criticisms which have been profitably taken into account in the preparation of this manuscript. All the speakers and listeners who participated unselfishly in these experiments as well as friends from Lecce, Mallorca and Madrid who put us in contact with potential participants deserve a special mention for their time and dedication. We would also like to thank the Servei d’Estadística Aplicada (Universitat Autònoma de Barcelona) and Joan Borràs-Comes for their help in the statistics used in this research. Finally, we would also like to thank the Editor, Adrian Simpson, as well as two anonymous reviewers for their many suggestions for improving this manuscript. This research has been funded by projects FFI2009–07648/FILO, FFI2011–23829/FILO, CONSOLIDER-INGENIO 2010 CSD2007–00012,
BFU2012–31995 (awarded by the Spanish Ministry of Science and Innovation and the Spanish Ministry of Economy and Competitiveness) and 2009 SGR 701 (awarded by the Generalitat de Catalunya), and by a short research stay subsidised under the Subprogram FPI-MICINN. The paper describes a piece of research which is relevant for the general project on Romance languages InAPoP (PTDC/CLE-LIN/119787/2010), funded by the FCT (Portugal), to which the authors belong to.

References


Estebas Vilaplana, Eva. 2009. The use and realization of accentual focus in Central Catalan with a comparison to English. Munich: Lincom Europa.


