

# A tonal scaling contrast in Majorcan Catalan interrogatives

MARIA DEL MAR VANRELL BOSCH

## Abstract

*This paper reports the application of the Categorical Perception paradigm to a pitch height contrast in the nuclear accent between yes-no and what-questions in Majorcan Catalan. Using two natural tokens produced by a female speaker, two intonational continua were created, from yes-no to what-question contour and vice versa, by shifting the peak in 4 steps of 15 Hz each. 42 Majorcan Catalan listeners participated in a two-part experiment, consisting of an identification and a discrimination task. The results from the identification task showed that it is possible to switch the perceived category by manipulating the pitch height of the leading tone. Also, Reaction Times were shorter within categories and longer between categories. Discrimination results revealed that the shift in the identification function corresponded to the peak in the discrimination function. The comparisons between obtained and predicted discrimination results indicated that discrimination can be predicted from identification results on the basis of phonetic categorization. These results confirmed that the difference in pitch height of the leading tone in nuclear accent for yes-no and what-questions in Majorcan Catalan is discrete and has a phonological character. In addition, the discrimination results revealed that Majorcan listeners are more sensitive to F0 differences when the first token is lower in frequency than the second.*

## 1. Introduction

In Majorcan Catalan,<sup>1</sup> yes-no questions and what-questions are characterized by a falling nuclear accent H+L\*, that is, an H leading tone aligned with the

---

<sup>1</sup> Majorcan Catalan is a dialect of Catalan spoken in Majorca, the largest of the Balearic Islands, with a population of roughly 750,000 inhabitants. Majorcan Catalan belongs to the Eastern Catalan dialect group along with the other Balearic subdialects, Central Catalan, Rossellonese and Algerese.

pretonic syllable and a L\* tone associated with the word-final stressed syllable. Typically, yes-no questions are headed by the unaccented interrogative particle *que*<sup>2</sup> ('that') (Figure 1, upper panel), and this can be compared with what-questions headed by the accented wh-particle *què* ('what') (Figure 1, bottom panel). Results from a Map-Task recording by Payà & Vanrell (2005), Vanrell (2006) and my own speech demonstrated that in yes-no questions the H leading tone is upstepped (Figure 1, upper panel). As can be seen in Figure 1, the difference in tonal height of the H leading tone consequently involves a difference in the intonation pattern of these two types of questions in terms of realization. Thus, even though both types of questions are characterized by a falling final intonation, what-questions show a steady high tone which extends from the beginning of the sentence to the last stressed syllable, with the falling pitch movement aligned with the last stressed syllable in the utterance (i.e., *mol*, in *mol dre*). By contrast, yes-no questions have a well-defined rising slope over the two syllables preceding the nuclear accent (i.e., *mol*). In Figure 1, bottom panel, the accented syllable *li* of *volia* is realized in a high tone, there is a further rise in pitch on the last syllable of this word, and then there is a steeper falling slope until the end of the nuclear syllable (*mol*). In both types of sentences, we observe low boundary tones.

This intonational difference is particularly relevant in Majorcan Catalan because both sentences in Figure 1 are homophonic at the segmental level, since the two interrogative particles (accented and unaccented) are pronounced as [ˈkə]/[kə]. This does not occur in other varieties of Catalan, in which the accented interrogative particle *què* is pronounced as [ˈkɛ] and the unaccented interrogative particle *que* is pronounced as [kə].

Pitch height variation has been assumed to be paralinguistic by the standard Autosegmental-metrical (AM) model (Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988), that is, it exclusively expresses a change in the degree of meaning corresponding usually to differences in emphasis or prominence. However, some studies have shown that the difference in pitch height can also trigger categorical effects. Ladd & Morton (1997) applied the Categorical Perception paradigm (CP) to a contrast between two pitch-accents in English, the normal high accent and the emphatic high accent. Evidence was found for a shift between these two categories in identification but a related discrimination peak was not found. As the main assumption of classical CP is that discrimination is easier at the category boundary and more difficult within categories, they concluded that this contrast was not categorically perceived. On the basis of the results of Ladd & Morton (1997), Chen (2003) argued that the absence of peak discrimination might be related not to the nonexistence of categorical perception but rather to a hypothetical

---

<sup>2</sup> See Rigau & Prieto (2005) for a syntactic and prosodic description of Catalan yes-no questions headed by interrogative marker *que*.

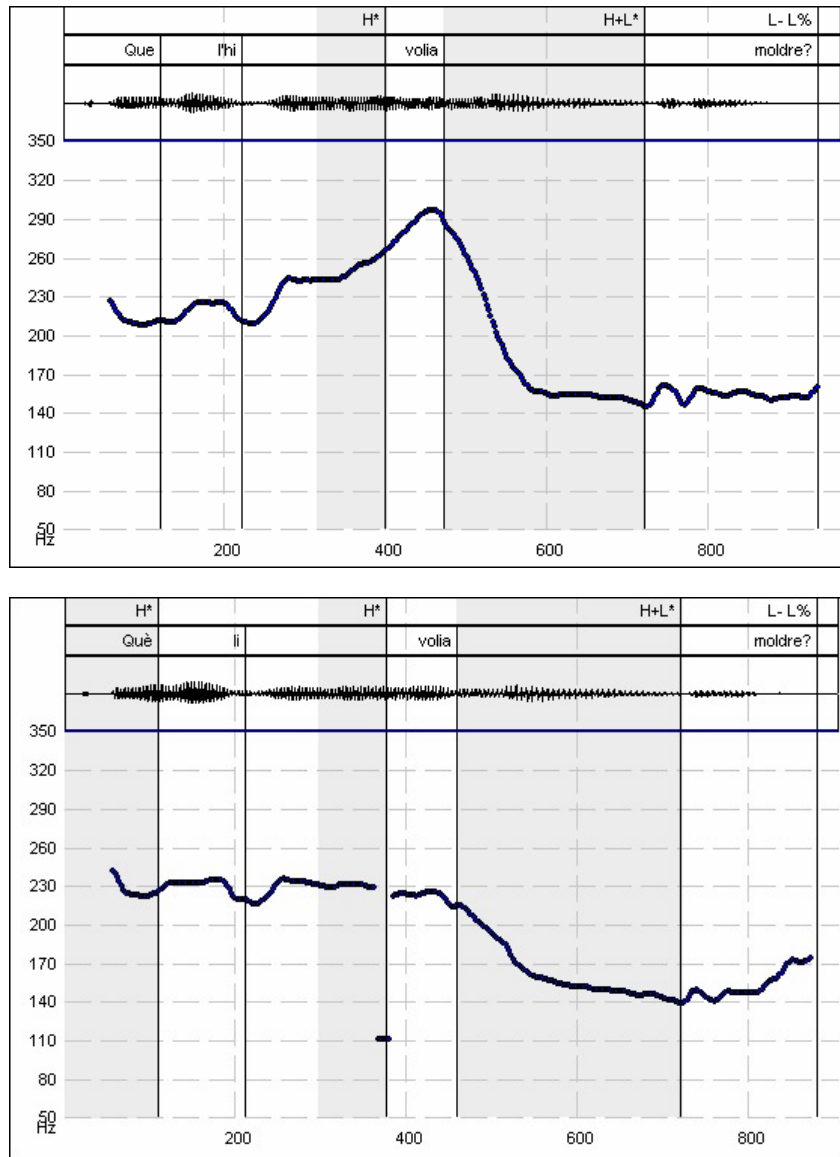


Figure 1. Waveforms and fundamental frequency contours of a yes-no question *Que l'hi volia moldre?* [kə li vu,liə 'mɔɫdrə] ('Did s/he want to grind it for him/her?'), upper panel, and a what-question *Què li volia moldre?* ('What did s/he want to grind for him/her?') [kə li vu,liə 'mɔɫdrə], bottom panel. Accented syllables are shaded.

unsuitability of applying the CP paradigm to a pitch height contrast. This unsuitability derives from the fact that according to CP, listeners would be

incapable of perceiving differences between two stimuli when they belong to the same category; yet in Ladd & Morton (1997) listeners were indeed able to perceive differences in F0 across the full continuum. For that reason, according to Chen (2003), the CP paradigm may be inappropriate for examining discreteness in differences in peak height. Chen thus proposed an alternative method, which she called the Reaction Time (RT) measurement, to examine the nature of two intonational contrasts: normal high accent vs. emphatic high accent and early peak alignment vs. late peak alignment. Noting that mean RT was shortest for within-category identification and longest for across-category identification for the peak height continuum but not for the peak alignment contrast, and basing herself on previous studies (Pisoni & Tash, 1974), Chen concluded that this peak height contrast is discrete. Prieto (2003) studied the effects of sentence type (statements, yes-no questions, wh-questions, commands, and exclamatory sentences) on scaling variation of the peak of the IP-initial pitch accent in Castilian Spanish, and found that pitch height is not just related to paralinguistic usage, since sentence-type also has a strong effect on the scaling of the first pitch accent in the utterance in Spanish. In a production and a perception experiment, Calhoun (2003) 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. Face (2005) carried out a gating perception experiment to examine the disambiguating role of intonation in the perception of two sentence types in Castilian Spanish. It was found that the different scaling of the first F0 peak between declaratives and absolute interrogatives is the cue that leads to 95% accuracy in the perception of sentence type in Castilian Spanish. In this context, Majorcan Catalan provides a good test case to show that degrees of pitch height may not be uniquely related to paralinguistic variation, but can play a decisive role in differentiating two different utterance types: yes-no questions and what-questions.

The Categorical Perception paradigm (CP) has been one of the most commonly used methods to examine the nature of tonal contrasts and has been applied to differences in peak alignment (Kohler, 1987; D'Imperio & House, 1997; Chen, 2003) and differences in pitch height in both tonal languages (Francis et al., 2003; Francis & Ciocca, 2003) and intonational languages, for boundary tones (Remijsen & van Heuven, 1999; Post, 2000; Schneider & Linfert, 2003; Cummins et al., 2006; Falé & Faria, 2006) as well as for pitch accents (Ladd & Morton, 1997; Chen, 2003).

The application of the CP paradigm involves the presence of asymmetries in discrimination results. Asymmetries in tonal perception occur when the discrimination of a tonal change presented in one direction is easier compared to the same change presented in the reverse direction. These asymmetries have been found repeatedly in the application of CP to the perception of not only intonational languages (Kohler, 1987; Ladd & Morton, 1997; Remijsen & van Heuven, 1999; Schneider & Linfert, 2003; Cummins et al., 2006; Falé

& Faria, 2006) but also tonal languages (Francis & Ciocca, 2003). Studies that report asymmetries in the application of the CP paradigm to pitch height contrast seem to agree that two different contours are more successfully discriminated when the second one – whatever two points in a contour are compared – has a higher pitch.

Accordingly, the primary goal of the present study is to test through an experimental procedure based on the Categorical Perception paradigm (CP) whether listeners make categorical linguistic use of F0 scaling differences in perceiving yes-no questions as opposed to what-questions in Majorcan Catalan.

Since asymmetries have been reported repeatedly in the literature on the application of the CP paradigm to pitch height contrasts, a second goal of the study is to verify the presence of asymmetries. It is predicted that these effects do occur, such that it will be easier to discriminate between pairs of stimuli when the direction of change is upwards in pitch level than when it is downwards (Ladd & Morton, 1997; Remijsen & van Heuven, 1999; Schneider & Linfert, 2003; Cummins et al., 2006; Falé & Faria, 2006).

## **2. Method**

In the CP paradigm, subjects perform two tasks: an identification task and a discrimination task. In the identification task, subjects listen to randomly ordered stimuli constructed from a continuum and judge which of two categories each stimulus represents. In the discrimination task, the subjects listen to the stimuli again, but this time they are asked to identify the test stimulus in terms of a reference stimulus. In the pair-wise AX task, the subjects hear the test stimulus and a single reference stimulus, and decide whether the two stimuli are two instances of the same stimulus or different stimuli. The patterns of results expected are shown in Figure 2. The idealized functions of responses to the identification task have an S-shape (solid lines), i.e. an abrupt shift from one category to the other rather than a gradual transition. Figure 2 also shows the idealized discrimination function (dashed line). If perception is categorical, discrimination is easier when the two stimuli straddle the boundary between the categories than when the two stimuli are from within the same category. A stimulus continuum is typically considered categorical if listeners' responses match two criteria (Francis et al., 2003): first, identification proportions should predict discrimination accuracy (Liberman et al., 1957); second, peaks of discrimination should correspond to the location of the category boundaries determined by identification (Repp et al., 1979). Typically, discrimination results are predicted through a formula taken originally from Liberman et al. (1957) and the boundary between categories in sigmoid response curves is calculated by Probit analysis, which fits a cumulative normal curve to probability estimates as a function of stimulus level by the method of least squares (Finney, 1971), estimating the

mean (which marks the identification crossover) and standard deviation for each distribution. In this study the CP paradigm has not been applied in the classical sense. Firstly, in the identification task, reaction time as well as the response rate is measured. Secondly, another formula is used in addition to the formula taken from Liberman et al. (1957) to predict the discrimination performance from the identification proportions. And finally, the statistical analyses employed are different from the ones that researchers typically use in this type of study (see details on these alternative or additional tools in the relevant subsections).

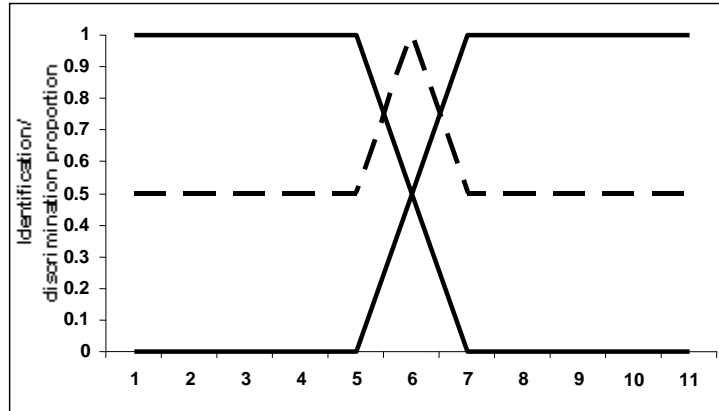


Figure 2. Idealized identification (solid lines) and discrimination (dashed lines) functions.

### 2.1. Stimuli

One token of the yes-no question *Que l'hi duries?* (Would you take it to him/her?) and one token of the what-question *Què li duries?* (What would you take him/her?), based on the production results of Vanrell (2006), were produced by a 24-year-old native female speaker of Majorcan Catalan. Both tokens are homophonic at the segmental level (see Figure 1). In the yes-no question token the leading tone was 263 Hz while in the what-question token it was 203 Hz. A linear stylization of the rising-falling movement was carried out. Three points were interpolated: a point at rising onset L1, a point at the peak H, and a point at the falling offset L2. L1 was aligned in both tokens at the onset of the pretonic syllable DU, H at the offset of the vowel of the stressed syllable DU, and L2 at the offset of the vowel of the syllable RI. From these two base tokens, ten stimuli were created by means of PSOLA synthesis: two synthesized base tokens (one from the yes-no question token and one from the what-question token), four stimuli created by shifting the peak downwards from the yes-no question synthesized token (Figure 3, left panel) and four stimuli by shifting the peak upwards from the what-question

synthesized token (Figure 3, right panel) in four steps of 15 Hz each. Stimulus 1 is always 203 Hz, stimulus 2 is 218 Hz, stimulus 3 is 233 Hz, stimulus 4 is 248 Hz and stimulus 5 is 263 Hz (see Table I). The choice of 15 Hz for the step-size used in this study was motivated by the verification through a pilot test that this difference in Hz would not be too easily perceptible, especially in the discrimination task, in which the listeners reported perceiving hardly any difference between the two stimuli in a pair.

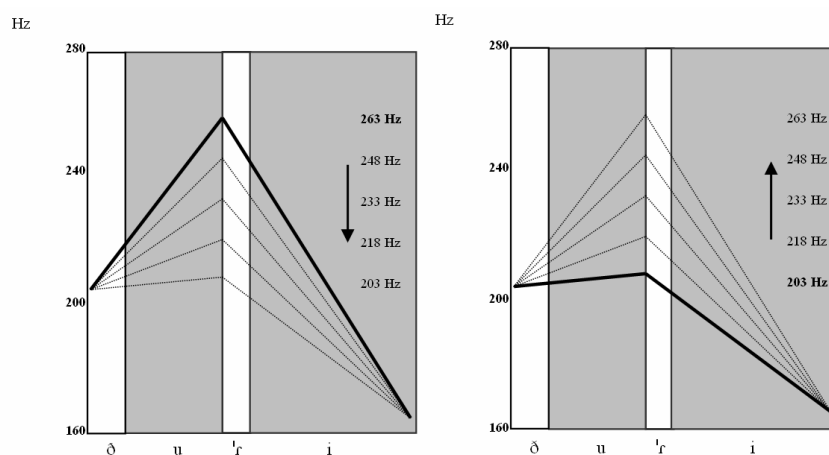


Figure 3. Schematized creation of the stimuli from yes-no question base stimulus (left panel) and what-question base stimulus (right panel).

## 2.2. Tasks and experimental procedures

Subjects were seated in front of a laptop in a quiet room and heard the stimuli over headphones. The perception test was played by means of PERCEVAL (Laboratoire Parole et Langage), software for performing computerized auditory and visual perception experiments, which also records RTs. As we were also interested in RT measurements, the listeners were instructed to rest their hands near the keyboard and to press the keys as fast as they could, but never before the end of the utterance. The identification task preceded the discrimination task and there was no break between the two tasks. In both tasks, subjects were given written instructions about how they were to respond. There was a practice block before both the identification and the discrimination test block. The full test lasted approximately 30 minutes.

### 2.3.1. Identification task

The materials for the identification task consisted of 4 repetitions of each of the 10 stimuli (five stimuli from the yes-no question token and five stimuli from the what-question token). These 40 stimuli (5 stimuli x 2 question types

x 4 repetitions – see Table I) were presented in blocks of 10 in random order. There was a practice block before the test session made up of the two synthesized base stimuli plus the eight stimuli created from the synthesized base tokens by shifting the peak. There was no break between the blocks. The subjects were asked to respond after each stimulus according to how they would answer the question in a real situation. In other words, if they perceived the yes-no question *Que l'hi duries?* (Would you bring it to him/her?), they were to press the “S” key on the keyboard (for Sí= “Yes”), whereas if they perceived the what-question *Què li duries* (What would you bring him/her?), they were to press the “A” key (for Això= “That”).

Stimulus	F0 value of the peak
1	203 Hz
2	218 Hz
3	233 Hz
4	248 Hz
5	263 Hz
5 stimuli	
5 stimuli x 2 question types x 4 repetitions + 10 practice stimuli = 50 stimuli (presented in blocks of 10 in random order).	

*Table I. Stimuli for the identification task.*

### 2.3.2. Discrimination task

The materials for the discrimination task consisted of pairs of stimuli taken from the identification task. Eight pairs of stimuli were created in AB order, meaning that stimulus B is always higher in frequency than stimulus A (four from the yes-no-question-based continuum and four from the what-question-based continuum): pair 203-218 Hz, pair 218-233 Hz, pair 233-248 Hz and pair 248-263 Hz. Eight pairs of stimuli were created in BA order in which stimulus A was lower in frequency than stimulus B (four from the yes-no-question-based continuum and four from what-question-based continuum): pair 218-203 Hz, pair 233-218 Hz, pair 248-233 Hz and pair 263-248 Hz. Additionally, 10 AA pairs were created which contained two identical stimuli (five from the yes-no-question-based continuum and five from the what-question-based continuum): pair 203-203 Hz, pair 218-218 Hz, pair 233-233 Hz, pair 248-248 Hz, pair 263-263 Hz. Two repetitions of all these stimuli



were randomized. The practice block before the test session was based on the yes-no-question-based continuum for half the listeners and on the what-question-based continuum for the other half. Note that there were no blocks made up of pairs of stimuli created from the yes-no question and what-question base stimuli all together in order to avoid blocks of 26 pairs of stimuli. It was felt that this would have tired the subjects and that it was important to keep the duration of the experimental session under an hour without breaks. Thus, each subject heard a total of 65 pairs of stimuli (4 AB pairs + 4 BA pairs + 5 AA pairs x 2 question types x 2 repetitions + one practice block of 13 pairs – see Table II). Subjects were asked to decide whether they heard the pair of stimuli as same or different. If the two stimuli sounded the same, they were to press the “I” key on the keyboard (for *Igual* ‘same’) and if the stimuli sounded different, they were to press the “D” key (for *Diferent* ‘different’). The interval between the two stimuli in each pair was 0.5 seconds. The order of the blocks was counterbalanced, that is, the half of the listeners whose practice block was based on the yes-no question base stimulus, started with the block based on the what-question base stimulus and went on to the block based on the yes-no question base stimulus, alternating thus until they had listened to the four test blocks; the other half of the listeners heard a practice block based on the what-question and thereafter heard the test blocks in the reverse order relative to the first group of subjects.

AB	4 pairs
Pair 1_2	203-218 Hz
Pair 2_3	218-233 Hz
Pair 3_4	233-248 Hz
Pair 4_5	248-263 Hz
BA	4 pairs
Pair 2_1	218-203 Hz
Pair 3_2	233-218 Hz
Pair 4_3	248-233 Hz
Pair 5_4	263-248 Hz
AA	5 pairs
Pair 1_1	203-203 Hz
Pair 2_2	218-218 Hz
Pair 3_3	233-233 Hz
Pair 4_4	248-248 Hz
Pair 5_5	263-263 Hz
13 pairs of stimuli	
13 pairs of stimuli x 2 question types x 2 repetitions + 13 pairs of practice stimuli = 65 pairs of stimuli (presented in blocks of 13 in random order).	

Table II. Pairs of stimuli for the discrimination task.

### 2.3. *Subjects*

Forty-two native speakers of Majorcan Catalan (twenty-five female speakers and seventeen male speakers), between 16 and 41 years old, participated in the experiment. None of them reported a history of hearing disability. Subjects had to achieve a pre-established level of identification accuracy whereby 80% of the base stimuli had to be recognized. The responses of those listeners who failed to identify 80% of the base stimuli were rejected. The data from 10 subjects was discarded for that reason. Finally, only the data of 32 subjects were analyzed.

### 2.4. *Statistical analyses*

We used non-parametric tests because they do not require the assumption of a normal population.

The statistical test used was the Wilcoxon matched pairs signed rank test (see e.g.: Blalock, 1979). The significance level was fixed at 0.05 and the results were obtained by SPSS 14 statistics software.

The Wilcoxon matched pairs signed rank test was used to:

- compare the identification rate between two conditions (e.g.: stimulus 1 on the continuum vs. stimulus 2 on the continuum) in determining the location of the boundary shift.
- compare the values of RT between two conditions (e.g.: stimulus 1 on the continuum vs. stimulus 2 on the continuum) in determining whether subjects were significantly faster at identifying stimuli across categories relative to stimuli within categories.
- compare the “different” response rate between two conditions (e.g.: pair 1\_2 in the continuum vs. pair 2\_3 in the continuum) in each order of presentation in determining where the real discrimination peak was located in the discrimination results.
- analyze the rate of “different” responses between AB (low-high) and BA (high-low) orders of presentation for each pair of stimuli (e.g.: pair 1\_2 vs. pair 2\_1).
- compare the discrimination rate predicted by the formulas taken from Liberman et al. (1957) and Godfrey et al. (1981) with the actual discrimination rate.
- compare the difference between hit rate and false alarm rate ( $d'$ ) between AB and BA orders of presentation for each pair of stimuli.

The tables that report the results from the Wilcoxon matched pairs signed rank tests include the  $z$ - and  $p$ -values. Since in most cases the Wilcoxon matched pairs signed rank statistic was used for multiple tests, the post-hoc Bonferroni correction was applied by adjusting the  $p$ -values.

### 3. Results

#### 3.1. Identification results

Figure 4 shows the identification rate for the two continua created from the yes-no question (in black) and what-question (in grey) base stimuli. The “identification rate” is defined as the number of “yes-no question” responses (in yes-no-question-based stimuli) or “what-question” responses (in what-question-based stimuli) over the total responses. As can be seen, the functions present the expected S-shape. The frequency range from 203 Hz to 218 Hz (i.e., stimuli 1-2) seems to correspond to the “what-question” category and the 248-263 Hz range (i.e., stimuli 4-5) to the “yes-no question” category. Thus, the transition between the two categories would correspond to the 218-248 Hz range (i.e., stimuli 2-4). However, if we examine Figure 4 carefully, we see that the exact boundary between the two categories may be located at the specific range values between 218 Hz (stimulus 2) and 233 Hz (stimulus 3) because it is between these two stimuli where the biggest difference in identification rate is observed for both continua created. That the midpoint between stimuli 2 and 3 is the crossing point or boundary between the two categories is also apparent by the fact that it is in this point where the two categories are harder to identify, shown by the identification rate of 0.6.

It is worth noting that the responses to the two continua behave differently. This is particularly noteworthy in two respects. Firstly, in the yes-no-question-based continuum, stimulus 3 (233 Hz) triggered a high rate of yes-no question responses (0.76), whereas for the what-question-based continuum the identification rate is around 0.5 for this stimulus. Secondly, we see that stimulus 5 (263 Hz) created from the what-question base stimulus triggered a higher rate of identification as what-question than would be expected. We should recall that, as noted above, the real boundary between the categories would be located between the range of 218 and 233 Hz. Thus, stimulus 1 (203 Hz) and stimulus 2 (218 Hz) would correspond to the “what-question” category while stimulus 3 (233 Hz), stimulus 4 (248 Hz) and stimulus 5 (263 Hz) lie in principle within the “yes-no question” category. This indeed seems to be the case at least for the yes-no question based stimuli. However, for the what-question-based continuum, it is not so clear that stimulus 3 (233 Hz) belongs to the “what-question” category, since its identification rate as “what-question” is higher than the identification rate as yes-no question of stimulus 3 in the yes-no-question-based continuum. Consequently, stimulus 3 receives identification rates that set it neither in the what-question category neither in the yes-no-question category. That is, it receives identification rates that make it lie in between stimulus 2 (identified clearly as what-question) and stimulus 4 (clearly yes-no-question, or non-what-question). It is speculated that this difference in the location of the boundary in the what-question-based continuum may be caused by the presence of the accented interrogative particle, which would weaken the effect of peak height and would lead

listeners to identify stimulus 3 more as “what-question” than the identification rate as yes-no question of stimulus 3 in the yes-no-question-based continuum.

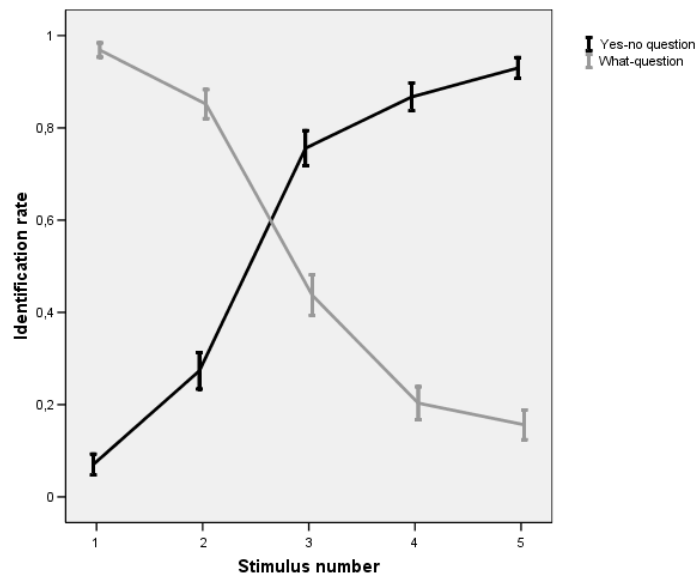


Figure 4. Identification functions for the continua created from yes-no question and what-question base stimuli. Error bars represent the standard error of the mean.

Likewise, the high identification rate as what-question for stimulus 5 created from the what-question base stimulus could be explained by the accented interrogative particle exerting a similar effect.

This explanation is supported by the results of Vanrell (2006), in which the identification results were analyzed according to the subjects' gender and musical background, since gender differences have been reported in production and perception experiments (Jensen & Carlin, 1981; Johnson et al., 1999; Rogers, 2003), as well as differences in the accuracy of perception depending on the degree of musical training (Glenn Schellenberg, 2002; Cummins et al., 2006). The identification results for qualified musicians displayed a clear frontier region between the two categories, that is, between stimulus 2 (218 Hz) and stimulus 3 (233 Hz) in both continua, and there was no high identification rate as what-question for either stimuli 3 or stimulus 5 in the what-question-based continuum. In trying to provide an explanation it was suspected that, because of their occupational training, these subjects were more attuned to changes in pitch height and during the experiment focused on tonal changes, while paying little attention to the presence of the accented interrogative particle.

Although there are clear differences between the respective responses to the two continua, it is important to emphasize that the boundary shift in identification is very robust: the functions are unmistakably S-shaped in spite of the effect on subjects' perception of the accented interrogative particle.

Table III reports the standard error of the mean for every stimulus. Stimulus 1 and stimulus 5 for both base stimuli have lower standard errors. This means that for stimulus 1 and 5 subjects agreed in their responses because these stimuli represent the canonical categories. On the other hand, stimulus 3, which would be the crossover stimulus particularly in the what-question-based continuum, has a higher standard error, which shows less agreement among subjects. However, for the yes-no-question-based continuum it is stimulus 2 that gets the highest standard error of all stimuli. Note that identification rates get higher standard errors as they get closer to the most ambiguous rate, 0.5. Stimulus 2 in yes-no-question-based continuum receives an identification rate of approximately 0.3, so it shows a higher standard error than stimulus 2 of what-question-based continuum, which gets a rate of approximately 0.8-0.9. Stimulus 3 in yes-no gets a rate of almost 0.8, so its standard error is not as high as that of stimulus 2, or as high as that of stimulus 3 in what-question-based continuum, which gets a rate between 0.4 and 0.5.

	Stimulus 1	Stimulus 2	Stimulus 3	Stimulus 4	Stimulus 5
Yes-no-question-based stimuli	0.023	0.040	0.038	0.030	0.023
What-question-based stimuli	0.015	0.032	0.044	0.036	0.032

*Table III. Standard error of the mean of the identification rate for yes-no-question- and what-question-based stimuli.*

Table IV shows the results of four Wilcoxon matched pairs signed rank test comparing the identification rate between adjacent stimuli (e.g.: stimulus 1 vs. stimulus 2, stimulus 2 vs. stimulus 3, and so on) for each continuum. According to the CP paradigm, there should be significant differences between the response rate for stimulus 3 and stimulus 2, as they belong to different categories in the case of yes-no-question-based continuum. However, in the case of the what-question-based continuum, since stimulus 3 is the crossover stimulus, there should be significant differences between stimuli 2 and 3 and between stimuli 3 and 4. As can be seen in Table IV, the

results are nearly what we would expect: stimulus 3 is significantly different from stimulus 2 in the yes-no-question-based continuum and significantly different from both stimulus 2 and stimulus 4 in the what-question-based continuum.<sup>3</sup>

		Stimulus 1-2	Stimulus 2-3	Stimulus 3-4	Stimulus 4-5
Yes-no- question- based stimuli	z-value	-4.333	7.410	-2.271	-2.000
	p-value	<0.004	<0.004	NS	NS
What- question- based stimuli	z-value	-3.441	-6.786	-4.243	-1.500
	p-value	0.004	<0.004	<0.004	NS

*Table IV. z- and p-values of four Wilcoxon matched pairs signed rank tests comparing the identification rate between adjacent stimuli for yes-no-question- and what-question-based stimuli.*

### 3.1.2. Reaction Time

In order to claim the categoriality of a contrast, it is not enough that identification results have a clear category boundary, since this category boundary could be task-induced. For that reason, some researchers (Chen, 2003; Falé & Faria, 2006 for intonation) propose a Reaction Time approach to test the hypothetical discreteness of a contrast. According to Chen (2003), if the categories that result from the identification task are not task-induced, it is expected that the subjects will need approximately the same time to identify the stimuli that belong to the same category, while subjects will require more time to identify the stimuli that are in the crossover region between categories, the across-category stimuli. Thus, the within-category stimuli will be less demanding than the across-category stimuli in terms of cognitive load. Figure 5 plots the mean of RT measurements of the peak height continua created from the yes-no question (black bars) and what-question (grey bars) stimuli. As in Chen (2003), the results show that listeners are faster at identifying stimuli within categories and slower at identifying stimuli across categories. Observe that while there is a peak in RT measurements corresponding to the stimulus 3 for the continuum created from the what-question base stimulus, in the continuum created from the yes-no question base stimulus RT measurements of stimuli 2 and 3 are balanced, so we find a sort of plateau instead of a clear peak. This agrees with the results from the identification task, where it was found that the category boundary would be located between stimuli 2 and 3 in yes-no-question-based stimuli but specifically at stimulus 3

<sup>3</sup> Note that we find significant-differences between stimuli 1-2 in both continua, although the difference in identification rate between them is small and that they can be assumed to fall in one category, separate from stimulus 3.

in what-question-based stimuli. Consequently, stimuli 2 and 3 (from yes-no-question-based continuum) require the same time to be identified because they both flank the frontier region while stimulus 3 (from what-question-based continuum) require more time to be identified because it lies in the frontier region.

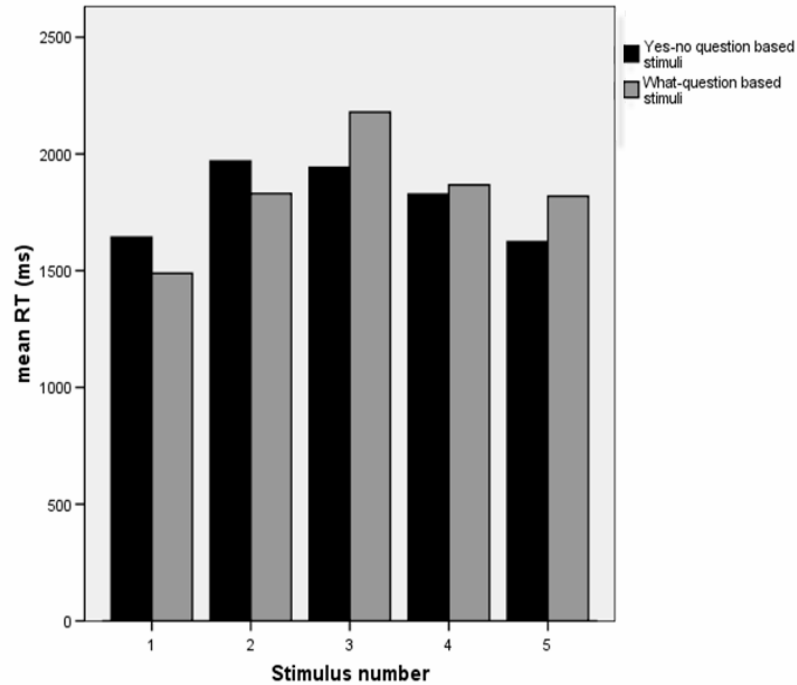


Figure 5. Mean RTs of the peak height continua created from yes-no question and what-question base stimuli.

Table V shows the results of four Wilcoxon matched pairs signed rank test comparing mean RTs between adjacent stimuli for yes-no-question- and what-question-based continua. As we can see, there are no differences between mean RTs of the stimuli 2 and 3 from the yes-no-question-based continuum because they both are located at the sides of the frontier region. By contrast, we find significant differences between stimuli 2 and 3 in the what-question-based continuum, since stimulus 3 is the across-category stimulus.

		RT 1-2	RT 2-3	RT 3-4	RT 4-5
Yes-no-question-based stimuli	z-value	-4.377	-1.251	-1.331	-1.913
	p-value	<0.004	NS	NS	NS
What-question-based stimuli	z-value	-4.313	-2.658	-2.263	-1.805
	p-value	<0.004	0.032	NS	NS

*Table V. z- and p-values of four Wilcoxon matched pairs signed rank tests comparing mean RTs between adjacent stimuli for yes-no-question- and what-question-based continua.*

### 3.2. Discrimination results

The material for the discrimination task was made up of pairs of stimuli in AB and BA order (see section on Methods) and pairs in which two stimuli were identical.

Figure 6 shows the rate of “different” responses (number of “different” responses over the total responses) to the various pairs of stimuli corresponding to the continuum created from the yes-no question base stimulus (peak at 263 Hz). AB hits are shown in light grey (“different” responses to dissimilar pairs where the second stimulus, stimulus B, had a pitch value 15 Hz higher than the first stimulus, stimulus A: 203 vs. 218 Hz, 218 vs. 233 Hz, 233 vs. 248 Hz, 248 vs. 263 Hz), BA hits in dark grey (“different” responses to dissimilar pairs where the second stimulus, stimulus A, has a pitch value 15 Hz lower than the first stimulus, stimulus B: 218 vs. 203 Hz, 233 vs. 218 Hz, 248 vs. 233 Hz, 263 vs. 248 Hz) and false alarms in black (“different” responses to identical pairs: 203 vs. 203 Hz, 218 vs. 218 Hz, 233 vs. 233 Hz, 248 vs. 248 Hz, 263 vs. 263 Hz). The actual “different” responses rates are indicated by the intermediate points between two adjacent stimuli. AB order should be interpreted from left to right, whereas BA order should be interpreted from right to left.

We find two discrimination peaks at the pairs 218 vs. 233 Hz (AB pair) and 233 vs. 218 Hz (BA pair). The identification results suggested that this frequency range did indeed represent the crossover between the categories. The most striking feature of these results has to do with the BA hits. Note that the discrimination peak in AB pairs in which the second stimulus has a higher peak than the first has a higher rate of “different” responses. These results suggest that listeners have more trouble discriminating between pairs of stimuli presented in BA order (the pairs in which the second stimulus has a lower peak than the first).



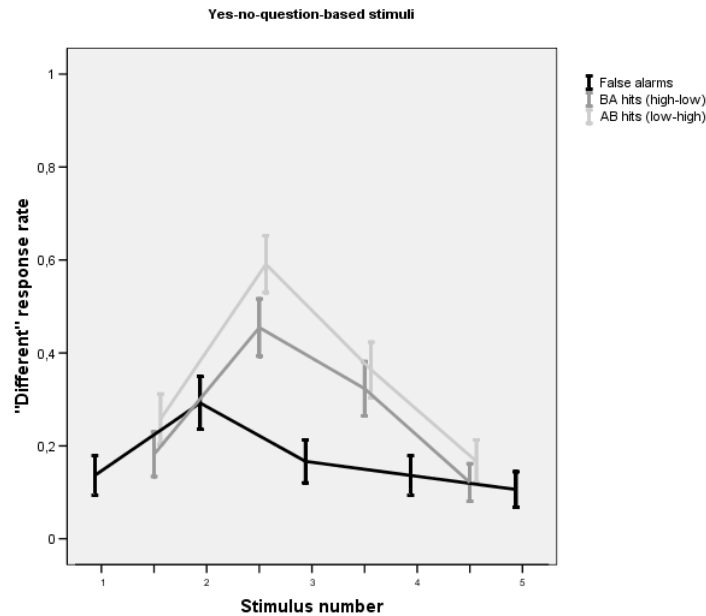


Figure 6. Rate of “different” responses for pairs that were actually different (hits) and pairs that were identical (false alarms) corresponding to the continuum created from the yes-no question base stimulus. Error bars represent standard error of the mean.

Table VI shows the results of three Wilcoxon matched pairs signed rank tests for each order of presentation comparing the response rate between two different conditions: pair 1\_2 vs. pair 2\_3, pair 2\_3 vs. pair 3\_4, and so on. Table VII shows the results of four Wilcoxon matched pairs signed rank tests comparing the response rate between the conditions AB vs. BA for each pair of stimuli. From the results of table VI we can infer that the most important discrimination peak is in the AB function because pair 2\_3 (218-233 Hz), where the discrimination peak is located, is significantly different from both pair 1\_2 and pair 3\_4. Although we observe differences in discrimination with regards to order of presentation (Figure 6), these differences are not statistically significant (Table VII).

Results of the discrimination task using stimuli created from the what-question base stimulus (peak at 203 Hz) are plotted in Figure 7 as the rate of AB hits in light grey (203 vs. 218 Hz, 218 vs. 233 Hz, 233 vs. 248 Hz, 248 vs. 263 Hz), BA hits in dark grey (218 vs. 203 Hz, 233 vs. 218 Hz, 248 vs. 233 Hz, 263 vs. 248 Hz) and false alarms in black (203 vs. 203 Hz, 218 vs. 218 Hz, 233 vs. 233 Hz, 248 vs. 248 Hz, 263 vs. 263 Hz).

AB order		1_2 vs. 2_3	2_3 vs. 3_4	3_4 vs. 4_5
Yes-no-question-based stimuli	z-value	-3.773	-3.000	-2.711
	p-value	<0.003	0.009	0.021

BA order		2_1 vs. 3_2	3_2 vs. 4_3	4_3 vs. 5_4
Yes-no-question-based stimuli	z-value	-3.411	-1.257	-2.449
	p-value	0.003	NS	0.042

Table VI. *z- and p-values of three Wilcoxon matched pairs signed rank tests for each order of presentation comparing the “different” response rate between adjacent pairs for yes-no-question-based stimuli.*

		1_2 vs. 2_1	2_3 vs. 3_2	3_4 vs. 4_3	4_5 vs. 5_4
Yes-no-question-based stimuli	z-value	-1.091	-2.043	-0.894	-0.728
	p-value	NS	NS	NS	NS

Table VII. *z- and p-values of four Wilcoxon matched pairs signed rank tests comparing the “different” response rate between two conditions (AB order and BA order) for each pair of yes-no-question-based stimuli.*

A major discrimination peak can be seen at AB pair 218 vs. 233 Hz (the rate of “different” responses reaches nearly 0.7), which agrees with the results of the identification task (we found the shift between categories around 233 Hz). Note that there is hardly any difference between the rate for BA hits and AA false alarms. These results confirm the findings shown in Figure 6, that is, it appears that subjects have trouble discriminating between stimuli when the direction of change in frequency is downwards.

Table VIII shows the results of three Wilcoxon matched pairs signed rank tests for each order of presentation comparing the response rate between two different conditions: pair 1\_2 vs. pair 2\_3, pair 2\_3 vs. 3\_4, and so on. Table IX shows the results of four Wilcoxon matched pairs signed rank tests comparing the response rate between conditions AB vs. BA for each pair of stimuli. The discrimination peak is located again at the 218 vs. 233 Hz interval in AB order. This can be seen in the results shown in Table VIII, where the differences between pair 2\_3 and adjacent pairs in the AB order are significant. By contrast, differences between pair 3\_2 and adjacent pairs in BA order are not significant. Effects of order of presentation are confirmed from the results shown in Table IX, where there are significant differences between pairs 2\_3 vs. 3\_2 and 3\_4 vs. 4\_3.

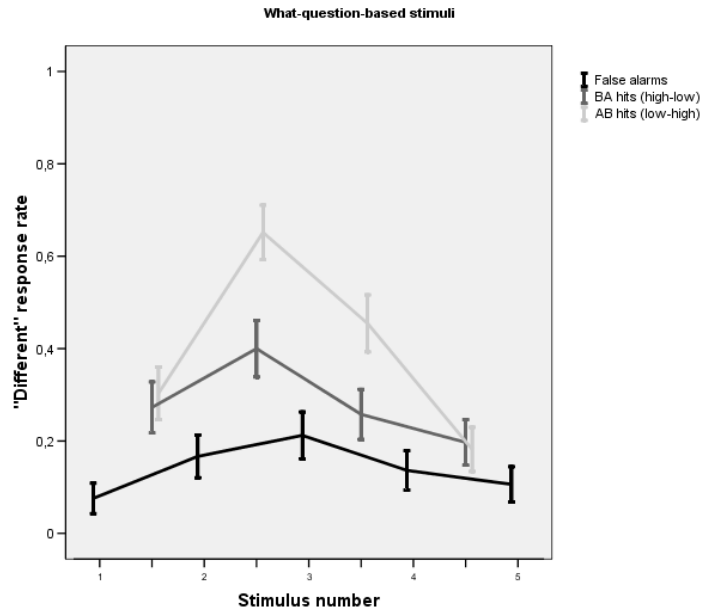


Figure 7. Rate of “different” responses for pairs that were actually different (hits) and pairs that were identical (false alarms) corresponding to the continuum created from the what-question base stimulus. Error bars represent standard error of the mean.

AB order		1_2 vs. 2_3	2_3 vs. 3_4	3_4 vs. 4_5
What-question-based stimuli	z-value	-4.226	-2.535	-3.024
	p-value	<0.003	0.033	0.006

BA order		2_1 vs. 3_2	3_2 vs. 4_3	4_3 vs. 5_4
What-question-based stimuli	z-value	-1.461	-2.043	-0.655
	p-value	NS	NS	NS

Table VIII. z- and p-values of three Wilcoxon matched pairs signed rank tests for each order of presentation comparing the “different” response rate between adjacent pairs for what-question-based stimuli.

This difference with respect to the significance of the effects of order of presentation could be related to the nature of the original stimuli. When listeners hear the accented interrogative particle in the what-question-based

stimuli, subjects expect a high tone on the syllable preceding the one with the nuclear accent. In the order AB, in the second token of the pair the pitch in this pretonic syllable is even higher in frequency than expected, so this difference is very noticeable. By contrast, after the unaccented interrogative particle in the yes-no-question-based stimuli, listeners would expect the super-high variant of the high tone in the syllable preceding the nuclear accent, but only the second token has a pitch level on that syllable that could be close to what is expected. Perhaps this difference would not be so noticeable because the pitch height of the prenuclear syllable of the first token is contradictory with the absence of accent in the interrogative particle. This could explain not only why we get significant effects of order of presentation only for the what-question-based continuum, but also why we get generally more sharply differentiated discrimination results in the what-question-based stimuli.

		1_2 vs. 2_1	2_3 vs. 3_2	3_4 vs. 4_3	4_5 vs. 5_4
What-question-based stimuli	z-value	-0.200	-2.959	-2.746	-0.258
	p-value	NS	0.012	0.024	NS

*Table IX. z- and p-values of four Wilcoxon matched pairs signed rank tests comparing the “different” response rate between two conditions (AB order and BA order) for each pair of what-question-based stimuli.*

### 3.2.1. *Models of discrimination performance*

#### 3.2.1.1. *Relating discrimination responses to identification responses*

We see in the previous section that the continua tested in this study fulfill one of the criteria which according to Francis et al. (2003) that identification and discrimination results should obey in order to consider that the contrast tested is categorical, that is, the peaks of discrimination correspond to the location of the category boundary as determined by identification. The other criterion is that identification results should predict discrimination accuracy (Liberman et al., 1957). It can be said, hence, that the extent to which discrimination performance can be predicted from classification is what is referred to as categorical perception. Thus, in order to determine whether discrimination performance can be predicted by identification results, two formulas for predicting discrimination were applied. The first formula was taken from Liberman et al. (1957), who used it to predict the results of an ABX discrimination task. However, Pollock & Pisoni (1971) showed that the same equation can also be used to predict performance in a same/different discrimination task. The equation is:

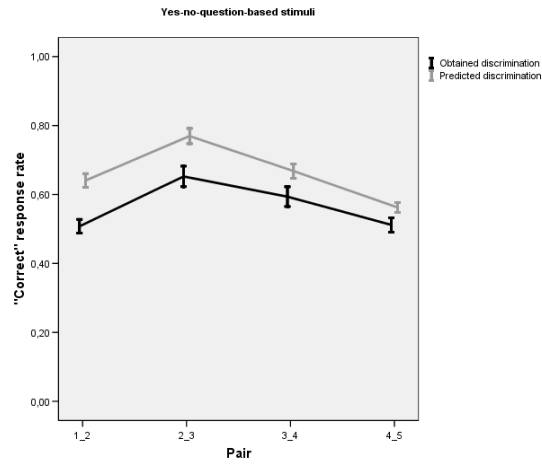


Figure 8. Predicted and obtained discrimination rates for the continuum created from the yes-no question base stimulus. Error bars represent the standard error of the mean.

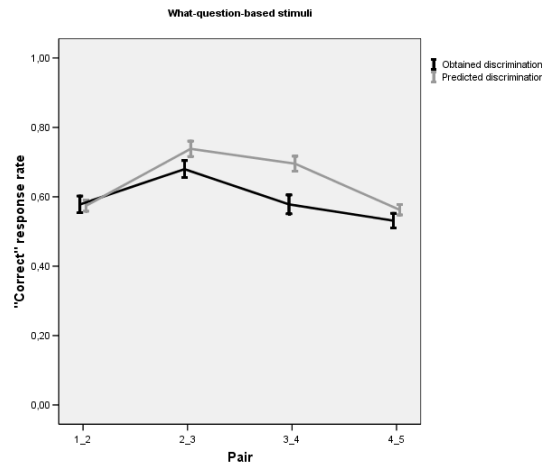


Figure 9. Predicted and obtained discrimination rates for the continuum created from the what-question base stimulus. Error bars represent the standard error of the mean.

$$(1) \quad P(\text{disc}_{12}) = 0.5[1 + (p_1 - p_2)^2]$$

$P_1$  is the probability of identifying Stimulus 1 as Category A and  $p_2$  is the probability of identifying stimulus 2 as category A. This formula assumes that when listeners do not hear a difference they respond “same” or “different”

randomly, so that performance is by chance. As pointed out by Macmillan et al. (1977): “If the resulting classification led to a decision (i.e., if A and B were classified differently and X as one of them –in an ABX discrimination task), the observer would respond as indicated; if it did not lead to a decision, he would guess, choosing each response with probability 0.5”.

The second formula is taken from Godfrey et al. (1981) and is a more general formula that predicts discrimination on the basis of phonetic categorization without guessing probabilities. This formula was also used by other studies of children’s categorical perception such as Wolf (1973) and Brandt & Rosen (1980) that employed same/different discrimination tasks:

$$(2) \quad \text{Proportion discriminated} = (P_{1a} \times P_{2b}) + (P_{1b} \times P_{2a}), \text{ where}$$

$P_{1a}$  = proportion of time that stimulus 1 was identified as “a”,  
 $P_{2b}$  = proportion of time that stimulus 2 was identified as “b”,  
 $P_{1b}$  = proportion of time that stimulus 1 was identified as “b”,  
 $P_{2a}$  = proportion of time that stimulus 2 was identified as “a”.

Figures 8 and 9 show the obtained (black) and predicted (grey) discrimination functions as a result of the application of formula (1). The rate of correct discrimination for each pair was calculated as in Francis & Ciocca (2003), which uses the same kind of task as in this study, that is, as the average of the rate of “different” responses for different pairs and the rate of “same” responses for same pairs. For example, the rate of “correct” responses for the 1-2 pair was the average of the rate of “different” responses for the 1-2 and 2-1 pairs and the rate of “same” responses for the 1-1 and 2-2 pairs.

		Obtained 1_2 vs. predicted 1_2	Obtained 2_3 vs. predicted 2_3	Obtained 3_4 vs. predicted 3_4	Obtained 4_5 vs. predicted 4_5
Yes-no- question- based stimuli	z-value	-2.905	-3.148	-1.766	-1.526
	p-value	0.016	NS	NS	NS
What- question- based stimuli	z-value	-0.0251	-0.971	-2.999	-1.360
	p-value	NS	NS	0.012	NS

*Table X. z- and p-values of four Wilcoxon matched pairs signed rank tests for each continuum comparing the obtained and predicted discrimination rate for each pair of stimuli.*

Four Wilcoxon matched pairs signed rank tests were carried out for each continuum in order to compare the obtained and predicted discrimination rates

(Table X) for each pair of stimuli. In order to claim that the formula accurately predicts the discrimination function, predicted and obtained discrimination functions cannot be significantly different. The results show that the difference between predicted versus obtained discrimination is significant for pair 1\_2 with stimuli created from the yes-no question base stimulus and for pair 3\_4 with stimuli created from the what-question base. Thus, it was concluded that discrimination results cannot be predicted accurately from the identification results if we assume that listeners respond randomly when they do not hear a difference.

Figures 10 and 11 show the obtained (in black) and predicted (in grey) discrimination functions as a result of the application of formula (2). The real discrimination values have been calculated as the rate of different pairs which were correctly called “different”, as in Godfrey et al. (1981), who use this formula with the same kind of task as in this study.

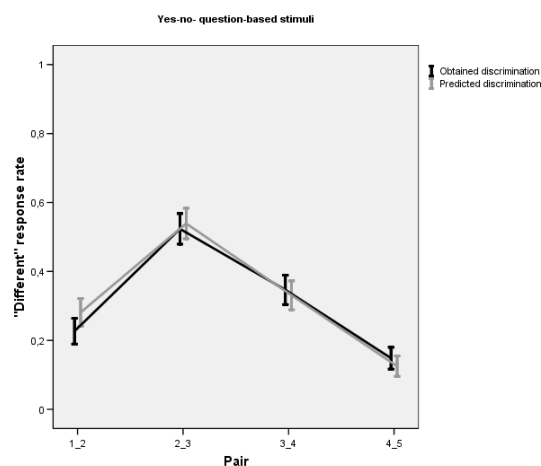


Figure 10. Predicted and obtained discrimination rates for the continuum created from the yes-no question base stimulus. Error bars represent the standard error of the mean.

The results of four Wilcoxon matched pairs signed rank tests (Table XI) for each continuum show that the differences between predicted versus obtained discrimination is significant only for pair 1 with stimuli created from the what-question base stimulus. This means that formula (2) is suitable for our data and that discrimination data can be predicted from identification data only on the basis of phonetic categorization, without making assumptions about guessing.

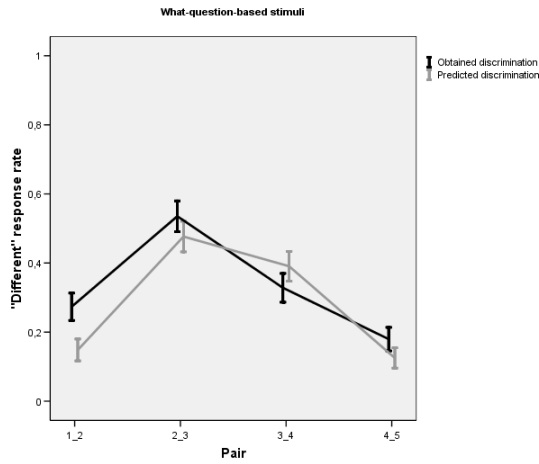


Figure 11. Predicted and obtained discrimination rates of the continuum created from the what-question base stimulus. Error bars represent the standard error of the mean.

		Obtained 1_2 vs. predicted 1_2	Obtained 2_3 vs. predicted 2_3	Obtained 3_4 vs. predicted 3_4	Obtained 4_5 vs. predicted 4_5
Yes-no- question- based stimuli	z-value	-1.021	-0.258	-0.258	-0.539
	p-value	NS	NS	NS	NS
What- question- based stimuli	z-value	-2.596	-0.927	-1.000	-1.219
	p-value	0.036	NS	NS	NS

Table XI. *z-* and *p-values* of four Wilcoxon matched pairs signed rank tests for each continuum comparing the obtained and predicted discrimination rate for each pair of stimuli.

### 3.4.2. Signal Detection Theory

The model of discrimination performance described above by formula (1) presupposes that when listeners do not hear a difference or do not know how to respond, they respond “same” or “different” at random. But it is by no means certain that listeners act in this way. According to Keating (2004), some subjects may tend to give a “different” response most of the time while, on the other hand, other subjects may be very conservative and only give a



“different” response when they are completely sure that they hear a difference. This means that in the former case the results for same pairs are not reliable and in the latter case the same will be true for different pairs. The point is that the percentage of correct discriminations between different pairs is highly susceptible to subjects who tend to give only “different” (or “same”) responses all the time, and it should be interpreted in terms of the listener’s response bias, that is, his or her tendency to qualify stimuli pairs as “same” or “different”.

Signal Detection Theory attributes responses to a combination of sensitivity and bias. Sensitivity is the variable that is being investigated and bias is what we must take into account so that the sensitivity measure is meaningful. The statistical expression  $d'$  ( $d$  prime) is a measure of the difference between the hit rate (proportion of different pairs to which subjects responded “different” and false alarm rate (proportion of same pairs to which subjects responded “same”). However,  $d'$  is not just the difference between the hit rate and the false alarm rate, rather, it is defined in terms of  $z$ , the inverse of the normal distribution functions, as shown in (3).

$$(3) \quad d' = z(H) - z(F)$$

$D'$  has been used in the discrimination literature for obtained and predicted discrimination by Best et al. (1981) and in addition to percentage of correct / percentage of different responses to different pairs by Francis & Ciocca (2003). In order to validate the results plotted in Figures 6 and 7 related to the presence of order of presentation effects, Signal Detection Theory was applied to our data. Figures 12 and 13 show the discrimination results presented as  $d'$  for each stimulus pair in low-high order (black lines) and high-low order (grey lines).  $D'$  scores were calculated on the basis of “different” responses to the pairs that were truly different (hits) and “different” responses to the pairs that were actually the same (false alarms). Following Macmillan & Creelman (1991),  $d'$  was calculated using roving methods<sup>4</sup> (using Table A5.4, pp. 338-354). This table was generated by varying the response threshold (the value of  $k$ ) in equation (4):

---

<sup>4</sup> According to Macmillan & Creelman (1991), there are two different strategies used by subjects in responding to “same-different” experiments and therefore different appropriate models for  $d'$  depending on how stimuli are across trials in a block. In roving designs, the two stimuli vary from trial to trial and “subjects are likely to apply a differencing strategy, applying a threshold of difference to decide if 2 stimuli are different enough to count as ‘different’” (Keating, 2004).

(4)

$$H = \left[ \frac{(-k + d')}{\sqrt{2}} \right] + \left[ \frac{(-k - d')}{\sqrt{2}} \right]$$

$$F = 2\Phi(-k\sqrt{2})$$

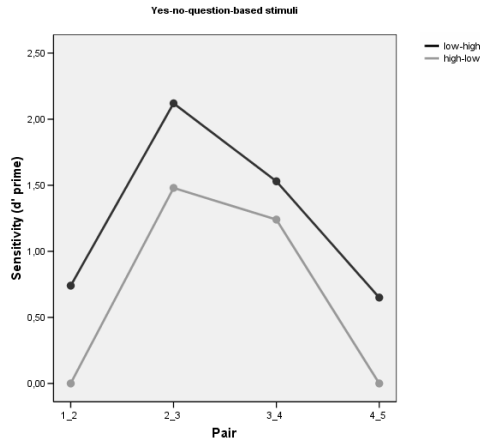


Figure 12.  $d'$  calculated for both orders of presentation of pairs from the continuum created from the yes-no question base stimulus.

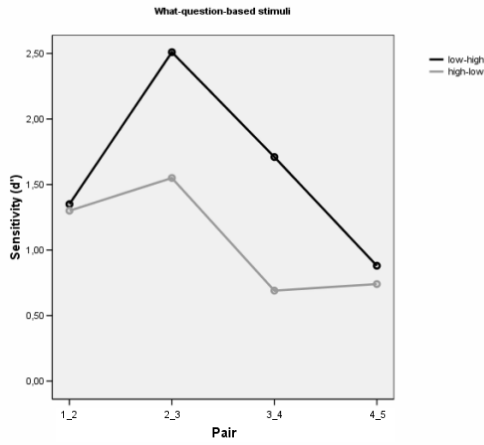


Figure 13.  $d'$  calculated for both orders of presentation of pairs from the continuum created from the what- question base stimulus.

Figures 12 and 13 show that Majorcan listeners are more sensitive to F0 differences in speech stimuli when the second stimulus in a pair has higher F0. Observe that this difference in sensitivity is especially important in the pair that represents the boundary between the categories and which corresponds to the discrimination peak, pair 2\_3.

Four Wilcoxon matched pairs signed rank tests (Table XII) were carried out on hit rate minus false alarm rate for each pair with the conditions AB and BA orders of presentation. Significant differences were found only for the second and third pairs for stimuli created from the what-question base stimulus. For stimuli created from the yes-no question base stimulus, the differences in order of presentation were significant for none of the pairs. Thus, the application of  $d'$  in addition to the rate of “different” responses to different/same pairs has confirmed the presence of order of presentation effects which are statistically significant only for pair 2\_3 vs. pair 3\_2 and pair 3\_4 vs. pair 4\_3 for stimuli created from the what-question base stimulus.

		Pair 1_2 vs. pair 2_1	Pair 2_3 vs. pair 3_2	Pair 3_4 vs. pair 4_3	Pair 4_5 vs. pair 5_4
Yes-no- question- based stimuli	z-value	-1.213	-1.851	-0.832	-0.688
	p-value	NS	NS	NS	NS
What- question- based stimuli	z-value	-0.098	-2.265	-2.688	-0.225
	p-value	NS	0.032	0.028	NS

Table XII.  $z$ - and  $p$ -values of four Wilcoxon matched pairs signed rank tests comparing the difference between hit rate and false alarm rate for each pair with the conditions AB and BA orders of presentation.

#### 4. Discussion

The present study has provided evidence that Majorcan Catalan listeners make categorical linguistic use of F0 scaling differences in perceiving yes-no questions as opposed to what-questions in Majorcan Catalan. This evidence comes from different sets of results. The identification results show that it is possible to switch the perceived category by manipulating the pitch height of the leading tone from an H tone to a super-high tone and vice versa. We observed in Figure 4 that the presence/absence of the accented what-particle in the two continua does not interfere in the categorical perception of this contrast, that is, the identification functions appear unmistakably S-shaped with an identification rate that goes from 0.85 to about 0.2 (in the case of the continuum created from the what-question base stimulus) and from less than

0.27 to about 0.87 (in the case of the continuum created from the yes-no question base stimulus) within 2 steps of the 5-step continuum.

Evidence of this linguistic contrast is also provided by RT measurements. A mean RT peak/plateau can be observed in Figure 5 at the identification boundary for both continua; hence, the mean RTs were shorter for within categories and longer for across categories. According to Chen (2003), these are essential properties of linguistically real categories and not task-induced.

Moreover, by comparing through statistical analyses the magnitude of difference in the identification rate for adjacent stimuli in the identification task, it has been observed that, although there is a shift from one response to the other in the range of values between 218 and 248 Hz, the threshold between the “yes-no question” and the “what-question” category would be located between 218 and 233 Hz. This is actually the pair of stimulus that is best discriminated and, hence, where the discrimination peak is situated. However, for classic CP it is not enough that an abrupt shift in the identification function is observed and that this shift in the identification function corresponds to the discrimination peak; rather, discrimination results should also be predicted from identification results. As it turns out, the application of the Haskins formula (the formula taken from Liberman et al., 1957) and the more general formula taken from Godfrey et al. (1981) do indeed permit us to predict the shape of the discrimination curve. From Figures 8, 9, 10 and 11 and from statistical analyses, it can be said that the formula used by Godfrey et al. (1981) based simply on phonetic categorization, without assumptions about guessing, is the one that best fits our data. Because the results of the Wilcoxon matched pairs signed rank tests show that the differences between predicted and obtained discrimination are not significant, it is concluded that this contrast is categorically perceived.

On the basis of this evidence, we claim that the difference in pitch height of the leading tone of the nuclear accent is a strong perceptual cue that Majorcan listeners use when distinguishing yes-no questions from what-questions. This does not mean that pitch height is the only cue; on the contrary, identification results seem to suggest that there is a supplementary cue, the accented interrogative particle. The effect of the accented interrogative particle is especially noticeable in the what-question-based continuum, in which it appears to delay the switch from one category to the other. This effect can be observed not only in our identification results but also in the results of RT measurements, in which the difference in the crossover boundary between the two kinds of stimuli can be seen in the presence of a mean RT plateau in the yes-no-question-based continuum on the one hand and a mean RT peak in the what-question-based continuum on the other. Thus, it would be of great interest to know how listeners would respond to the continua if the effect of accent were neutralized. It is expected that we would obtain results similar to the identification performance of qualified musicians (see page 18). In any case, this indicates the need for further

research to confirm that the effect of the interrogative particle is exactly the effect suggested by these results.

Finally, it is worth trying to account for the discrimination asymmetries that our results report (it is easier to perceive differences between two stimuli when the second one has a higher pitch than the first). These asymmetries have been related in the literature to the F0 declination or downdrift, the gradual declination of fundamental frequency over the course of an utterance (Pierrehumbert, 1979; Gussenhoven & Rietveld, 1988). F0 declination has been argued to be a universal characteristic of speech production. Evidence for compensation of this declination effect has been provided for American English listeners (Pierrehumbert, 1979), Dutch listeners (Gussenhoven & Rietveld, 1988) and Cantonese listeners (Wong, 1999). According to Francis & Ciocca (2003), these asymmetries may be explained in terms of a compensation for an expected declination in F0 over the course of an utterance. Thus, listeners are able to compensate for this decline by taking into account the position of the accent within the utterance so that the meaning conveyed by the speaker is correctly identified. Given two tokens, when the second token has lower pitch than the first, this compensation would ensure that the two tokens sound identical; by contrast, when the second token has a higher pitch than the second, this raising in pitch of the second token would enhance the perception of the difference between the two tokens. Notice that for our case a putative effect of declination would have to be interpreted as applying across utterances, rather than within utterances. However, further research is necessary to test whether declination exists in Majorcan Catalan or whether Majorcan Catalan listeners compensate perceptually for this expected declination.

## **5. Conclusions**

The results of this perception study confirm the categorical perception of the difference in pitch height between H and upstepped H within the H+L\* nuclear accent and consequently the phonological role of scaling of the H leading tone, since it permits listeners to distinguish yes-no questions from what-questions in Majorcan Catalan.

The results also indicate that there are discrimination asymmetries that depend on the direction of change, it being easier to distinguish between the stimuli pair when the direction of change is upwards.

## **Acknowledgments**

I am grateful to Pilar Prieto for guiding my work, to Francesc Felipe for help with computer and technical issues, to the subjects for participating in this experiment, to Aaju Chen, Mariapaola D'Imperio, José Ignacio Hualde and Marta Ortega for their helpful advice, to Laboratoire Parole et Langage for

PERCEVAL software, to Caterina Petrone and Mariapaola D'Imperio for help with PERCEVAL, to participants at the Phonetics and Phonology in Iberia (PaPI 2005) preview and conference and the Speech Prosody 2006 conference for comments and suggestions, to the *Servei d'Estadística de la Universitat Autònoma de Barcelona* and especially to Anna Espinal. Finally, I would like to thank Gorka Elordieta and two anonymous reviewers for their useful comments.

## References

- Best, C.T., Morrongiello, B., & Robson, R. (1981) Perceptual equivalence of acoustic cues in speech and nonspeech perception, *Perception & Psychophysics*, **29**, 191-211.
- Blalock, H. (1979) *Social statistics*. New York: McGraw-Hill.
- Brandt, J., & Rosen, J.J. (1980) Auditory phonemic perception in dyslexia: categorical identification and discrimination of stop consonants, *Brain and Language*, **9**, 324-337.
- Calhoun, S. (2003) The nature of theme and rheme accents. In *Proceedings of One-Day Meeting for Young Speech Researchers*. University College, London.
- Chen, Aojun. (2003) Reaction time as an indicator of discrete intonational contrasts in English. In *Proceedings of Eurospeech 2003*, 97-100. Geneva: ISCA Archive.
- Cummins, F., Doherty, C., & Dilley, L. (2006) Phrase-final pitch discrimination in English. In *Proceedings of Speech Prosody* (Hoffmann, R. & H. Mixdorff, editors), pp. 467-470. Dresden: TUDpress Verlag der Wissenschaften GmbH.
- D'Imperio, M., & House, D. (1997) Perception of questions and statements in Neapolitan Italian. In *Proceedings of Eurospeech 1997* (G. Kokkinakis, N. Fakotakis & E. Dermatas, editors), pp. 251-254. Rhodes, Greece: ISCA Archive.
- Face, T.L. (2005) F0 peak height and the perception of sentence type in Castilian Spanish, *Revista Internacional de Lingüística Iberoamericana (RILI)*, **III**, 49-66.
- Falé, I., & Hub Faria, I. (2006) Categorical perception of intonational contrasts in European Portuguese. In *Proceedings of Speech Prosody* (R. Hoffmann & H. Mixdorff, editors), pp. 69-72. Dresden: TUDpress Verlag der Wissenschaften GmbH.
- Finney, D.J. (1971) *Probit analysis*. Cambridge: Cambridge University Press.
- Francis, A., & Ciocca, V. (2003) Stimulus presentation order and the perception of lexical tones in Cantonese, *Journal of Acoustical Society of America*, **114**, 1611-1621.
- Francis, A., Ciocca, V., & Kei Chit Ng, B. (2003) On the (non)categorical perception of lexical tones, *Perception & Psychophysics*, **65**, 1029-1044.
- Glenn Schellenberg, E. (2002) Asymmetries in the discrimination of musical intervals: going out-of-tune is more noticeable than going in-tune, *Music Perception*, **19**, 223-248.
- Godfrey, John J., K., Syrdal-Laskym Ann, Millay, Kathleen K., & Knox, Carol M. (1981) Performance of dyslexic children on speech perception tests, *Journal of Experimental Child Psychology*, **32**, 401-424.

- Gussenhoven, C., & Rietveld, A. (1988) Fundamental frequency declination in Dutch: Testing three hypothesis, *Journal of Phonetics*, **16**, 355-369.
- Gussenhoven, C. (2004) *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Jensen, Marvin D., & Carlin, Phyllis Scott. (1981) Sex differences in the accuracy of intonation perception: a study of oral interpretation students, *Missouri Speech Journal*, **12**, 17-21.
- Johnson, K., Strand, E. A., & D'Imperio, M. (1999) Auditory-visual integration of talker gender in vowel perception, *Journal of Phonetics*, **27**, 359-384.
- Keating, P. (2004) Statistics (<http://www.linguistics.ucla.edu/faciliti/facilities/statistics/statistics.html>).
- Kohler, K. J. (1987) Categorical pitch perception. In *Proceedings of 11th International Congress of Phonetic Sciences* (U.E. Viks, editor), pp. 331-333. Tallinn: Academy of Sciences of the Estonian S.S.R.
- Ladd, D.R., & Morton, R. (1997) The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics*, **25**, 313-342.
- Liberman, A., Harris, K., Hoffman, H., & Griffith, B. (1957) The discrimination of speech sounds within and across phoneme boundaries, *Journal of Experimental Psychology*, **54**, 358-368.
- Macmillan, N.A., & Creelman, C. Douglas. (1991) *Detection theory: a user's guide*. Cambridge: Cambridge University Press.
- Macmillan, Neil A., Kaplan, Howard L., & Creelman, C. Douglas. (1977) The psychophysics of categorical perception, *Psychological Review*, **84**, 452-471.
- Payà, M., & Vanrell, M.M. (2005) Yes-no questions and echo-questions intonation in Majorcan and Minorcan Catalan. Paper presented at *Phonetics and Phonology in Iberia*. Barcelona: 20-21 June.
- Pierrehumbert, J.B. (1979) The perception of fundamental frequency declination, *Journal of the Acoustic Society of America*, **66**, 363-369.
- Pierrehumbert, J.B. (1980) *The Phonetics and phonology of English intonation*. New York: Garland Press.
- Pierrehumbert, J.B., & Beckman, M.E. (1988) *Japanese tone structure*. Cambridge: MIT Press.
- Pisoni, D.B., & Tash, J. (1974) Reaction times to comparisons within and across phonetic categories, *Perception and Psychophysics*, **15**, 285-290.
- Pollock, I., & Pisoni, D.B. (1971) On the comparison between identification and discrimination tests in speech perception, *Psychonomic Science*, **24**, 299-300.
- Post, B. (2000) *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics.
- Prieto, P. (2003) Scaling of H1 peaks in Spanish: evidence from five sentence types. Paper presented at *15th International Congress of Phonetic Sciences*. Barcelona: 3-9 August.
- Remijsen, B., & van Heuven, V. (1999) Gradient and categorical pitch dimensions in Dutch: diagnostic test. In *Proceedings of 14th International Congress of Phonetic Sciences* (J.J. Ohala, Y. Hasagawa, M. Ohala, D. Granville & A.C. Bailey, editors), pp. 1865-1868. San Francisco: University of California.
- Repp, B.H., Healy, A.F., & R.G., Crowder. (1979) Categories and context in the perception of isolated steady-state vowels, *Journal of Experimental Psychology: Human Perception and Performance*, **5**, 129-145.

- Rigau, G., & Prieto, P. (2005) A typological approach to Catalan interrogative sentences headed by *que*, *Probus*, **16**, 113-150.
- Rogers, H. (2003) Male-female phonetic bibliography: <http://ccat.sas.upenn.edu/~haroldfs/bibliogs/malefeml.html>.
- Schneider, K., & Linfert, B. (2003) Categorical perception of boundary tones in German. In *Proceedings of 15th International Congress of Phonetic Sciences* (D. Recasens, M.J. Solé & J. Romero, editors), pp. 1-4. Barcelona: Causal Productions Pty Ltd.
- Vanrell, M.M. (2006) The phonological role of tonal scaling in Majorcan Catalan interrogatives. Unpublished Master's thesis, Universitat Autònoma de Barcelona (<http://seneca.uab.es/ggt/Tesis/MVanrell.pdf>).
- Wolf, C.G. (1973) The perception of stop consonants by children, *Journal of Experimental Child Psychology*, **8**, 351-361.
- Wong, P.C.M. (1999) The effect of downdrift in the production and perception of Cantonese level tone. In *Proceedings of 14th International Congress of Phonetic Sciences* (J.J. Ohala, Y. Hasagawa, M. Ohala, D. Granville & A.C. Bailey, editors), pp. 2395-2398. San Francisco: University of California.

Maria del Mar Vanrell  
UAB – Dept. Filologia Catalana  
Universitat Autònoma de Barcelona – Edifici B  
08193 Bellaterra (Barcelona) – Spain  
[MariadelMar.Vanrell@uab.es](mailto:MariadelMar.Vanrell@uab.es)