The role of pitch range in establishing intonational contrasts

Joan Borràs-Comes, Maria del Mar Vanrell and Pilar Prieto

Journal of the International Phonetic Association / Volume 44 / Issue 01 / April 2014, pp 1 - 20
DOI: 10.1017/S0025100313000303, Published online: 21 March 2014

The role of pitch range in establishing intonational contrasts

Joan Borràs-Comes

Universitat Pompeu Fabra

joan.borras@upf.edu

Maria del Mar Vanrell

Freie Universität Berlin

mariadelmar.vanrell@fu-berlin.de

Pilar Prieto

Institució Catalana de Recerca i Estudis Avançats (ICREA) & Universitat Pompeu Fabra

pilar.prieto@upf.edu

One of the unresolved issues in the field of intonational phonology is whether pitch accent range differences are used by languages to express discrete linguistic distinctions. In Catalan, as in some other Romance languages, a rising-falling nuclear pitch contour – i.e. a rising pitch accent associated with the utterance-final stressed syllable followed by a Low boundary tone – can be used to convey three different pragmatic meanings depending on its pitch range properties: information focus statements (IFSs), corrective focus statements (CFSs), and counter-expectational questions (CEQs). In order to investigate how these pragmatic meanings are distributed across the pitch range continuum and whether Catalan listeners use these tonal scaling distinctions to identify such meanings, we performed an identification task and a congruity test. The results show that CEQs differ from both IFSs and CFSs in a discrete way, yet the perceived difference between IFSs and CFSs cannot be exclusively explained by scaling differences. These findings provide further evidence that pitch range differences can be used to make intonational distinctions in some languages, and strengthen the argument that pitch range features need to be represented descriptively at the phonological level.

1 Introduction

As is well known, intonational languages use pitch variation to express differences in pragmatic and discourse meanings. Though early approaches distinguished among four or three level tones (Trager & Smith 1951, Stockwell, Bowen & Silva-Fuenzalida 1956, respectively), the Autosegmental-Metrical (AM) model takes as a central assumption that only two tones, Low and High, are necessary to distinguish pitch accent and boundary tone categories in English. This means that all remaining pitch range variation exclusively expresses differences in emphasis or prominence (e.g. Pierrehumbert 1980, Beckman &

Journal of the International Phonetic Association (2014) 44/1
doi:10.1017/S0025100313000303

Pierrehumbert 1986, Bolinger 1986, Dilley 2010). This assumption relies on a version of the so-called Free Gradient Hypothesis (Ladd 1994, 2008; Gussenhoven 1999), which holds that one of the most common effects of gradually expanding the pitch range of a given pitch accent is the pragmatic reinforcement of the utterance (namely an increase in the degree of the speaker's involvement in the speech act). In line with this, Liberman & Pierrehumbert (1984) demonstrated in their study of English tonal range that a gradual increase in emphasis was correlated with an increase in tonal range of the pitch accent.[1]

Nevertheless, work on English and other languages has revealed that pitch range variation can express categorical differences in meaning (Ward & Hirschberg 1985; Hirschberg & Ward 1992; Ladd 1994, 2008; Ladd & Morton 1997; Chen 2003; Braun 2006; Savino & Grice 2011; Vanrell 2006, 2011). Whereas it is generally accepted that tones in tone languages behave as phonemic units, a traditional belief exists that pitch variations in non-tone languages do not behave that way. In the last decades, work within the intonational phonology field has shown that intonational contrasts apply to intonational languages, conveying 'meanings that apply to phrases or utterances as a whole, such as sentence type or speech act, or focus and information structure' (Ladd 2008: 7). For example, Ladd & Morton (1997) investigated the contrast between normal and emphatic rising pitch accents in English. Though an abrupt shift in identification from normal to emphatic interpretations was found as pitch range increased, little evidence was provided of an associated peak in discriminability between stimulus pairs. In a replication study of the experiment, Chen (2003) claimed that taking the identification results together with an analysis of reaction time (RT) data revealed that the perceived distinction between a normal High accent and an emphatic High accent is of a discrete nature. Hirschberg & Ward (1992) showed that a larger pitch range of the English rise-fall-rise tune can change the interpretation of an utterance from one of uncertainty to one of incredulity. Finally, Calhoun (2004) found that themes and rhemes are marked by distinctive pitch accents and that the most reliable cue to the theme and rheme accents is pitch height.

Some recent work on Romance languages has found that pitch range variation can also convey discrete intonational contrasts. Savino & Grice (2011) demonstrated that the pitch range of a rising pitch accent was responsible for the difference between information-seeking and counter-expectational questions in Bari Italian (where the latter are produced with an expanded pitch range). The listeners' responses and reaction times obtained by means of a semantically motivated identification task provided clear evidence for the categorical use of pitch range variation in Bari Italian question interpretation. Similarly, by using the results of a gating experiment, Face (2005, 2007, 2011) claimed for Spanish that the height of the initial f0 peak of an utterance allows listeners to distinguish between declaratives and *yes–no* questions, thus arguing for the phonologization of pitch range. This was consistent with Prieto (2004), who found that the height of the initial f0 peak varies depending on sentence type; specifically, *yes–no* questions, *wh*-questions, exclamatives, and imperatives all have significantly higher initial f0 peaks than declaratives. Moreover, Vanrell (2011) showed for falling nuclear pitch accents (H+L* L%) that the pitch height of the High leading tone is the main cue used by Majorcan Catalan listeners to distinguish between a *wh*-question and two types of *yes–no* questions. That is, an upstepped leading High tone signals a *yes–no* question in which the speaker has no previous knowledge about the answer, whereas a non-upstepped leading tone

[1] Terms such as 'height', 'scaling', and 'range' will be used throughout the paper. All these terms are related to the notion of pitch. While the pitch 'height' is determined by vocal fold vibration, the 'scaling' refers to changes in pitch height. In turn, 'pitch range' refers to the distance or span between the lowest and the highest f0 values observed in an utterance (a valley and a peak, respectively). In addition, varying the 'register' of an utterance implies that a whole contour is raised or lowered in the f0 space. For more information on these concepts, see Gussenhoven (2004). In the present article, the values concerning register are kept constant, and so are the values of the lowest f0 values.

signals that the speaker is asking a *yes–no* question about mutually shared information; in addition, a downstepped leading tone signals a *wh*-question.[2]

In general, these investigations demonstrate that pitch range variation can be perceived in a discrete fashion in some languages and thus strengthen the arguments in favor of treating pitch range differences in phonological terms in these languages. The idea of enriching the traditional High–Low dichotomy with a finer differentiation of pitch range was already advocated by researchers such as Ladd (1994: 60), who pointed out that 'the Bruce-Pierrehumbert approach to intonational phonology must be enriched with a notion of categorical distinctions of pitch range. We need to get rid of the idea that any distinction that is orthogonal to the basic opposition between High and Low tones is ipso facto gradient: both gradient factors and categorical ones play a role in the vertical scale of any given tone'.

In this paper, we investigate more extensively the role of pitch accent range variation in conveying intonational contrasts in Catalan. In our previous descriptive studies based on the analysis of Catalan dialectal data from the *Interactive Atlas of Catalan Intonation* (Prieto & Cabré 2007–2013; see also Prieto 2002) using Cat_ToBI (Aguilar, De-la-Mota & Prieto 2009; Prieto et al. 2009; Prieto, 2014;) we observed that the rising pitch accent of information focus statements (IFSs) was produced with a narrow pitch range, while that of corrective/contrastive focus statements (CFSs) and counter-expectational questions (CEQs) was produced with a wider pitch range.[3] In these three types of utterances, the alignment properties of the tones are found to be the same, i.e. a Low tone is aligned with the beginning of the accented syllable, the rising movement occurs within this accented syllable, and the peak of this rise is aligned approximately with the end of the accented syllable.[4] Similar observations have been made for other Romance languages, such as Friulian (Roseano, Vanrell & Prieto 2014), Sardinian (Vanrell et al. 2014) and Castilian Spanish (Estebas-Vilaplana & Prieto 2010). Examples of linguistic contexts eliciting these three types of pragmatic meanings are shown in (1).[5]

---

[2] This specific case of upstep is an example of a paradigmatic use of upstep, which has the function of creating a categorical distinction between pitch accents. We also note that the classic paradigmatic use of upstep (and downstep), as a phrasal mechanism, is also used in this language. Similar conclusions have been drawn when examining boundary tones. Some crosslinguistic studies have reported active Mid-level boundary tones contrasting with High-level tones in the phonological domain of English (Beckman & Ayers Elam 1997), Greek (Arvaniti & Baltazani 2004), German (Grice et al. 2005), Spanish (Beckman et al. 2002), Korean (Lee 2004), and Catalan (Vanrell 2011).

[3] By information focus statement, we refer here to a carrier of new information in which there is a particular constituent which is focalized with respect to the background. On the other hand, corrective focus statement refers to the marking of 'a constituent that is a direct rejection of an alternative' (Gussenhoven 2007). This type of focus typically corrects 'the value of the alternative assigning a different value' (Cruschina 2011). The main difference between the two focus types is that while corrective focus is dependent on the previous assertion which is denied/corrected by the new focalized item, information focus is not. Though not entirely uncontroversial (Brunetti 2004), the distinction between information and corrective/contrastive focus has been widely accepted by the linguistic community.

[4] The AM representation adopted for this rising accent is L+H* L%, as stated in Prieto et al. (2009) and Prieto (2014). Even though there can be little variation in the alignment of the later peak, this variation is phonetic-like and does not convey differences in meaning, as reported in these publications (see also Prieto 2005).

[5] Even though Romance languages such as Catalan, Italian, and Spanish have been said to mark CFS through syntactic mechanisms (Vallduví 1991, Ladd 2008), this does not exclude an active role for intonation, especially in those cases in which word order remains the same (Estebas-Vilaplana 2009 for Catalan; Face & D'Imperio 2005 for Italian and Spanish). According to previous research on this issue (Solà 1990, Vallduví 1991), since prominence shift is a less-used strategy in Catalan to make the focused constituent fall under prominence, other syntactic mechanisms such as dislocation (*NO LES TINC, les claus*, lit. 'NOT THEM I.HAVE, the keys', 'I do not have the keys') or elision (*NO LES TINC*, lit. 'NOT THEM I.HAVE', 'I do not have them') of the non-focal material of a sentence (Solà 1990, Vallduví 1991, Prieto & Cabré 2007–2013), focus fronting (*NEGRES, són, i no blanques*, lit. 'BLACKS, they.are, and not whites', 'They are black, not white') or clefting (*És EL MARÇAL* (*que/el que/qui/el qui*) *no suporto*, lit. 'IS THE MARÇAL

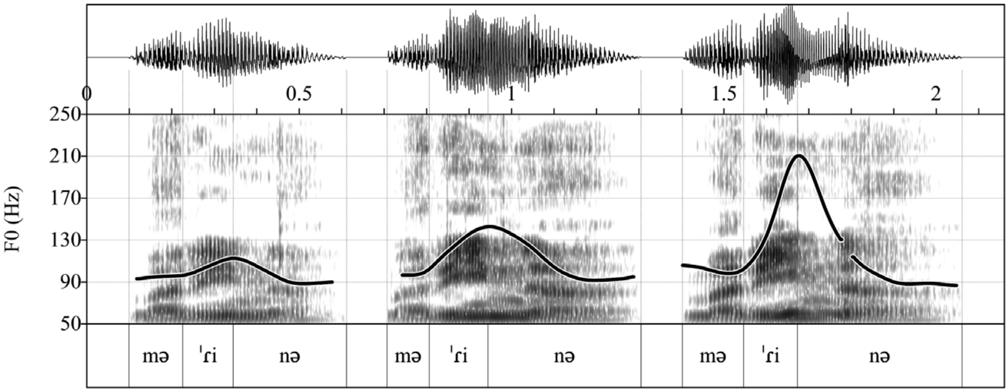

Figure 1 Waveforms, spectrograms and f0 contours of the proper name *Marina* produced with an IFS meaning (left), a CFS meaning (central position), and a CEQ meaning (right).

(1)	a.	IFS	Com es diu, la seva filla?	'What's their daughter's name?'
			Marina.	'Marina.'
	b.	CFS	Es diu Júlia, ella, no?	'Her name's Júlia, isn't it?'
			MARINA!	'[No! It's] MARINA!'
	c.	CEQ	Li posaran Marina.	'They'll call her Marina.'
			Marina?	'Marina? [Really?]'

Figure 1 shows the waveforms, spectrograms and f0 contours of the proper noun *Marina* ([məˈɾinə]) obtained as responses to the contexts in (1).

With the aim of investigating the role of pitch range in the interpretation of rising pitch accents in Catalan, we initially carried out two identification tasks with twenty native speakers of Catalan, the results of which are reported in Borràs-Comes, Vanrell & Prieto (2010). These tasks were identification tasks with binomial identification responses (two-way identification tasks), the first dealing with the contrast between IFSs and CEQs and the second with the contrast between IFSs and CFSs. The identification results showed an S-shaped function for both comparisons, thus suggesting a discrete perception for three types of pragmatic meanings. However, an analysis of the reaction times revealed a significant reaction time peak only when IFSs was compared with CEQs at the presumed category boundary. As Chen (2003: 98) pointed out, 'if the identification categories emerging from the response frequencies are not task-induced but linguistically real, we will expect that the within-category stimuli are comparable in terms of cognitive load and therefore will trigger similar mean RTs for identification', and vice versa. This close correlation has also been found in many other experiments (e.g. Falé & Hub Faria 2005 for European Portuguese, and Vanrell 2006, 2011 for Catalan). The fact that we found no peaks in RTs in the IFS vs. CFS comparison (Borràs-Comes et al. 2010) was interpreted as providing initial evidence for both a categorical effect in pitch range (i.e. the phonological difference between an IFS and a CEQ) and a gradient effect (i.e. the difference in scaling between an IFS and a CFS). In addition, recent results

who not I.stand', 'It is Marçal who I cannot stand') (Solà 1990, Vallduví 1991) are proposed. Such sentence types are characterized by a similar intonation pattern L+H* L%, either produced in isolation or accompanied by non-focal material, which tends to undergo tonal compression.

from a mismatch negativity analysis by Borràs-Comes et al. (2012) back up this result using very similar materials. This study found a stronger MMN brain response when inter-category stimuli (IFSs and CEQs) were presented than when listeners heard pairs of intra-category stimuli having the same physical distance in terms of pitch range between them (either two types of IFSs or two types of CEQs).

The goal of the present study is to investigate more deeply the role of pitch accent range in conveying the above-mentioned pragmatic meaning distinctions in Catalan (IFS, CFS, and CEQ) by using two tasks. First, we will use an identification task allowing for the simultaneous comparison of the three categories (Experiment 1), and then we will take linguistic context explicitly into account in order to test for the congruity of each target sentence occurring in a typical linguistic context for each pragmatic meaning (Experiment 2). These experiments are complemented with the results of reaction time measures, as these measures have been found to be significantly useful to investigate the discreteness of different intonational contours. Following our initial findings showing that the comparisons between IFS/CFS and IFS/CEQ do not behave alike, we initially hypothesized that the three categories would not be distributed in three well-differentiated areas of the pitch height continuum depending on the scaling of the H tone, but rather in only two such areas.[6]

Another goal of the study is to assess the utility of these tasks for the investigation of the role of intonational differences in conveying pragmatic meaning distinctions. A triple identification task and a congruity task were thus conducted to test for the presence and hierarchy of this potential three-way distinction between rising pitch accents in Catalan. This would give us more information about the suitability of binomial identification tasks for the investigation of pragmatic meanings. In other words, we want to know if such a three-way contrast in identification will lead to similar results as two separated two-way contrasts and whether the results of such a study can be corroborated by using a congruity task. Experiment 1 consisted of a semantically motivated identification test in which participants had to identify each of the three meanings (IFS, CFS, and CEQ) for a set of isolated stimuli, allowing for a triple response. To our knowledge, no similar triple identification tasks have been previously applied to intonation, and so this is the first study approaching the analysis of intonational contrasts that allows for more than two responses at a time. Experiment 2 consisted of a congruity test which tested participants' acceptance of each stimulus occurring within a typical communicative context. This type of task allows us to investigate whether listeners are aware of the semantic appropriateness of a particular intonation contour to a given communicative context and can detect an incongruous use of this contour. This methodology has been used successfully by other researchers investigating intonation contrasts (see Rathcke & Harrington 2006, Rathcke & Harrington 2010, Crespo-Sendra 2011, Vanrell 2011). A set of twenty native speakers participated in the two experiments.

---

[6] Note that a three-way distinction in pitch height does not represent a very marked situation cross-linguistically if we consider the tonal height distinctions reported for tone languages. For example, in some African languages there is a distinction between lexical tones that are High and Overhigh (see McHugh 1990 on Chaga). Likewise, Francis, Ciocca & Ng (2003) report a three-way distinction between lexical tones in Cantonese. In this tone language, the same syllable /ji/ means 'doctor' when produced with a High-level tone, 'two' when produced with a low-level tone, and 'spaghetti' when produced with a Mid-level tone. The results of two identification experiments showed that the perception of Cantonese level tones is qualitatively similar to that presented by Abramson (1979) for Thai level tones. The listeners showed evidence of the presence of category boundaries in an identification task, but no corresponding peaks in discrimination accuracy. Just as there are tone languages with two or three distinct level tones, it would not be surprising if some intonation languages can make use of more than two level tones to express a variety of pragmatic meanings.

2 Experiment 1: Identification task

2.1 Method

This experiment consisted of an identification task with three possible response options. Specifically, participants had to classify each of the auditory stimuli as conveying one of the three pragmatic meanings of interest in our study, namely IFS, CFS, and CEQ. As noted above, as far as we know, no similar triple identification task has thus far been used to investigate potential differences in intonational pitch perception. We initially hypothesized that the triple response procedure would be able to test whether Catalan listeners would be capable of distributing the acoustic pitch range continuum into three or two discrete categories.

2.1.1 Participants and materials

A set of twenty native speakers of Central Catalan participated in the experiment.[7] All subjects were undergraduates studying journalism or translation at the Campus de la Comunicació of the Universitat Pompeu Fabra in Barcelona and were paid for their participation. They were seven men and 13 women. All were right-handed and none of them had previous experience with linguistic perception tasks. The age of the participants was between 19 and 37 years (average = 21.6, standard deviation = 4.1). The average Catalan dominance of the participants (taken from a report on the daily interactions in Catalan provided by the participants themselves) was 86% (standard deviation = 13%).

We first recorded the three short dialogs shown in (2) below in order to produce an appropriate context for an IFS (2a), a CFS (2b), and a CEQ (2c). A male Catalan native speaker was recorded using a Marantz PMD-660 digital recorder in a quiet room at the Universitat Pompeu Fabra. The productions were elicited using the Discourse Completion Test method (Nurani 2009, Félix-Brasdefer 2010). Following this methodology, participants were asked to respond to a situational prompt that was offered by the interviewer. (Text in brackets corresponds to those fragments that can be omitted in Catalan spontaneous speech.)

(2)	a.	IFS	Com la vols, la cullera?	‘What type of spoon do you want?’
			Petita, [sisplau].	‘[I want a] little [spoon, please].’
	b.	CFS	Volies una cullera gran, no?	‘You want a big spoon, don’t you?’
			PETITA, [la vull, i no gran].	‘[I want a] little [one, not a big one].’
	c.	CEQ	Jo la vull petita, la cullera	‘I want a little spoon.’
			Petita? [N’estàs segur?]	‘[A] little [one]? [Are you sure?]’

We then created a synthesized continuum for the noun phrase *petita* [pə.ˈti.tə] ‘little.FEM’ by modifying the f0 peak height in 11 steps (distance between each one = 1.2 semitones). This target word (which contains voiceless plosives) was selected so that we would be able to compare our results with those of Borràs-Comes et al. (2012) and thus enhance the comparison between the experimental methodologies used in the three studies. A single item was used so that listeners could easily keep in mind the three linguistic contexts provided at the beginning of the task. The speech manipulation was performed on a single [pə.ˈti.tə] recording by means of the Pitch Synchronous Overlap and Add (PSOLA) resynthesis routine available in Praat (Boersma & Weenink 2008). This routine keeps the segmental information invariable, thus making it possible to test for only the changes in pitch height. Figure 2 shows an idealized

[7] In order to constrain the sample of population in our experiments, we only included speakers of Central Catalan, which is the Catalan dialect spoken in the province of Barcelona, the province of Girona, and the eastern part of the province of Tarragona. Specifically, 15 of our subjects came from the province of Barcelona, three from the province of Tarragona, and two from the province of Girona.

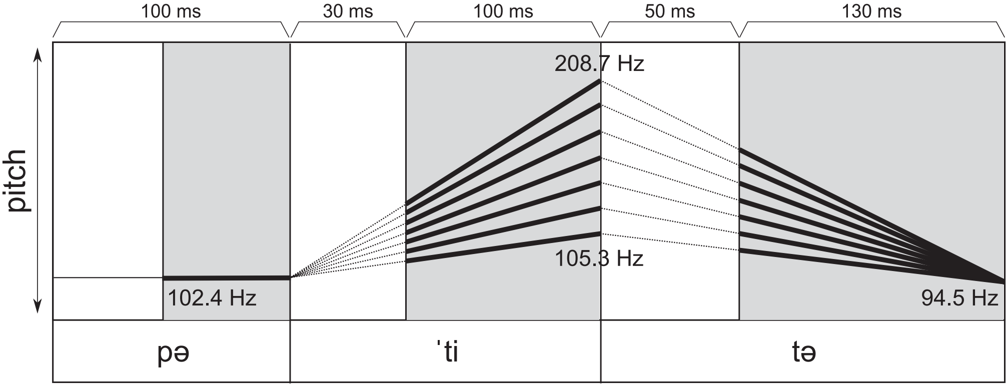

Figure 2 Idealized schema of the pitch manipulation in the noun phrase *petita* [pə.ˈti.tə] 'little.FEM'. Duration of the segments is shown at the top, and the correspondence with each segment is shown at the bottom. The Hz values at the center of the image represent the final frequencies of the extreme stimuli (steps 1 and 11). Thin lines correspond to the idealized contours that would be found within the closure periods for [t] in [ti] and [tə]. Please note that even though only seven idealized pitch contours are represented in the figure, eleven stimuli were used in the identification test.

Table 1 Values in Hz for the H tone of our stimuli and distance in semitones between the L and H targets of each one.

Stimulus	Hz for H tone	Semitone difference between L and H
1	105.3	0.48
2	112.8	1.68
3	120.8	2.85
4	129.2	4.02
5	138.3	5.20
6	148.1	6.38
7	158.5	7.56
8	169.6	8.74
9	181.8	9.94
10	194.9	11.14
11	208.7	12.33

schema of the pitch manipulation in the target noun phrase. As shown in the figure, pitch movements were realized with a rising tonal movement starting at the onset of the accented syllable [ˈti], which was preceded by a low plateau for the syllable [pə] (102.4 Hz, 100 ms). The posttonic syllable [tə] was realized with a falling tonal movement (94.5 Hz, 180 ms). The peak height continuum ranged from 105.3 Hz to 208.7 Hz, and the total duration of each stimulus was 410 ms (see Table 1 for specific values).[8]

[8] Note that, by manipulating the peak height and maintaining the f0 value of the L target, the slope is in turn altered. The investigation of the combinatory effects of slope and height, which implies taking into account both duration and alignment, is a topic that falls outside the scope of the present paper but needs to be investigated in further experiments.

2.1.2 Procedure

Participants were orally instructed to pay attention to the intonation of the stimuli and indicate which interpretation was more likely for each stimulus by pressing the corresponding computer key, namely 'A' for *Afirmació* ('Statement', i.e. IFS), 'C' for *Correcció* ('Correction', i.e. CFS) and 'P' for *Pregunta* ('Question', i.e. CEQ). These three labels were chosen because they would suggest intuitive response labels to participants with no previous experience with linguistic perception tasks. Prior to the experiment, subjects gave verbal confirmation to the experimenter of their understanding of the three different linguistic contexts.

The task consisted of six blocks in which all stimuli in the continuum were presented to the subjects in a randomized order, i.e. the order of the stimuli of each trial list was different for each block (with no order constraints) and for each subject. An interval of 15 seconds of silence was inserted between blocks. Subjects were instructed to press the button as quickly as they could. The experiment was set up in such a way that the next stimulus was presented only after a response had been given. The inter-stimulus interval was set at one second. We obtained a total of 1,320 responses for this experiment (11 steps × 6 blocks × 20 listeners). The experiment lasted approximately eight minutes. This included a brief training session intended to get subjects used to the stimuli and the task, which consisted of the same procedure as the experimental task with the difference that subjects were asked only to identify isolated instances of extreme and central stimuli (specifically, stimuli 1, 2, 5, 7, 10, and 11). No feedback was provided.

The experiment was set up by means of the psychology software E-prime version 2.0 (Psychology Software Tools Inc. 2009), and identification responses and RTs were automatically recorded using this software. Reaction time measures are provided relative to the offset of the stimuli.

2.2 Results

2.2.1 Identification results

Figure 3 shows the results of Experiment 1. The y-axis represents the absolute number of responses given to each stimulus. The x-axis represents the steps of the acoustic continuum. Different line types represent the different identification responses given. The graph presents a summary of how the participants categorized the acoustic space into three parts. On the one hand, it shows that the distribution of IFS and CFS responses are closer and more frequent for the lower stimuli, roughly differentiable between stimuli 1 and 4. On the other hand, the distribution of CEQ responses is clearly different from that of statements and shows a great frequency between stimuli 8 and 11. In other words, the graph shows that responses present an unsettled distribution between stimuli 5 and 7.

A Generalized Linear Mixed Model (GLMM) analysis (multinomial distribution) was performed with IDENTIFICATION of the three possible categories as the dependent variable.[9] STIMULUS was set as the fixed factor, and SUBJECT × BLOCK were set as crossed random factors (thus avoiding at the same time inter-subject variation and possible effects of fatigue, boredom, and practice). Results showed a significant effect of STIMULUS over the response given ($F_{20,\ 1299} = 19.014, p < .001$).

---

[9] All responses and RTs were analyzed through a GLMM using IBM SPSS Statistics 19.0 (IBM Corporation, 2010). As Baayen, Davidson & Bates (2008) and Quené & van den Bergh (2008) point out, mixed-effects modeling offers considerable advantages over repeated measures ANOVA. Specifically for our data, they are suitable to analyze non-continuous dependent variables, such as binomial and multinomial responses. On the other hand, we can control for both fixed and random factors (in our case, SUBJECT and BLOCK) at the same time.

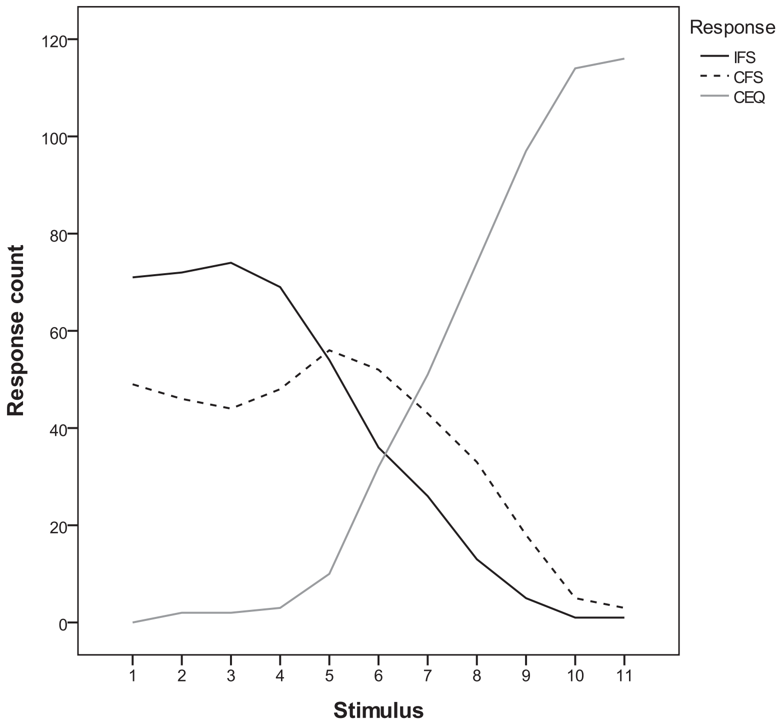

Figure 3 Absolute number of responses for each stimulus, for Experiment 1. IFS = solid black line; CFS = dashed line; CEQ = solid grey line.

Because a multinomial distribution of the dependent variable does not allow the extraction of estimated means, new GLMM analyses were conducted for each possible pair of allowed responses in the experiment (namely IFS vs. CFS, IFS vs. CEQ, and CFS vs. CEQ) in order to determine whether identification responses, when compared to one another, would show a significant distribution among the stimuli in the continuum. The overall test results showed a lower result of the Fisher's F test when applied to the comparison between the two types of statements: IFS vs. CFS ($F_{10,\,808} = 4.706$, $p < .001$), IFS vs. CEQ ($F_{10,\,913} = 24.878$, $p < .001$), and CFS vs. CEQ ($F_{10,\,888} = 25.891$, $p < .001$). This means that the different distribution of IFS and CFS among the stimuli is less clear than when each of these given responses are compared with the distribution of CEQ responses.

Table 2 shows the results of the Bonferroni deviation contrasts within each stimulus in the continuum. These results provide important information for detecting that each pair of categories has a significantly different distribution along the acoustic continuum, i.e. the distribution of responses is significantly different between the three categories for each step in the continuum. We have two exceptions to this generalization, namely, that (a) as expected, at stimulus number 6, the comparisons between IFS vs. CFS, IFS vs. CEQ, and CFS vs. CEQ are not significantly different; and (b) at stimuli 7–11, there is no significant difference between IFS and CFS, revealing that the distributions of their responses are similar.

In sum, the results of the triple response identification task indicate that Catalan listeners clearly associate the higher end of the pitch range continuum with a CEQ interpretation, and that they perceive a greater degree of ambiguity when processing the lower end of the pitch range continuum, with a very similar distribution between IFS and CFS interpretations.

Table 2 Results of the Bonferroni deviance contrasts (over each possible pair of allowed responses in the experiment: namely IFS vs. CFS, IFS vs. CEQ, and CFS vs. CEQ) within each stimulus of Experiment 1. Bold indicates $p < .05$.

	IFS vs. CFS		IFS vs. CEQ		CFS vs. CEQ	
Stimulus	*t*	Sig.	*t*	Sig.	*t*	Sig.
1	−3.332	**.008**	−4.761	**<.001**	−3.822	**.001**
2	−3.630	**.003**	−5.508	**<.001**	−4.114	**<.001**
3	−3.918	**.001**	−5.260	**<.001**	−4.110	**<.001**
4	−3.281	**.009**	−5.501	**<.001**	−4.371	**<.001**
5	−1.546	.613	−4.549	**<.001**	−3.564	**.001**
6	−0.092	1.000	−0.118	.906	0.001	.999
7	0.420	1.000	2.759	**.012**	2.905	**.008**
8	1.703	.533	5.814	**<.001**	5.748	**<.001**
9	1.973	.342	7.099	**<.001**	8.251	**<.001**
10	1.324	.743	5.600	**<.001**	8.670	**<.001**
11	0.699	1.000	5.669	**<.001**	7.954	**<.001**

IFS = information focus statement; CFS = corrective focus statement; CEQ = counter-expectational question

2.2.2 Reaction time measures

Figure 4 shows the averaged RTs for each pair of responses obtained in Experiment 1. The y-axis represents the mean RT, and the x-axis represents the steps in the pitch continuum. The graph shows a clear RT peak at stimulus 6, with a more pronounced slope towards the high end of the continuum than towards the low end.

A GLMM was applied with the RT measures as the dependent variable, STIMULUS, RESPONSE GIVEN and their interaction as fixed factors, and SUBJECT × BLOCK as crossed random factors. There were significant effects for STIMULUS ($F_{10,\,1211} = 2.732$, $p = .003$), RESPONSE GIVEN ($F_{2,\,1254} = 5.402$, $p = .005$), and their interaction ($F_{20,\,1227} = 2.379$, $p = .001$). In order to determine whether STIMULUS had a significant effect within each response given, deviation contrasts were extracted. The overall test results showed an effect of stimulus for IFS ($F_{10,\,1288} = 5.917$, $p < .001$) and CEQ ($F_{10,\,1288} = 2.318$, $p = .011$), but not for CFS ($F_{10,\,1288} = 1.766$, $p = .062$). This means that we can only argue for a RT peak when IFS and CEQ responses were analyzed.

In sum, the results of the triple response identification task indicate that Catalan listeners associate the higher end of the pitch range continuum with a CEQ interpretation, and that they display more perceptual confusion in the lower end of the pitch range continuum, which is distributed between the IFS and CFS responses. Taking into account the RT measures, this suggests a fairly close association of the lower end of the continuum with IFS responses, but no clear conclusions about the role of pitch range in determining a CFS interpretation.

3 Experiment 2: Congruity task

3.1 Methodology

This experiment consisted of a congruity task which had the goal of assessing participants' preference for a particular stimulus as more acceptable in a given communicative context. As noted above, this task makes it possible to investigate whether listeners are aware of the

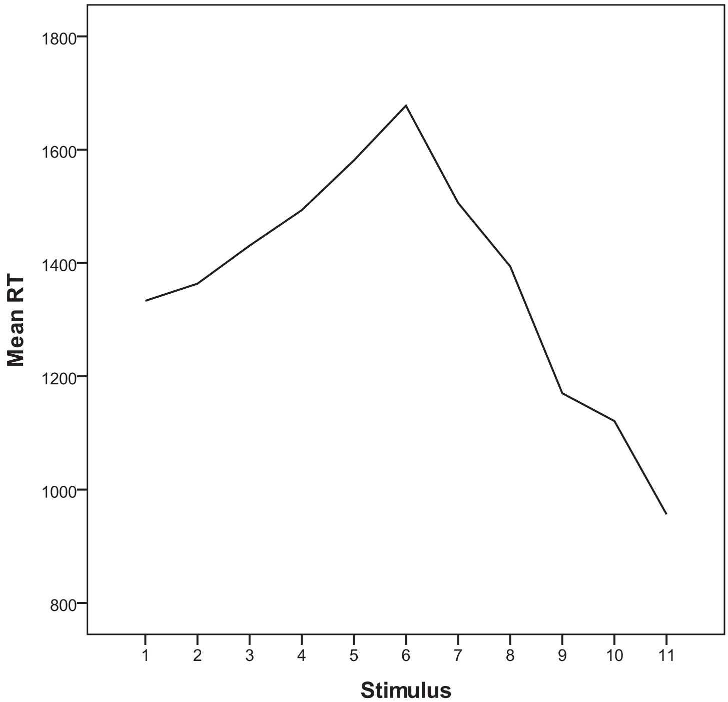

Figure 4 Averaged reaction time (RT) measures (in ms) for Experiment 1.

semantic degree of appropriateness of a particular intonation contour to a given discourse context and whether they are able to detect an incongruous use of this contour.

3.1.1 Participants and materials

For this experiment, the same set of participants was presented with the three types of linguistic contexts shown again in (3) below, each time followed by the target utterance *Petita* 'little.FEM'. The same set of subjects participated in both experiments because this would increase the comparability between the results of the two tasks. The context recordings were of a female native speaker of Central Catalan. Each context was systematically combined with all the target utterances. Their duration was approximately 1,450 ms and their pitch range was between 176.27 Hz and 299.17 Hz. An interval of approximately 300 ms of silence was inserted between the context and the target utterance.

(3)	a.	IFS	Com la vols, la cullera?	'What type of spoon do you want?'
			Petita, [sisplau].	'[I want a] little [spoon, please].'
	b.	CFS	Volies una cullera gran, no?	'You want a big spoon, don't you?'
			PETITA, [la vull, i no gran].	'[I want a] little [one, not a big one].'
	c.	CEQ	Jo la vull petita, la cullera	'I want a little spoon.'
			Petita? [N'estàs segur?]	'[A] little [one]? [Are you sure?]'

In this experiment we used six stimuli only, specifically, stimuli 1–3–5–7–9–11 from the continuum used in Experiment 1. Thus, the distance between each step in the continuum in this case was 2.4 semitones rather than 1.2.

3.1.2 Procedure

Subjects were asked to rate the target word as being semantically 'appropriate' or 'inappropriate' within that specific linguistic context by pressing the corresponding computer key, namely 'A' for *Adequat* 'appropriate' and 'I' for *Inadequat* 'inappropriate'. Thus, we obtained information about the perceived congruity of each combination of linguistic context + target stimulus. A brief training session was conducted prior to the task, consisting of rating the acceptability of stimuli 3 and 9 within each of the three communicative contexts. As in Experiment 1, the aim of the training session was merely to get participants used to the task and they received no feedback. (Stimuli 3 and 9 were chosen because they were neither extreme nor central in the auditory continuum and were equidistant from the midpoint.)

The task consisted of five blocks in which all stimuli in the continuum were presented twice within each of the three linguistic contexts in a randomized order. We thus obtained a total of 3,600 responses for this experiment (6 steps × 3 linguistic contexts × 5 blocks × 2 repetitions × 20 listeners). The experiment lasted approximately 22 minutes. No counterbalancing was used between Experiments 1 and 2 (see description below), and subjects performed a distractor behavioral task between the two experimental segments which consisted of identifying which one was the stressed syllable of a set invented words produced with seven different intonational contours.

3.2 Results

3.2.1 Identification results

Figure 5 shows the semantic congruity results of our experiment. The y-axis represents the mean perceived appropriateness between the linguistic context and the x-axis represents target stimulus. Different line types represent the linguistic contexts heard. For instance, stimulus 1 was accepted at a rate of .97 (i.e. 97% of the time) when occurring in an IFS linguistic context, .77 when occurring in the CFS context, and only .09 when occurring in the CEQ context. The opposite pattern of results was obtained for stimulus 11. Interestingly, the results reveal that stimuli 1–5 are generally rated as appropriate for both IFS and CFS contexts: while the IFS and CFS functions are similar, they sharply contrast with the function found for the CEQ linguistic context. Subjects seem to divide the six-point continuum into two general categories, i.e. 'statement' and 'question', with the boundary located at stimulus 7 (which corresponds to 158.5 Hz), thus assigning both IFS and CFS to stimuli 1–5 and CEQ to 9–11.

A GLMM analysis (binomial distribution) was conducted with appropriateness as the dependent variable, LINGUISTIC CONTEXT, STIMULUS, and their interaction as fixed factors, and SUBJECT × BLOCK as crossed random factors. Main effects of LINGUISTIC CONTEXT ($F_{2,3582} = 8.810$, $p < .001$) and STIMULUS ($F_{5,3582} = 29.284$, $p < .001$) were found and, crucially, an interaction between LINGUISTIC CONTEXT and STIMULUS ($F_{10,3582} = 92.269$, $p < .001$) was also detected.

In order to know how the three meanings are distributed in the pitch range continuum, we must analyze which part of the continuum contains a significant number of 'appropriate' and 'inappropriate' responses for each discourse context separately. To this end, Bonferroni deviation contrasts were extracted (over the two available responses, i.e. 'appropriate' and 'inappropriate') within each stimulus. The results of the deviation contrasts are presented in Table 3. The first column for each meaning contains the results of the *t*-tests (where a positive value indicates a preference for 'appropriate' responses), and the second column contains the significance of this preference (all < .001, except when stimulus 7 is presented with a CFS context). More specifically, it is shown that stimuli 1–5 were significantly categorized as 'appropriate' for an IFS context, and 7–11 were considered 'inappropriate'. For CFS, stimuli 1–5 were considered 'appropriate', stimulus 7 was not associated with any response, and stimuli 9–11 were considered 'inappropriate'. For CEQ, stimuli 1–5 were considered 'inappropriate', and stimuli 7–11 were considered 'appropriate'. The roughly parallel results for IFS and CFS indicate that both meanings share the lower part (stimuli 1–5) as appropriate

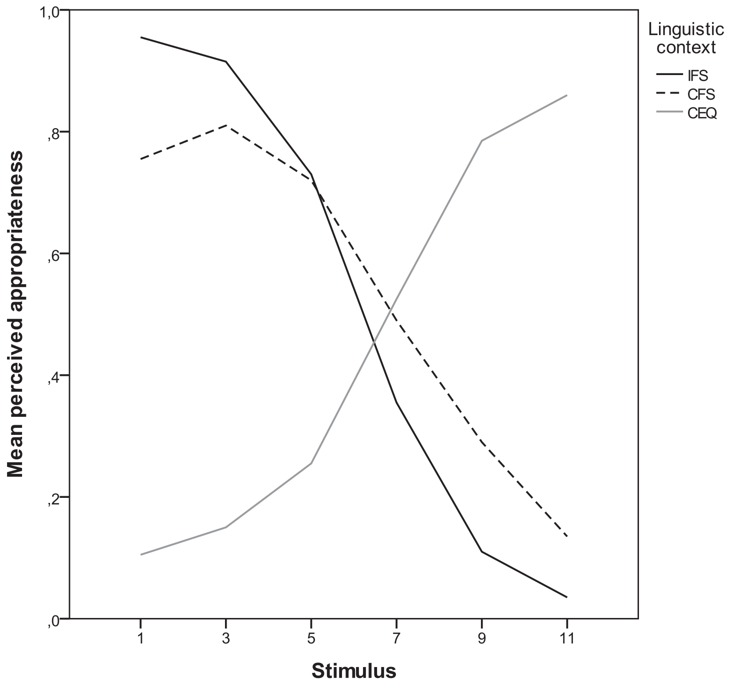

Figure 5 Mean rate of appropriateness for each type of communicative situation (IFS context: solid black line, CFS context: dashed line, CEQ context: solid grey line).

Table 3 Results of the Bonferroni deviance contrasts (applied to 'appropriate' and 'inappropriate' responses) within each stimulus, for the three linguistic contexts. Bold indicates $p < .05$ for appropriate responses.

	IFS context		CFS context		CEQ context	
Stimulus	*t*	Sig.	*t*	Sig.	*t*	Sig.
1	**10.094**	<.001	**6.735**	<.001	−9.227	<.001
3	**9.981**	<.001	**8.272**	<.001	−8.178	<.001
5	**5.538**	<.001	**5.712**	<.001	−5.181	<.001
7	−4.153	<.001	−1.207	.227	**2.849**	<.001
9	−10.210	<.001	−7.104	<.001	**10.030**	<.001
11	−10.304	<.001	−10.972	<.001	**11.637**	<.001

IFS = information focus statement; CFS = corrective focus statement; CEQ = counter-expectational question

pitch range values, whereas CEQ occupy the higher part (stimuli 7–11). This means that (a) the location of the boundary between statements (IFS and CFS) and questions (CEQ) falls immediately before stimulus 7, and (b) IFS and CFS contexts share the same perceptual behavior, both contrasting with the distribution of CEQ responses.

3.2.2 Reaction time measures

Figure 6 shows the averaged RTs obtained for each linguistic context in our congruity test. The y-axis represents the mean RT, and the x-axis represents the steps in the acoustic continuum. Specifically, the analysis of RT measures in a congruity test can shed light on the potential perceptual confusion of associating a given pitch range with a specific linguistic context (i.e. an RT peak for a specific meaning can be interpreted as indicating that that meaning has a specific pitch range for its production). The graph indicates a clear increase in RTs observed

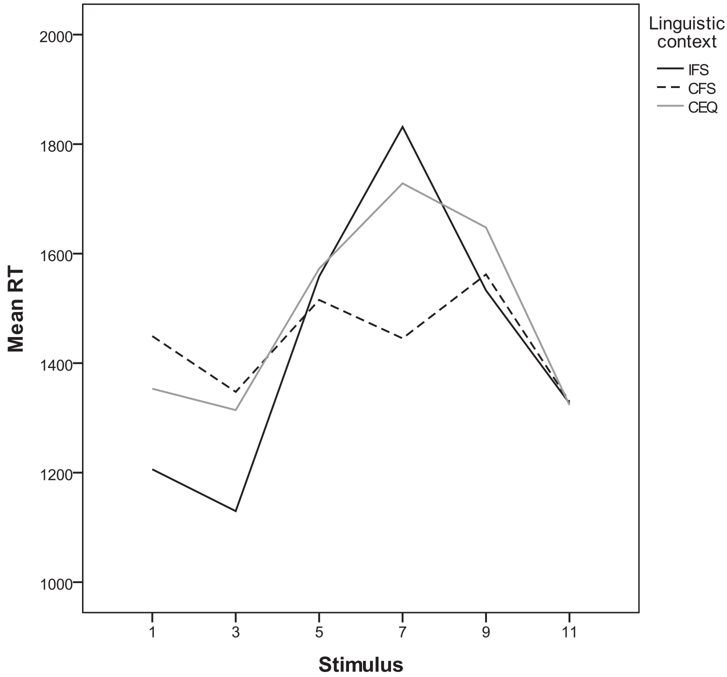

Figure 6 Averaged reaction time (RT) measures (in ms), according to linguistic contexts (IFS: solid black line, CFS: dashed line, CEQ: solid grey line).

near stimulus 4 for both IFS and CEQ contexts, but not for CFSs. This coincides with our analysis of the RT of the identification task.

A GLMM analysis (binomial distribution) was conducted, with the RT as the dependent variable, LINGUISTIC CONTEXT and STIMULUS as fixed factors, and SUBJECT × BLOCK as crossed random factors. A main effect of STIMULUS ($F_{5,\,3383} = 11{,}024$, $p < .001$) was found. There was no effect of LINGUISTIC CONTEXT ($F_{2,\,3383} = 0.796$, $p = .451$) and only a near-significant interaction between LINGUISTIC CONTEXT and STIMULUS ($F_{10,\,3383} = 1.801$, $p = .055$).

In order to analyze the patterns of RT obtained for each discourse context, deviation contrasts were extracted, with a sequential Bonferroni adjusted significance level at .05. The overall test results showed an effect of stimulus for IFS ($F_{5,\,3582} = 9.081, p < .001$) and CEQ ($F_{5,\,3582} = 4.437$, $p < .001$), but not for CFS ($F_{5,\,3582} = 1.108$, $p = .354$). The sequential Bonferroni deviation contrasts (over the RT) showed that there was a significant RT peak in stimulus 7 for IFS ($t_{3383} = 5.078, p < .001$) and CEQ ($t_{3383} = 3.021, p = .015$), but not for CFS ($t_{3383} = 0.047, p = 1$).

In sum, the robustness of the RT results of the congruity test and its coincidence with the results of Experiment 2 shows that this type of task is very informative and useful when trying to uncover the phonologically relevant contrasts in intonation.

4 Discussion and conclusions

The main goal of this study was to investigate the role of pitch accent range in conveying intonational differences in a language with a potential three-way pitch range contrast. We have investigated the potential phonological distinction between information focus statements (IFSs), corrective focus statements (CFSs), and counter-expectational questions (CEQs) in Central Catalan by performing two complementary experimental tasks.

Experiment 1 tested the participants' interpretation of each isolated stimulus using a triple response identification task. The results of this experiment showed how participants distributed the acoustic continuum across the three possible responses. They associated IFS and CEQ with the lower and higher ends of the continuum respectively, while CFS responses were less consistently associated and skewed towards the lower stimuli (see Figure 3). In order to corroborate the results from this triple identification task and also take explicitly into account the linguistic context in which these three meanings can occur, a semantic congruity test was also conducted (Experiment 2). The results showed that the lower stimuli (1–5) were judged significantly more appropriate for both IFS and CFS contexts, while the higher stimuli (7–11) were the most congruent within the CEQ context (see Figure 5). Thus, these results confirm the results from Experiment 1, namely that Catalan listeners associate the lower end of the pitch range continuum with statements (i.e. IFSs and CFSs) and the higher end of the continuum with questions.

Concerning the analysis of RT measures, as expected, they were found to correlate with the identification results and to increase for the stimuli located in the acoustic frontier between phonological distinctions. Experiment 1 showed a significant peak located at stimulus 6 and a significant role of pitch range only for IFS and CEQ interpretations. The analysis of RT measures from Experiment 2 clarifies this result because only two RT peaks were found, again for IFS and CEQ contexts (both at stimulus 7). Interestingly, the analysis of the RT from both experiments shows no significant role for pitch range when CFS is involved. Following Chen (2003; see also Pisoni & Tash 1974, Birch & Clifton 1995), if a RT peak located at a boundary between categories is taken as an indication of the discreteness of a perceived contrast, we cannot claim that participants' decisions on the appropriateness of CFS sentences are discretely distributed depending on pitch range. Our results reveal that CFS behaves approximately like IFS in terms of pitch range values, and the analysis of RT measures from our congruity experiment shows that there is no RT peak between the 'appropriate' and 'inappropriate' decisions that affect the role of pitch range for CFS marking. Therefore, this means that these two responses are not discretely divided by native listeners and so the role of pitch range for CFS marking is simply a gradient phenomenon. The IFS-like behavior and absence of a RT peak might thus be interpreted as meaning that pitch range distinguishes CFSs from IFSs in a gradient fashion. We argue that the detection of an utterance as being a CFS relies to a greater extent on a pragmatic inferencing process, such that CFS is understood when contrastive information is added to the discourse in normal conversation. Finally, the speaker can also mark the corrective status of that utterance with morphosyntactic strategies like focus fronting, as well as with postfocal prosodic reduction.

Borràs-Comes et al. (2010) tested the participants' interpretation of similar isolated stimuli in a binomial way by comparing the perception of IFS vs. CEQ and IFS vs. CFS. No differences were found between the two identification functions, which meant that, according to identification responses, CFS and CEQ would be associated with similar pitch range values.[10] However, the results of our present study show that, when participants are allowed

[10] In order to compare Borràs-Comes et al.'s (2010) results with the ones presented in this paper, we ran three binomial identification tasks with the same set of stimuli (i.e. identification tasks involving only two categories). These three binomial identification tasks took into account the three possible combinations of pragmatic interpretations (i.e. Task 1: IFS vs. CFS, Task 2: IFS vs. CEQ, and Task 3: CFS vs. CEQ). The results of the three binomial tasks were analyzed through a GLMM, with STIMULUS, TASK, and their interaction as fixed effects. A main effect of STIMULUS was found ($F_{10,3267} = 87.435$, $p < .001$), but neither a main effect of TASK ($F_{2,3267} = 0.870$, $p = .419$) nor an interaction between TASK × STIMULUS ($F_{20,3267} = 1.224$, $p = .223$) was found. This means that binomial identification tasks' results are influenced by the fact that they are forced-choice tasks in which two specific categories

to give any of the three possible responses, IFS and CFS show a similar distribution in the pitch range continuum. In line with this, we suggest that we need to use binomial identification tasks with caution, as they might be unsuitable for investigating differences in intonational categories if no additional measures (e.g. RTs or congruity tasks) are taken into account (see Chen 2003). In our view, if listeners have only two responses available for responding they can easily train themselves to categorize the given acoustic space into the two categories available (Ladd, p.c.), which is directly related to the nature of forced choices and the non-allowance of alternatives (Repp 1984, Dhar & Simonson 2003). Despite this, the results of the congruity task are slightly different from those of the triple-identification task, especially in the distribution of IFS and CFS responses among lower stimuli (different in Experiment 1, but similar in Experiment 2). In this case, the triple-identification task might still lead participants to over-categorize the stimuli among all available responses. By contrast, congruity tasks crucially take into account linguistic context, i.e. the stimuli are always evaluated for their congruity or incongruity with the preceding context.

Thus, concerning methodology, we would like to highlight the usefulness of using the response values and reaction times from semantic congruity tests to investigate the phonological status of intonational contrasts. First, this task is especially suitable for assessing linguistic contrasts, as listeners are not simply dividing the acoustic space into three categories. Second, one of the main advantages of using congruity tasks is that they take pragmatic context into account, by evaluating the degree of linguistic appropriateness of different intonation patterns within different discourse contexts. We thus argue that the exclusive use of identification tasks, even if they are complemented with reaction time measures, might not be suitable to assess the phonological role of intonation differences.

Taken together, the two experiments have crucially shown that variation in pitch range is the main phonological strategy that Catalan listeners use to discriminate between IFSs and CEQs, i.e. there is a specific threshold along a continuum of pitch range beyond which a CEQ meaning is consistently understood (according to our experiments, approximately six semitones of difference between the L and H targets). The results from the mismatch negativity analysis by Borràs-Comes et al. (2012) back up this analysis. This study found a stronger MMN brain response when inter-category stimuli (IFSs and CEQs) were presented than when listeners heard pairs of intra-category stimuli having the same physical distance between them (either two types of IFSs or two types of CEQs).

In line with our results, it is important to note that the identification of CFS in Catalan does not crucially rely on scaling differences. Recent production results reported by Vanrell et al. (2013) showed that tonal scaling is not a stable cue in distinguishing non-focal vs. contrastive focal accents in Catalan. The absence of a categorical difference between IFSs and CFSs with respect to pitch scaling might thus be related with the reported preference for Catalan to use changes in syntactic structure for corrective/contrastive focus marking (Vallduví 1991).[11] Moreover, contextual pragmatic inference can be important to detect CFS online. As stated in Levinson (2010; see also Heritage 2013), pragmatic inference works well enough to detect more than half of all the polar questions that appear in English spontaneous speech (see also Stivers 2010). In our specific case, it would be possible to classify as a CFS any contradictory utterance provided as simply the last word in a conversation (i.e. when someone contradicts the assumption of the interlocutor, then it is assumed that they know the information at issue;

---

are given to listeners. It is thus puzzling to observe how the same stimuli can be heard as having different meanings depending on the task, that is, while trinomial identification tasks identify two phonological contrasts, paired binomial tasks identify three phonological contrasts. In this sense, listeners in binomial tasks are probably just dividing the acoustic space into two parts and are less able to rely on the specific pragmatic properties of the pitch contours under comparison.

[11] Taking the example *Vull* TARONGES 'I want ORANGES' (extracted from Prieto & Cabré 2007–2013), if the speaker wants to focalize the constituent *taronges* 'oranges' (i.e. s/he wants ORANGES and not some other fruit), s/he will resort most often to clause-external detachment (*TARONGES, vull* 'ORANGES, I want').

e.g. SPEAKER A: 'Mary will come tomorrow.' SPEAKER B: 'John will come tomorrow.'). Finally, as one of the reviewers of this article pointed out, it is important to consider another possible reason for the lack of contrast found between IFSs and CFSs, namely their semantic similarity (see Krifka 2007).

It is thus clear that given the specifics of Catalan intonation, we need to be able to signal a phonological distinction between upstepped and non-upstepped rising pitch accents (i.e. counter-expectational questions [L+¡H*] and statements [L+H*]; Aguilar et al. 2009). This paradigmatic use of upstep does not correspond to the classic use of upstep and downstep in the AM system as a syntagmatic procedure. The AM system originally considered upstep to represent the raising of an H tone caused by the presence of a preceding H tone in the same prosodic phrase (see Beckman & Ayers Elam 1997, Nibert 2000, Henriksen 2012). By including a category L+¡H* in the Cat_ToBI phonological analysis (and in any other ToBI analyses), the upstep representation becomes polysemic between its paradigmatic and its syntagmatic interpretation. This is also the case with the now common use of !H with a paradigmatic use (especially in boundary tones) rather than a syntagmatic one. Thus, the issue of the scope of the upstep and downstep mechanisms within the AM model should be considered with care. The issue of phonologically contrastive H levels has been also noted by Face (2011) for Castilian Spanish, who argues for an AM transcription system which takes pitch range into account without altering the dichotomy between L and H targets that exists in the ToBI system. He proposes that an intonational domain (which can range from a pitch accent to an intonational phrase) can be specified by a 'frame' that sets 'the space for the realization of the f0 rises and falls' (Face 2011: 89). Following Face, the Catalan IFS contour might be labeled $_{0}$[L+H*], while the Catalan CEQ contour might be labeled $_{H+}$[L+H*], which would indicate that the high end of the continuum would be extended. This is an alternative transcription strategy which should be evaluated with rigor but which is beyond the scope of this investigation.

All in all, the results presented here represent new empirical evidence that pitch accent range variation can express categorical differences in meaning (Ward & Hirschberg 1985; Hirschberg & Ward 1992; Ladd 1994, 2008; Ladd & Morton 1997; Chen 2003; Vanrell 2006, 2011; Savino & Grice 2011).[12] As mentioned above, the division between two levels of pitch height to distinguish statements from questions is very productive in other Romance languages (Savino & Grice 2011 for Bari Italian, Roseano et al. 2014 for Friulian, Vanrell et al. 2014 for Sardinian, and Estebas-Vilaplana & Prieto 2010 for Castilian Spanish), as well as in other languages, and this distinction needs to be reflected in the intonational phonology of such languages. Methodologically, we believe that congruity tasks (along with reaction time measures) can be profitably used to properly investigate intonational categories.

Acknowledgements

This research has been funded by projects BFU2012-31995 (awarded by the Ministerio de Economía y Competitividad), FFI2009-07648/FILO and FFI2011-23829/FILO (both awarded by the Ministerio de Ciencia e Innovación), and 2009 SGR 701 (awarded by the Generalitat de Catalunya). The authors would like to thank Amalia Arvaniti, Jonathan Barnes, Laura Dilley, Nicholas Henriksen, Louise McNally, Janet Pierrehumbert, Brechtje Post, Stefanie Shattuck-Hufnagel, Yi Xu, and three anonymous reviewers for their helpful comments. We would also like to thank the members of the Group of Prosodic Studies (UPF, Barcelona) for their comments on earlier versions of this manuscript.

[12] The use of higher f0 peaks can be related to the general finding that the average pitch in questions is higher than the average pitch in non-questions (Ohala 1983, Bolinger 1986), something that has been analyzed as a 'discretised' manifestation of the so-called Frequency Code (Gussenhoven 1999).

References

Abramson, Arthur. S. 1979. The noncategorical perception of tone categories in Thai. In Björn Lindblom & Sven E. G. Öhman (eds.), *Frontiers of speech communication research*, 127–134. London: Academic Press.

Aguilar, Lourdes, Carme de-la-Mota & Pilar Prieto (coords.). 2009. Cat_ToBI Training Materials. http://prosodia.upf.edu/cat_tobi/ (accessed 1 October 2012).

Arvaniti, Amalia & Mary Baltazani. 2004. Greek ToBI. In Jun (ed.), 84–117.

Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, 390–412.

Beckman, Mary E. & Gayle Ayers Elam. 1997. Guidelines for ToBI labeling (version 3). Ms., The Ohio State University. [http://www.cs.columbia.edu/~agus/tobi/labelling_guide_v3.pdf, accessed 1 October 2012.]

Beckman, Mary E., Manuel Díaz-Campos, Julia Tevis McGory & Terrell A. Morgan. 2002. Intonation across Spanish, in the Tones and Break Indices framework. *Probus* 14, 9–36.

Beckman, Mary E. & Janet Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3, 255–310.

Birch, Stacy & Charles Clifton Jr. 1995. Focus, accent, and argument structure effects on language comprehension. *Language and Speech* 38(4), 365–391.

Boersma, Paul & David Weenink. 2008. Praat: Doing phonetics by computer (version 5.0.09). Computer program.

Bolinger, Dwight L. 1986. *Intonation and its uses: Melody in grammar and discourse*. Palo Alto, CA: Stanford University Press.

Borràs-Comes, Joan, Jordi Costa-Faidella, Pilar Prieto & Carles Escera. 2012. Specific neural traces for intonational discourse categories as revealed by human evoked potentials. *Journal of Cognitive Neuroscience* 24(4), 843–853.

Borràs-Comes, Joan, Maria del Mar Vanrell & Pilar Prieto. 2010. The role of pitch range in establishing intonational contrasts in Catalan. In *The Fifth International Conference on Speech Prosody*, Chicago, 100103, 1–4.

Braun, Bettina. 2006. Phonetics and phonology of thematic contrast in German. *Language and Speech* 49(4), 451–493.

Brunetti, Lisa. 2004. *A unification of focus*. Padova: Unipress.

Calhoun, Sasha. 2004. Phonetic dimensions of intonational categories – the case of L+H* and H*. In Bernard Bel & Isabelle Marlien (eds.), *Speech Prosody 2004*, Nara, Japan. 103–106.

Chen, Aoju. 2003. Reaction time as an indicator of discrete intonational contrasts in English. In *8th EuroSpeech*, Geneva, 97–100.

Crespo-Sendra, Verònica. 2011. *Aspectes de l'entonació del valencià*. Ph.D. dissertation, Universitat Pompeu Fabra, Barcelona. [http://www.tdx.cat/handle/10803/51575, accessed 1 October 2012.]

Cruschina, Silvio. 2011. Focalization and word order in Old Italo-Romance. *Catalan Journal of Linguistics* 10, 92–135.

Dhar, Ravi & Itamar Simonson. 2003. The effect of forced choice on choice. *Journal of Marketing Research* 40(2), 146–160.

Dilley, Laura C. 2010. Pitch range variation in English tonal contrasts: Continuous or categorical? *Phonetica* 67, 63–81.

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

Estebas-Vilaplana, Eva & Pilar Prieto. 2010. Castilian Spanish intonation. In Pilar Prieto & Paolo Roseano (eds.), *Transcription of intonation of the Spanish language*, 17–48. Munich: Lincom Europa.

Face, Timothy L. 2005. F0 peak height and the perception of sentence type in Castilian Spanish. *Revista de Lingüística Iberoamericana* 2(6), 49–65.

Face, Timothy L. 2007. The role of intonational cues in the perception of declaratives and absolute interrogatives in Castilian Spanish. *Estudios de fonética experimental* 16, 185–225.

Face, Timothy L. 2011. *Perception of Castilian Spanish intonation: Implications for intonational phonology*. Munich: Lincom Europa.

Face, Timothy L. & Mariapaola D'Imperio. 2005. Reconsidering a focal typology: Evidence from Spanish and Italian. *Italian Journal of Linguistics* 17(2), 271–289.
Falé, Isabel & Isabel Hub Faria. 2005. Intonational contrasts in EP: A categorical perception approach. In *9th European Conference on Speech Communication and Technology*, Lisboa, 1705–1708.
Félix-Brasdefer, J. César (2010). Data collection methods in speech act performance: DCTs, role plays, and verbal reports. In Alicia Martínez-Flor & Esther Usó-Juan (eds.), *Speech act performance: Theoretical, empirical, and methodological issues*, 41–56. Amsterdam & Philadelphia, PA: John Benjamins.
Francis, Alexander L., Valter Ciocca & Brenda Kei Chit Ng. 2003. On the (non)categorical perception of lexical tones. *Perception & Psychophysics* 65(7), 1029–1044.
Frota, Sónia & Pilar Prieto (eds.). 2014. *Intonational variation in Romance*. Oxford: Oxford University Press.
Grice, Martine, Mariapaola D'Imperio, Michelina Savino & Cinzia Avesani. 2005. Towards a strategy for labeling varieties of Italian. In Jun (ed.), 55–83.
Gussenhoven, Carlos. 1999. Discreteness and gradience in intonational contrasts. *Language and Speech* 42(2–3), 283–305.
Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
Gussenhoven, Carlos. 2007. Types of focus in English. In Chungmin Lee, Matthew Gordon & Daniel Büring (eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation*, 83–100. Heidelberg: Springer.
Henriksen, Nicholas. 2012. The intonation and signaling of declarative questions in Manchego Peninsular Spanish. *Language and Speech* 55(4), 543–576.
Heritage, John. 2013. Action formation and its epistemic (and other) backgrounds. *Discourse Studies* 15(5), 551–578.
Hirschberg, Julia & Gregory Ward. 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics* 20, 241–251.
IBM Corporation. 2010. IBM SPSS Statistics (version 19.0.0). Computer program.
Jun, Sun-Ah (ed.). 2005. *Prosodic models and transcription: Towards prosodic typology*. Oxford: Oxford University Press.
Krifka, Manfred. 2007. Basic notions of information structure. In Caroline Féry, Gisbert Fanselow & Manfred Krifka (eds.), *The notions of information structure*, vol. 6, 13–55. Potsdam: Universitätsverlag Potsdam.
Ladd, D. Robert. 1994. Constraints on the gradient variability of pitch range, or, pitch level 4 lives! In Patricia A. Keating (ed.), *Phonological structure and phonetic form: Papers in Laboratory Phonology III*, 43–63. Cambridge: Cambridge University Press.
Ladd, D. Robert. 2008. *Intonational phonology*. Cambridge: Cambridge University Press.
Ladd, D. Robert & Rachel Morton. 1997. The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics* 25, 313–342.
Lee, Ho-Young. 2004. H and L are not enough in intonational phonology. *Eoneohag* 39, 71–79.
Levinson, Stephen C. 2010. Questions and responses in Yélî Dnye, the Papuan language of Rossel Island. *Journal of Pragmatics* 42(10), 2741–2755.
Liberman, Mark Y. & Janet Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In Mark Aronoff & Richard T. Oehrle (eds.), *Language sound structure: Studies in phonology presented to Morris Halle*, 157–233. Cambridge, MA: MIT Press.
McHugh, Brian D. 1990. The phrasal cycle in Kivunjo Chaga tonology. In Sharon Inkelas & Draga Zec (eds.), *The phonology–syntax connection*, 217–242. Chicago, IL: University of Chicago Press.
Nibert, Holly J. 2000. *Phonetic and phonological evidence for intermediate phrasing in Spanish intonation*. Ph.D. dissertation. University of Illinois at Urbana–Champaign.
Nurani, Lusia M. 2009. Methodological issue in pragmatic research: Is discourse completion test a reliable data collection instrument? *Jurnal Sosioteknologi* 17(8), 667–678.
Ohala, John J. 1983. Cross-language use of pitch: An ethological view. *Phonetica* 40(1), 1–18.

Pierrehumbert, Janet. 1980. *The phonetics and phonology of English intonation*. Ph.D. dissertation, MIT.
Pisoni, David B., & Jeffrey Tash. 1974. Reaction times to comparisons within and across phonetic categories. *Perception & Psycholinguistics* 15(2), 284–290.
Prieto, Pilar. 2002. Entonació. In Joan Solà, Maria Rosa Lloret, Joan Mascaró & Manel Pérez Saldanya (eds.), *Gramàtica del català contemporani*, vol. 1, 1395–1462. Barcelona: Empúries.
Prieto, Pilar. 2004. The search for phonological targets in the tonal space: H1 scaling and alignment in five sentence-types in Peninsular Spanish. In Timothy L. Face (ed.), *Laboratory approaches to Spanish phonology*, 29–59. Berlin: Mouton de Gruyter.
Prieto, Pilar. 2005. Stability effects in tonal clash contexts in Catalan. *Journal of Phonetics* 33(2), 215–242.
Prieto, Pilar. 2014. The intonational phonology of Catalan. In Sun-Ah Jun (ed.), *Prosodic Typology 2*, 43–80. Oxford: Oxford University Press.
Prieto, Pilar, Lourdes Aguilar, Ignasi Mascaró, Francesc Torres-Tamarit & Maria del Mar Vanrell. 2009. L'etiquetatge prosòdic Cat_ToBI. *Estudios de Fonética Experimental* 18, 287–309.
Prieto, Pilar & Teresa Cabré (coords.). 2007–2013. Atles interactiu de l'entonació del català. http://prosodia.upf.edu/atlesentonacio/ (accessed 1 October 2012).
Psychology Software Tools Inc. 2009. E-Prime (version 2.0). Computer program.
Quené, Hugo & Huub van der Bergh. 2008. Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language* 59, 413–425.
Rathcke, Tamara & Jonathan Harrington. 2006. Is there a distinction between H+!H* and H+L* in standard German? Evidence from an acoustic and auditory analysis. In Rüdiger Hoffmann & Hansjörg Mixdorff (eds.), *Speech Prosody 2006*, 783–786. Dresden: TUDpress Verlag der Wissenschaften GmbH.
Rathcke, Tamara & Jonathan Harrington. 2010. The variability of early accent peaks in Standard German. In Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio & Nathalie Vallée (eds.), *Laboratory Phonology 10*, 533–555. Berlin & New York: Mouton de Gruyter.
Repp, Bruno H. 1984. Categorical perception: Issues, methods, findings. In Norman J. Lass (ed.), *Speech and Language: Advances in basic research and practice*, vol. 10. New York: Academic Press.
Roseano, Paolo, Maria del Mar Vanrell & Pilar Prieto. 2014. Intonational phonology of Friulian and its dialects. In Frota & Prieto (eds.).
Savino, Michelina & Martine Grice. 2011. The perception of negative bias in Bari Italian questions. In Sónia Frota, Pilar Prieto & Gorka Elordieta (eds.), *Prosodic categories: Production, perception and comprehension*, 187–206. Heidelberg: Springer.
Solà, Joan. 1990. L'ordre dels mots en català. Notes pràctiques. In Joan Solà (ed.), *Lingüística i normativa*, 91–124. Barcelona: Empúries.
Stivers, Tanya. 2010. An overview of the question–response system in American English conversation. *Journal of Pragmatics* 42(10), 2772–2781.
Stockwell, Robert P., Donald J. Bowen & Ismael Silva-Fuenzalida. 1956. Spanish juncture and intonation. *Language* 32(4), 641–665.
Trager, George L. & Henry L. Smith (1951). *An outline of English structure* (Studies in Linguistics Occasional Papers 3). Norman, OK: Battenberg Press.
Vallduví, Enric. 1991. The role of plasticity in the association of focus and prominence. *The Eastern States Conference on Linguistics 7*, 295–306. Columbus, OH: Ohio State University Press.
Vanrell, Maria del Mar. 2006. A tonal scaling contrast in Majorcan Catalan interrogatives. *Journal of Portuguese Linguistics* 6(1), 147–178.
Vanrell, Maria del Mar. 2011. *The phonological relevance of tonal scaling in the intonational grammar of Catalan*. Ph.D. dissertation, Universitat Autònoma de Barcelona, Cerdanyola del Vallès. [http://prosodia.upf.edu/home/arxiu/tesis/doctorat/tesi_vanrell.pdf, accessed 1 October 2012.]
Vanrell, Maria del Mar, Francesc Ballone, Carlo Schirru & Pilar Prieto. 2014. Sardinian intonational phonology: Logudorese and Campidanese varieties. In Frota & Prieto (eds.).
Vanrell, Maria del Mar, Antonio Stella, Barbara Gili-Fivela & Pilar Prieto. 2013. Prosodic manifestations of the Effort Code in Catalan, Italian and Spanish contrastive focus. *Journal of the International Phonetic Association* 43(2), 195–220.
Ward, Gregory & Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61, 747–776.