THE PHONOLOGICAL ROLE OF TONAL SCALING IN MAJORCAN CATALAN INTERROGATIVES

Maria del Mar Vanrell Bosch

Supervisor:

Dr. Pilar Prieto Vives

Treball de Recerca

Departament de Filologia Catalana

Universitat Autònoma de Barcelona

2006

TABLE OF CONTENTS

Acknowledgments

I am profoundly grateful to my supervisor Pilar Prieto for her guidance, encouragement and constant support, but especially for having provided me the possibility of working with her and learning in a very active manner. Thanks again for introducing me to such a long list of people who have all contributed to the improvement of this research and to the raising of new questions that have stimulated my desire to learn. Thus, I owe special thanks to Aoju Chen, Mariapaola D'Imperio, José Ignacio Hualde and Marta Ortega for their helpful advice and discussion of different aspects of my work, as well as for providing me with important papers and references; to Laboratoire Parole et Langage for Perceval software; to Caterina Petrone and Mariapaola D'Imperio for help and suggestions about using Perceval; to Valter Ciocca and Alexander L. Francis for help and advice about models of discrimination performance.

Thanks also to the Departament de Filologia Catalana of the UAB for making me feel welcome from the very first moment. I would like to thank the teachers of my first year of doctoral studies for initiating me in such exciting fields: Anna Bartra, Laura Bosch, Maria Teresa Espinal, Maria Lluïsa Hernanz, Manuel Pérez Saldanya, Carme Picallo, Daniel Recasens, Joana Rosselló and Jaume Solà. Thanks also to Eulàlia Bonet, Teresa Cabré, Joan Mascaró and Gemma Rigau who beyond a simple professional relationship, have provided me confidence and support in my first steps in research and teaching.

I am very grateful to the UAB's Statistics Service and especially to Anna Espinal for the time and interest she devoted to my learning about statistics.

Part of this work was presented at the PaPI 2005 (Phonetics and Phonology in Iberia) and Speech Prosody 2006 conferences. I am grateful to the participants, as well as the participants of the PaPI preview, for their comments and suggestions.

All the subjects who participated unselfishly in the experiments as listeners and speakers deserve a special mention not only for their participation but for having replied to the "subject recruitment" e-mails and calls so fast. I would like to thank all the friends and students who were "suddenly" involved in a spiral of experiments and recordings.

I would also like to thank the people that have been close to me over the distance: my friends Xini, Magdalena, Miquel, Andreu, Jaume, Aina, Margalida, Apol·lònia, Francina, Toni, Tomeu, Victòria, Maria, Tòfol; and my sister and brother Margalida and Pep and their families. Also thanks to the people that have been close to me geographically as well as spiritually: Joan, Sebastià, Mònica, Maria Josep, Amadeu, Ramirada, Aina, Xico and Roger.

Great thanks to Francesc Felipe, my partner, for so many things: for help with computer and technical issues, for his methodological support, for his constant collaboration in the experiments—always being the first subject. But, especially, thanks for being the most important support at any time and for trusting in me when this adventure started. Finally, special thanks to my parents, Joan and Maria, for always understanding me, for being able to cut down the distance day after day, for remembering me in all they do (which is not always such an easy task).

LIST OF FIGURES

LIST OF TABLES

1. Introduction

1.1. Yes-no questions and wh-questions in Majorcan Catalan

In Majorcan Catalan[1], yes-no questions and wh-questions are characterized by a falling nuclear accent H+L*, that is, an H leading tone aligned with the pretonic nuclear syllable and a L* tone associated with the last stressed syllable. Typically, yes-no questions are headed by the unaccented interrogative particle *que* ('that') (Figure 1, left panel), and this can be compared with wh-questions headed by the accented wh-particle *què* ('what') (Figure 1, right panel). Results from a Map-Task recording (Payà and Vanrell, 2005), Vanrell (2003) and my own speech suggested that in yes-no questions the leading tone H is realized in a higher tone (Figure 1, left panel). As can be seen in Figure 1, the difference in tonal height of the leading tone H 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, wh-questions show a steady high tone which extends from the beginning of the sentence until the last stressed syllable with the falling pitch movement aligned with the last stressed syllable in the utterance. By contrast, yes-no questions have a well-defined slope that goes from the syllable that precedes the pretonic nuclear syllable realized in a high tone to the super-high tone of the pretonic nuclear syllable, where it is then followed by a steeper slope that goes down until the end of the nuclear syllable. In both types of sentences, we observe low boundary tones.

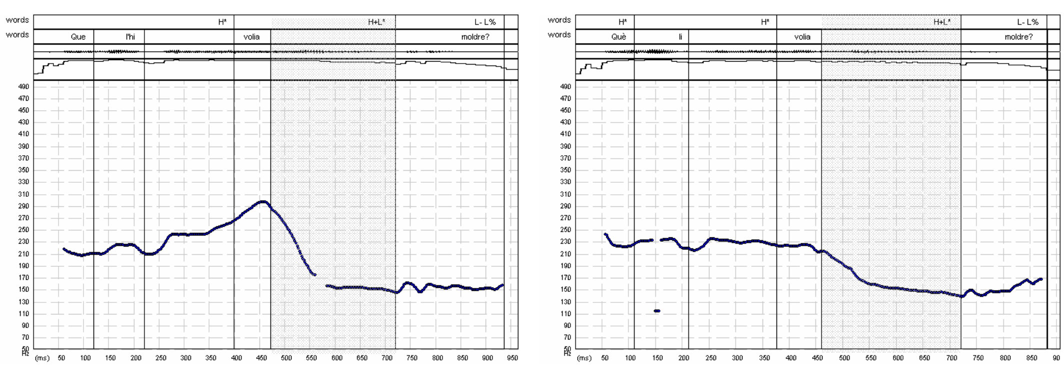


Figure 1. Waveforms and fundamental frequency contours of yes-no question *Que l'hi volia moldre?* [kə li vuˌliə ˈmɔl̪drə] ('Did s/he want to grind it for him/her?'), left panel, and wh-question *Què li volia moldre?* ('What did s/he want to grind for him/her?') [ˌkə li vuˌliə ˈmɔl̪drə], right panel.

This difference in pitch height of the leading tone H in the nuclear accent for yes-no questions has been described not only for Majorcan Catalan but also for other varieties of Catalan like

[1] Majorcan Catalan is a dialect of Catalan spoken in Majorca, the largest island of the Balearic Islands, with a population of 750,000 inhabitants. Majorcan Catalan is a member of the Eastern Catalan dialect group, together with the other Balearic subdialects (Minorcan, Ibizan and Formenteran Catalan), Central Catalan, Rossellonese and Alguerese.

Minorcan Catalan (Mascaró i Pons, 1986) and Eivissan Catalan (both subdialects of Balearic Catalan) (Figure 2), as well as Tarragonese Catalan, a subdialect of Central Catalan (Bonet i Lloret, 1988; Prieto, 2001). In Minorcan Catalan, Eivissan Catalan and Tarragonese Catalan, this tonal characteristic usually goes hand in hand with the unaccented interrogative particle *que.* According to Moll (1971) this interrogative particle *que* comes from the phrase *¿és que…?* (as in French *est-ce que…?*).[2] According to Mascaró i Pons (1986), the particle *que* is an interrogative marker that is very common in colloquial speech, and whose existence does not provoke a conflict with the interrogative particle *què* (‘what’), since the two particles display different intonation patterns. However, only in Majorcan Catalan do the pair made up of yes-no question and wh-question represent a minimal pair. That is, the two sentences in Figure 1 are homophonic at the segmental level. This is not true in other varieties of Catalan, in which the accented interrogative particle *què* is pronounced [ˈkɛ] while the unaccented interrogative particle *que* is pronounced [kə]. Moreover, wh-questions like yes-no questions, are characterized by a falling nuclear accent H+L* in Majorcan Catalan. Thus, in Majorcan Catalan, *què* (‘what’) and *que* (‘that’) do not produce different tonal contours. This is the reason why Majorcan Catalan provides a good test case to show that pitch height is not necessarily related to paralinguistic variation but can also play a decisive role in differentiating two different utterance types, in this case, yes-no questions and wh-questions.

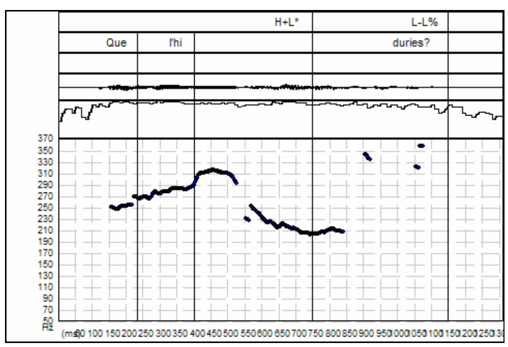


Figure 2. Waveforms and fundamental frequency contours of the yes-no question *Que l’hi duries?* [kə li ðuˈriəs] (‘Would you take it to him/her?’) in Ibizan Catalan.

Although pitch height variation has been assumed to be paralinguistic by the standard Autosegmental-Metrical (AM) model (Pierrehumbert, 1980; Pierrehumbert and Beckman, 1988), that is, it exclusively expresses differences in emphasis or prominence, some studies have shown that the difference in pitch height might also trigger categorical effects. Ladd and Morton (1997) applied the Categorical Perception paradigm (CP) to a contrast between two

[2] See Rigau and Prieto (2005) for a syntactic and prosodic description of Catalan yes-no questions headed by interrogative marker *que*.

pitch-accents in English, the normal high accent and the emphatic extra high accent. Evidence was found for a shift between these two categories in identification but a related discrimination peak was not. 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. From the results of Ladd and Morton (1997), Chen (2003) argued that the absence of peak discrimination might be related not to the inexistence of categorical perception but rather to a hypothetical unsuitability of the CP paradigm as applied to a pitch height contrast. Thus, Chen (2003) proposed an alternative method to examine the nature of two intonational contrasts between normal high accent vs. emphatic high accent and early peak alignment vs. late peak alignment, which she called the Reaction Time (RT) measurement. Noting that mean RT was shortest for within-category identification and longest for across-category identification only for the peak height continuum and not for the peak alignment contrast, and basing herself on previous studies (Pisoni and Tash, 1974), Chen concluded that this pitch height contrast is discrete. Face (2004) 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. Prieto (2003) studied the effects of sentence type on scaling variation of the first pitch accent peak in Castilian Spanish, and found evidence that pitch height is not solely related to paralinguistic variation, since sentence-type information also has a strong effect on H1 scaling in Spanish. In a production and a perception experiment, Calhoun (2003) found that themes and rhemes in English are marked by distinctive pitch accents and that the most reliable cue to the theme and rheme accents is pitch height.

1.2. Goals of the study

This study addresses the question of how differences in pitch height can be related not only to paralinguistic meaning but also to linguistic meaning. Majorcan Catalan provides a good test case to show this because it offers pairs of phrases that are homophonic at the segmental level and in which prosodic events play an important role in disambiguating them. The first goal of the study is to investigate the nature of the falling nuclear accent (H+L*) that characterizes both yes-no questions and wh-questions in Majorcan Catalan as well as the acoustic properties of the two interrogative particles *que* (‘that’) and *què* (‘what’) through a production study.

A second goal of the 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 the pretonic nuclear syllable in perceiving yes-no questions as opposed to wh-questions in Majorcan Catalan. The prediction is that if listeners' perception of differences in pitch height is categorical, that will mean that pitch scaling has a phonological character in Majorcan Catalan interrogatives.

Finally, since asymmetries have been reported repeatedly in the literature in the application of CP to intonational contrasts, a third goal of the study is to test if such asymmetries are present. It is predicted that these effects occur such that it will be easier to discriminate between a pair of stimuli when the direction of change is upwards than when it is downwards (Ladd and Morton, 1997; Remijsen and van Heuven, 1999; Falé and Faria, 2006; Cummins et al. (2006).

1.3. Methodology

1.3.1. The Categorical Perception (CP) paradigm

Different experimental approaches have addressed the question of whether a tonal difference has phonological categorical effects. One of the most common methods to examine the nature of tonal contrasts has been the Categorical Perception paradigm (CP). The CP paradigm has been applied to differences in peak alignment (Kohler, 1987; D'Imperio and House, 1997; Chen, 2003) and to differences in pitch height in both tonal languages (Francis et al., 2003; Francis and Ciocca, 2003) and intonational languages, for boundary tones (Remijsen and van Heuven, 1999; Post, 2000; Schneider and Linftert, 2003; Cummins et al., 2006; Falé and Faria, 2006) and for pitch accents (Ladd and Morton, 1997; Chen, 2003).

The CP paradigm is a particular two-fold experimental methodology which has come to be used to investigate how we manage to extract what is linguistically meaningful from the variability of the acoustic signal. This method involves, first, a classification task in which the listeners have to categorize stimuli taken from a continuum by giving them phoneme labels, and secondly, a discrimination task in which listeners are asked to judge pairs of stimuli as being either same or different. The results expected are shown in Figure 3. The functions of responses to the classification task have an S-shape, i.e. an abrupt shift from one category to the other rather than a gradual transition. Figure 3 also shows the idealized discrimination function (dashed line). If perception is categorical, discrimination between stimuli is much more accurate between categories than within them. In order for perception to be regarded as truly categorical, the accuracy of discrimination has to be predicted from the results of the classification task.

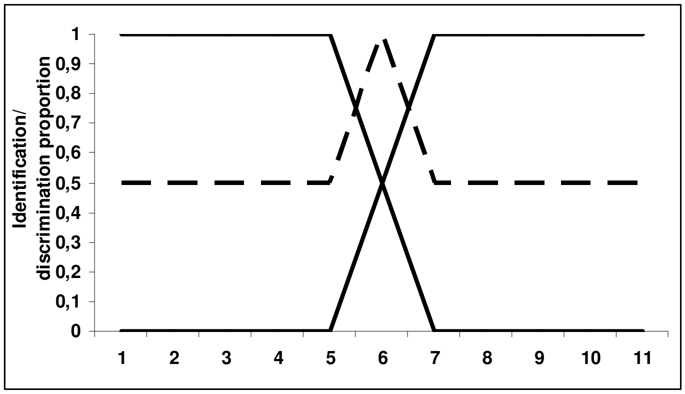


Figure 3. Idealized identification and discrimination functions.

Liberman et al. (1957) were the first to apply this method to speech perception. They created an acoustic continuum that varied across voiced consonants *b*, *d*, *g* in the extent and direction of the second formant transitions. In the identification task, they found that the stimulus continuum was divided into "three sharply bounded phonemic categories". In the discrimination task, they reported an important difference between the subjects: "[…] the data obtained with this S are not all so neat and striking as the particular examples chosen, and some of the other Ss were more variable, especially in their responses to the discrimination task […]". However, they concluded from the inspection of the data that those stimuli to which were attached different phonemic labels were more consistently discriminated between than those which were put within any one category. Thus, they obtained a discrimination peak of approximately 75% correct discrimination between stimuli separated by only one step in the continuum compared with 100% correct discrimination in two-step and three-step discrimination curves. This was only for a single subject CD. Finally, they compared obtained and predicted discrimination and concluded that discrimination curves predicted well the "occurrence and location of points of high and low discriminability in the obtained functions", but that subjects tended in general to discriminate between the stimuli better than had been predicted.

1.3.2. The CP applied to intonational contrasts

The CP paradigm has been applied not only to consonantal feature contrasts but to vowel feature contrasts too. It is well-known that there is a clear difference in degree of categorical perception between consonants and vowels, especially for those features related to the perception of stop consonants. According to Gerrits and Schouten (2004), who also quote Pisoni (1973), Pisoni (1974) and Tartter (1981), this difference may be due to differences in cue duration. The essential acoustic cues for stop consonants are rapidly changing formant transitions and a brief noise burst; in contrast, vowels are assumed to remain uniform over a much longer duration. This difference in cue duration has an effect on the availability of auditory memory for these two classes of speech sounds. According to Pisoni (1973) and Fujisaki and Kawashima (1971), discrimination may be performed either in an auditory mode or in a labeling mode (in the terminology of Gerrits and Schouten). The auditory mode only uses bottom-up stimulus information, whereas in the labeling mode a subject has to categorize the stimuli and then discriminate between the pair of stimuli based on the categories to which he has previously assigned the stimuli. In the case of consonants, the discrimination would be performed in a phonetic mode, since the change is too fast to allow an auditory comparison. This is not the case for vowels, however, which are perceived less categorically. The results of Pisoni (1973) and Schouten and van Hessen (1992) suggested that this low degree of categorical perception of vowels could be due to the nature of the stimulus material, which had been modeled on production in isolated words. Gerrits and Schouten (2004) hypothesized that, in running speech, temporal reductions and more complex spectral coding of the vowel would make vowel perception more categorical. For that reason, they studied the difference in perception between vowels spoken in isolated words and in a read text; however, they obtained no categorical perception in either condition.

Not much is known about the case that we are dealing with, namely the appropriateness of applying the CP paradigm to intonational contrasts. Kohler (1987) applied the paradigm to differences in peak alignment in German contours and concluded that the phenomenon of categorical perception also applied to the field of prosody. D'Imperio and House (1997) used the identification task to determine whether the primary cue in distinguishing interrogatives from declaratives in Italian involves only the alignment of the high level target with the syllable or if the category percept also depends on the presence of a rising or falling melodic movement within the syllable nucleus. Since the purpose of this experiment was not to determine if this contrast was categorical (they were actually two different categories) but rather what the primary cue was for questions and statement in Neapolitan Italian, the appropriateness of using this method to study intonational contrasts was not assessed

The CP paradigm has also been applied to differences in pitch height of boundary tones. Remijsen and van Heuven (1999) illustrated the application with a well-known intonational contrast in Dutch between a final fall (H*L- L%) and a final rise (H*L- H%). They interpreted the results as clearly showing CP for L% vs. H%. They concluded that tonal dimensions can be perceived categorically and that the CP technique is sensitive enough to determine categoriality in tonal dimensions. Post (2000) tested the hypotheses that there is a categorical distinction between H*0%[3] and H*H% in French and that changes in timing and pitch direction in IP-final

[3] According to Post (2000), H*H%, H*0% and H*L% represent Intonation Phrase final high rising, rising, and rising-falling movements respectively.

rises are gradient by applying the CP method. For that purpose, and following Gussenhoven (2004), Post (2000) used as well as the test continua, two control continua, that is, one continuum which was unequivocally continuous and another which was unequivocally discrete. She obtained mixed results. On one hand, the identification scores for the hypothesized categorical continuum appeared to reflect categorical perception, but the cross-over point in the identification function did not correspond to the discrimination peak. On the other hand, the results for the other continua all pointed towards gradient variation, even the continuum in which variation was undoubtedly categorical. Thus, she concluded that these mixed results may indicate that the identification results were task induced (subjects may have made an exclusively acoustic and not a linguistic decision) and that a discrimination task may not be suitable for the investigation of the intonational contrasts tested. Schneider and Linftert (2003) examined the contrast between low (L%) and high (H%) boundary tones between statements and questions in German. The curve for the identification function showed a clear S-shape with a clear crossover from one category to the other within 5 steps along the continuum. However, the discrimination function agreed with neither the notion of continuous perception, nor CP, since there was a plateau that correlated to a low but significant degree with a broad crossover in identification. According to Schneider and Linftert (2003), the results suggested that high and low boundary tones are indeed perceived categorically, but they think that the classical view of CP is not appropriate for dealing with intonational events perhaps because intonation transmits more information than segmental features do. Cummins et al. (2006) tested the discrimination of the phrase-final pitch contours from statement to question in English. As they were attempting to shed light on the relationship between discrimination sensitivity and category structure, they used speech and non-speech stimuli. They found that on one hand the boundary between categories did not coincide with the discrimination peak and that, on the other hand, the non-speech stimuli were better discriminated than either of the speech stimuli. Hence, they inferred that it is untenable to try to explain these results on the basis of the classical CP. Falé and Hub Faria (2006) tested the categorical nature of questions and statements in European Portuguese from the differences of boundary tones. They applied RT measurements to the identification task. From their identification results they deduced that the contrast is indeed categorical even though RT measurements point to continuous rather than categorical perception. The absence of a discrimination peak in the crossover between categories supports the continuous perception view. According to Falé and Hub Faria and as confirmed by the results of Remijsen and van Heuven (1999) and Schneider and Linftert (2003), the CP definition needs to be modified in order to better explain the data related to the application of CP methodology to intonational contrasts.

In applying the CP method to a peak height contrast in English, Ladd and Morton (1997) found that although an identification boundary occurred at a particular frequency in their stimuli, no discrimination peak was observed at the identification boundary. As this does not agree with the essence of the CP paradigm, they concluded that the distinction between normal high accent and extra high accent was not categorically perceived. Chen (2003), from the results of Ladd and Morton (1997), argued that the absence of peak could be due to the unsuitability of applying the CP paradigm to a pitch height contrast. Thus, Chen (2003) proposed an alternative method to the Reaction Time measurement. From the results of the application of the RT measurement to a classification task, she concluded that the combination of this method with response frequencies in the identification task is an effective way to distinguish the real linguistic identification categories from task-induced identification categories.

Table I presents a summary of the studies that have employed the CP paradigm to study intonational differences. It includes the criteria that according to Francis et al. (2003) a continuum must fulfill in order to be considered categorical (see §2 for a more detailed explanation). In an attempt to reformulate a new definition of CP paradigm in accordance with the results obtained from the application of this method to intonational contrasts, we should take into account some different points. First, it seems that, as in the case of vowels, categorical perception is poorer in intonational contrasts than in consonants. That is inferred not only from the discrimination results but also from the identification results. Identification results have a clear category boundary that could be task-induced, especially when the crossover point does not correspond to the discrimination peak. Second, as though there is not always peak discrimination, this fact, according to Chen (2003), does not necessarily mean that a contrast is non-categorical, some researchers propose a Reaction Time approach to test the hypothetical discreteness of a contrast. This method is endorsed by the findings of Pisoni and Tash (1974) in a continuum that ranged from /ba/ to /pa/. They found that RT was higher for the stimulus at the phonetic boundary and lower within categories. Chen (2003) obtained parallel results. Third, in attempting to provide an explanation for this difference in categorical perception between consonantal and intonational contrasts, Schneider and Linftert (2003) have pointed out that it could be because intonation transmits more information than segmental features do. However, it is worth noting that both Cummins et al. (2006) and Falé and Hub Faria (2006) used shortened stimuli to avoid problems in their discrimination task caused by auditory memory limitations. Observe that this caution did not improve their results.

Study	Language	Contrast	Identification predicts discrimination (*)	Discrimination peak corresponds to the crossover point between categories (!)	Discrimination rate at the peak	Other measurements
Kohler (1987)	German	"Established" vs. "new" vs. "emphatic" (peak alignment)	Not applied	Yes	Approx. 40% "different" (0 steps), approx. 50% "different" (1 step), approx. 90% "different" (2 steps)	
Ladd and Morton (1997)	English	Normal high pitch accent vs. emphatic high pitch accent (pitch height)	Not applied	Yes	Between 50% and 60% AB hit rate, approx. 20% BA hit rate and false alarm rate	
D'Imperio and House (1997)	Neapolitan Italian	H*, statement-narrow focus vs. L+H*, yes-no question (peak alignment)	Not applied	Only identification task is performed		
Remijsen and van Heuven (1999)	Dutch	Final fall, H*L-L% vs. final rise, H*L-H% (pitch height, boundary tones)	Yes	Yes	Approx. 80% AB hit rate, approx. 40% BA hit rate, approx. 60% (AB and BA hit rate averaged), approx. 20% (false alarm rate)	
Post (2000)	French	H*0% vs. H*H% (pitch height, boundary tones) and variation in timing and pitch direction in IP final rises	Not applied	Only in one of the five continua tested	Between 25-75% "different"	
Chen (2003)	English	Normal high accent vs. emphatic high accent (pitch height) and early peak alignment vs. late peak alignment (peak alignment)	Not applied	Only identification task is performed		Reaction Time
Schneider and Linftert (2003)	German	Phrase final pitch contour, statement vs. question (pitch, height, boundary tones)	Yes	Low but significant correlation	Approx. 80% AB hit rate, approx. 70 to 60% BA hit rate (plateau), approx. 75 to 70% (AB and BA hit rate averaged, plateau), approx. 20 % (false alarm rate)	
Cummins et al. (2006)	English	Phrase final pitch contour, statement vs. question (pitch height, boundary tones); shortened stimuli	Not applied	Only for one subject	Approx. 60% in speech, voiced, hum and buzz continuum (number of hits minus number of false alarms)	
Falé and Hub Faria (2006)	European Portuguese	Phrase final pitch contour, statement vs. question (pitch height, boundary tones); shortened stimuli	Not applied	No	Approx. 50% (AB hit rate)	Reaction Time

Table I. Studies that have employed the CP paradigm. It includes the criteria that according to Francis et al. (2003) a continuum must fulfill in order to be regarded as categorical —(*) and (!).

1.3.3. CP paradigm and presence of asymmetries

The application of the CP paradigm seems to involve the presence of asymmetries that have been found repeatedly in the application of CP to vowel perception (Polka and Bohn, 2003), perception in intonational languages (Kohler, 1987; Ladd and Morton, 1997; Remijsen and van Heuven, 1999; Cummins et al., 2006; Falé and Faria, 2006), perception in tonal languages (Valter and Ciocca, 2003). Asymmetries or order of presentation effects occur such that discrimination of a particular change in one direction is easier compared to the same change presented in the reverse direction.

With respect to vowel perception, Polka and Bohn (2003) review studies that report asymmetries in vowel perception in infants and propose that this data indicate that babies are predisposed to respond differently to vowels that occupy different positions in the articulatory/acoustic vowel space (defined by F1-F2), insofar as they perform better in vowel discrimination when the direction of the vowel change is from a less peripheral towards a more peripheral vowel in the vowel space. For instance, in testing discrimination of the German /u/-/y/ and /ʊ/-/ʏ/ contrasts by English-learning infants' in two age groups of 6-8 and 10-12 months, Polka and Werker (1994) found that both 6-8 and 10-12 months olds performed better when the direction of change was from /y/ to /u/ (rather than from /u/ to /y/) and when the direction of change was from /ʏ/ to /ʊ/ (rather than from /ʊ/-/ʏ/). They present new data that support their hypothesis and compare these results with findings obtained with birds and cats. According to them, these comparisons suggest that asymmetries evident in humans are unlikely to reflect general auditory mechanisms.

Regarding intonational languages, asymmetries have been found in the application of the CP paradigm to a contrast in peak alignment and to pitch scaling contrasts. Kohler (1987) applied the CP paradigm to a contrast in peak alignment in order to show that it can be applied to pitch contours as well as phonemic differences. From the discrimination data, he reported a strong order effect that results, according to him, from the perceptual testing procedures. However, we know nothing about this order effect besides the fact that it "disturbs the differentiation function and has to be taken into account". In tonal scaling, Ladd and Morton (1997) applied the categorical perception task to a contrast that had been considered gradient. Their results suggested that the normal/emphatic distinction may be categorically interpreted but not categorically perceived. Additionally, they reported order of presentation effects in the discrimination task. Listeners were more sensitive when the first stimulus of the pair had a lower peak than the second. Remijsen and van Heuven (1999) also reported an order of presentation effect in their discrimination data: when they applied the CP method to the contrast between (H%) and (L%) boundary tone in Dutch, they found that two different contours are more successfully discriminated when the second one has the higher terminal pitch. Likewise, Schneider and Lintfert (2003) tested the categorical status of high and low boundary tones in German and reported order of presentation effects. Thus, participants were more sensitive to stimuli pairs when the second stimulus had a higher final pitch. They concluded on this basis that the correlation between the boundary between categories and the discrimination peak was significant, and that this contrast had "categorical status". Cummins et al. (2006) investigated the discrimination of phrase final pitch contours within a continuum from statements to questions in English. In order to clarify the relationship between linguistic categories and the sensitivity of discrimination, they used speech and non-speech stimuli. However, the peak that they found in the discrimination results was not related to the category boundary revealed in the identification task. Moreover, discrimination was better for non-speech stimuli than for speech. Asymmetrical discrimination was also reported here; pairs whose higher stimulus was the second one were better discriminated again. They speculate that the absence of an interaction between stimulus type and presentation order suggests that the effect may be a general property of pitched stimuli, rather than speech specific. Falé and Faria (2005) tested the intonational contrast between statement and question in European Portuguese on a CP based paradigm. They found order of presentation effects when the direction of change was upwards; in that case, the pairs of stimuli were better discriminated.

With regard to tonal languages, Francis and Ciocca (2003) argued that listeners' auditory discrimination of vowel sounds depends in part on the order in which stimuli are presented. In their study, Cantonese listeners were more sensitive to a difference of 4 Hz in a linguistic tone continuum when the first syllable was lower in frequency than the second. American English-speaking listeners participated in this experiment in the same manner using the same stimuli but showed no such contrast effect. A non-speech continuum was used too in which tokens had the same fundamental frequencies as the Cantonese speech tokens but a non-speech-like timbre. Neither English nor Cantonese listeners showed any order of presentation effects. They concluded that tone presentation order effect may be language-specific.

There has been some debate about whether such presentation effects are language-specific or not. At this moment it seems to be an unresolved question. However, it is clear that perceptual asymmetries exist in other perceptual spaces like color, line orientations, and numbers (Polka and Werker, 1004; Rosch, 1975); and geometric figures and country concepts (Tversky and Gati, 1978). The efforts invested in clarifying whether these asymmetries are language independent or not seem to yield contradictory results. That is, while Francis and Ciocca (2003) did not find any order effect in the performance of the discrimination of a non-speech continuum by English and Cantonese listeners, in Cummins et al. (2006) the asymmetry was demonstrated for both speech and non-speech stimuli. However, regarding the asymmetries which have arisen in the application of the CP paradigm to pitch scaling contrasts, it seems that all the results point in the same direction, namely that asymmetries occur in such a way that a tonal change is easier to detect when the first stimulus is lower in frequency than the second.

1.4. Outline

This study comprises four chapters. The first chapter contains this introduction, which presents the contours studied, that is, yes-no questions and wh-questions in Majorcan Catalan, reviews the literature regarding the categorical effects of pitch height and next, states the aims and purposes of this study. Finally, it also contains a preliminary discussion on the implications of applying the CP paradigm to intonational contrasts.

The second chapter presents a production experiment designed to describe the contours of yes-no questions and wh questions in Majorcan Catalan, specifically the behavior of the two particles *que* and *què* that introduce these two types of questions and the acoustic characteristics of the falling nuclear accent that characterizes them.

Chapter three presents a perception experiment based on the Categorical Perception (CP) paradigm which tests whether listeners make categorical linguistic use of F0 scaling differences in perceiving yes-no questions as opposed to wh-questions in Majorcan Catalan. The application of the CP paradigm has been carried out bearing in mind the clear difference in degree of categorical perception that seems to exist between segmental and intonational contrasts and the conclusions which previous studies have come reached about the suitability of applying the CP paradigm to intonational differences as well as alternative or additional measurements or approaches.

Chapter four presents a summary of the major findings of this study and its general conclusions.

2. Production experiment

As mentioned in §1, there are two important differences between yes-no questions and wh-questions in Majorcan Catalan, namely a difference of accentuation in the two interrogative particles *que* ('that') and *què* ('what') and a tonal scaling difference in the H leading tone of the respective falling nuclear accents. Both characteristics will be investigated in this production study in order to determine how they interact with sentence type and utterance length factors. Five native speakers of Majorcan Catalan participated in this production experiment, reading a total of 450 utterances split into two sentence types (yes-no questions and wh-questions) of different lengths (from 1 to 3 pitch accents). It is expected that the presence/absence of accent in the particles *que* ('that') and *què* ('what') might trigger differences in pitch, vowel quality or duration. Additionally, this production study is expected to prove the existence of a super-high variant of the high pretonic nuclear tone in yes-no questions as opposed to a "normal" or plain high tone in wh-questions.

2.1. Method

2.1.1. Speech materials

The corpus that has been used in this study contains utterances of 2 sentence types (yes-no questions and wh-questions) of different lengths (from 1 to 3 pitch accents, without counting the accent of the interrogative particle *què* ('what'), in the case of wh-questions). The database was divided into blocks depending on the number of pitch accents of the utterances. Every block had pairs of utterances (yes-no question and wh-question) that were identical or very similar at the segmental level (see some examples in Table I below). Stressed syllables are marked in boldface. Note that the vowel that separates the third pitch accent from the second pitch accent (in utterances with three pitch accents) was generally elided, thus in most cases the second and third pitch accents were in clash condition (see examples in block III, Table I).

Yes-no questions	Wh-questions
BLOCK I (1 pitch accent)	
*Que en va **ven**dre?* ('Did s/he sell any of it?') [kəɱ va ˈvəɲ̟drə]	***Què** en va **ven**dre?* ('What did s/he sell of it?') [ˌkəɱ va ˈvəɲ̟drə]
BLOCK II (2 pitch accents)	
*Que n'espe**ra**va **ven**dre?* ('Did s/he expect to sell any of it?') [kə nəspeˌɾavə ˈvəɲ̟drə]	***Què** n'espe**ra**va **ven**dre?* ('What did s/he expect to sell of it?') [ˌkə nəspeˌɾavə ˈvəɲ̟drə]
BLOCK III (3 pitch accents)	
*Que hi vo**li**a tor**nar** a **ven**dre?* ('Did s/he want to sell there again?') [kəj vuˌliə torˌna ˈvəɲ̟drə]	***Què** hi vo**li**a tor**nar** a **ven**dre?* ('What did s/he want to sell there again?') [ˌkəj vuˌliə torˌna ˈvəɲ̟drə]

Table I. Examples of the utterances used in this experiment. Stressed syllables are marked in boldface.

Whenever possible, the words were composed of sonorant consonants (to avoid segmentally-induced effects on the F0 curve). The last word always contained the stress on the penultimate syllable to provide more room for nuclear accent and boundary tones realization. We obtained a total of 450 utterances (5 sentences x two types x 3 lengths x 3 repetitions x 5 speakers).

Note that the interrogative particle *què*/*que* is stressed in wh-questions and unstressed in yes-no questions. Since the interrogative particle *què* is stressed, a pitch accent is also associated with this syllable, whereas by contrast the unstressed interrogative particle *que* is unaccented. Thus, in our data we had a two-way contrast, that is, [+stress] [+accent] vs. [-stress] [-accent]. Consequently, if we rely solely on our test material, it is not possible to distinguish the effects of stress from those of accent.

2.1.2. Measurements

Various measurements were performed manually, and collected by Praat software automatically through a script into a file in .txt format.

Time measurements:

-Onset of the interrogative particles *que*/*què* (c0).

-Onset of the vowel of the interrogative particles *que*/*què* (v0).

-Offset of the interrogative particles *que*/*què* (c1).

-Onset of the nuclear stressed syllable (c2).

-Onset of the nuclear stressed vowel (v2).

-Offset of the nuclear stressed syllable (c3).

-End of the utterance (end).

Pitch measurements:

-F0, and F1 and F2 of the vowel of the interrogative particles *que*/*què* (f0v0/F1v0/F2v0).

-F0 peak within H+L* (f0H)

-F0 minimum within H+L* (f0L).

-F0 utterance-final value (f0end).

The choice of these measurements was motivated by the goals of this study. Thus, as we were interested especially in the tonal differences of the leading tone within the nuclear falling accent between yes-no questions and wh-questions as well as the different acoustic characteristics that might be triggered by the presence/absence of accent in the interrogatives particle *que*/*què*, no measurements related to prenuclear pitch accents were taken. This fact was favored by the impossibility of defining the prenuclear accents in terms of peaks and valleys, and also by the presence of clash between the second and third pitch accents in utterances of three pitch accents caused by the elision of the preposition *a* ('to'). This elision was always present in the utterances in Block III.

Figures 1, 2 and 3 show the waveforms and F0 contours of the utterances of Table I (speaker MV), as well as landmarks for time measurements (tier 1) and landmarks for pitch measurements (tier 2).

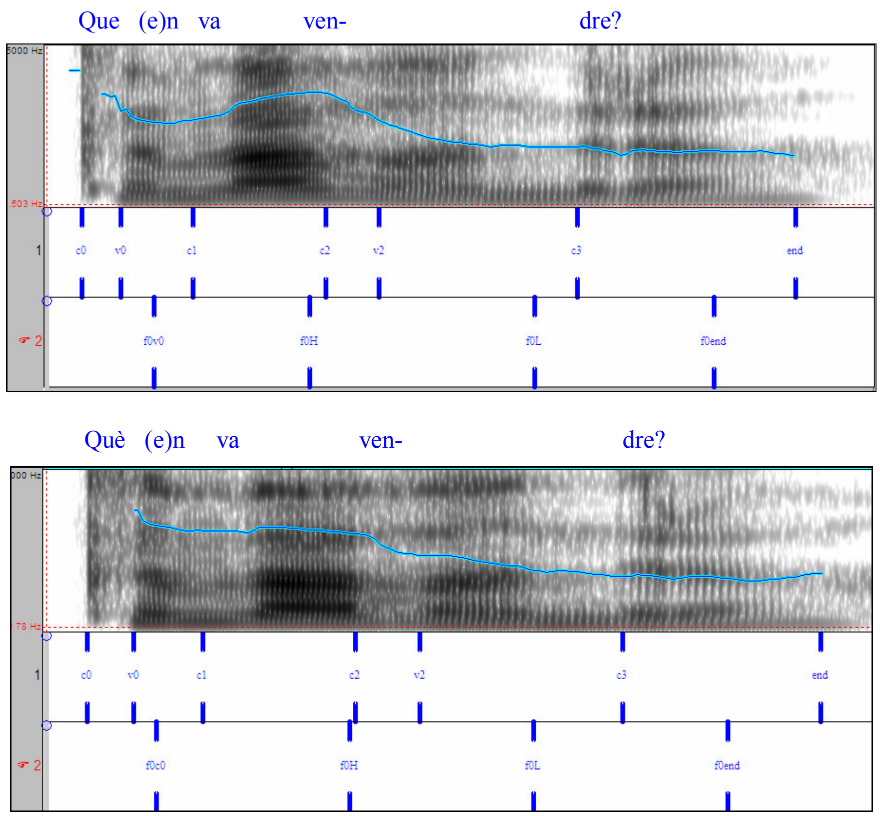


Figure 1. Waveform and fundamental frequency contour of yes-no question *Que en va vendre?* ('Did s/he sell any of it?'), upper panel, and wh-question *Què en va vendre?* ('What did s/he sell of it?'), bottom panel. Landmarks for time measurements are marked in tier 1 and landmarks for pitch measurements in tier 2.

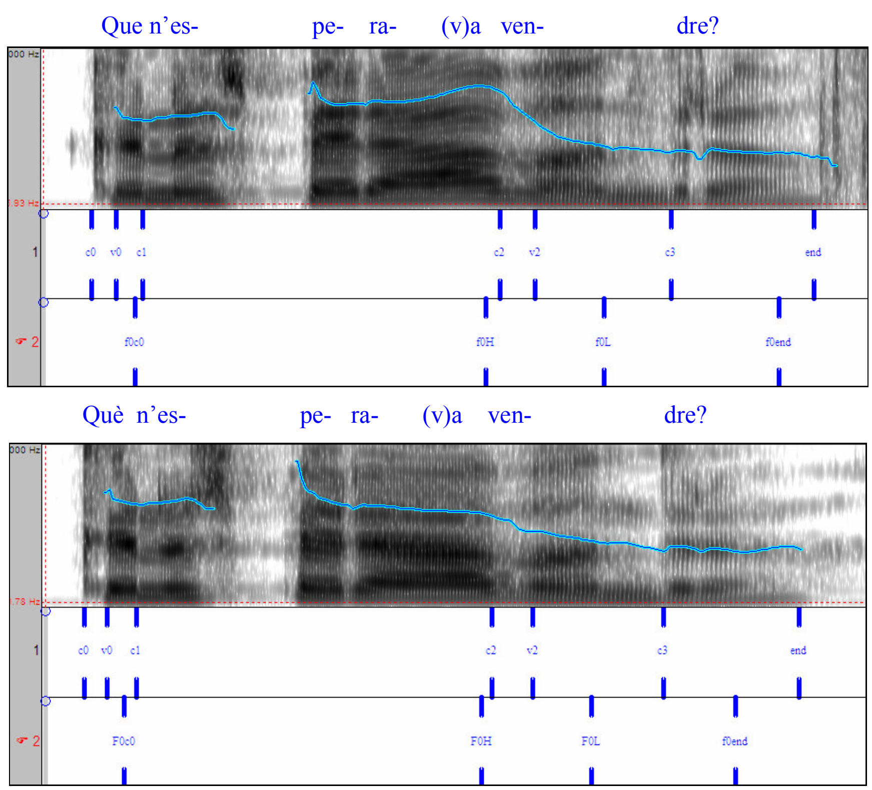


Figure 2. Waveform and fundamental frequency contour of yes-no question *Que n'esperava vendre?* ('Did s/he expect to sell any of it?'), upper panel, and wh-question *Què n'esperava vendre?* ('What did s/he expect to sell of it?'), bottom panel. Landmarks for time measurements are marked in tier 1 and landmarks for pitch measurements in tier 2.

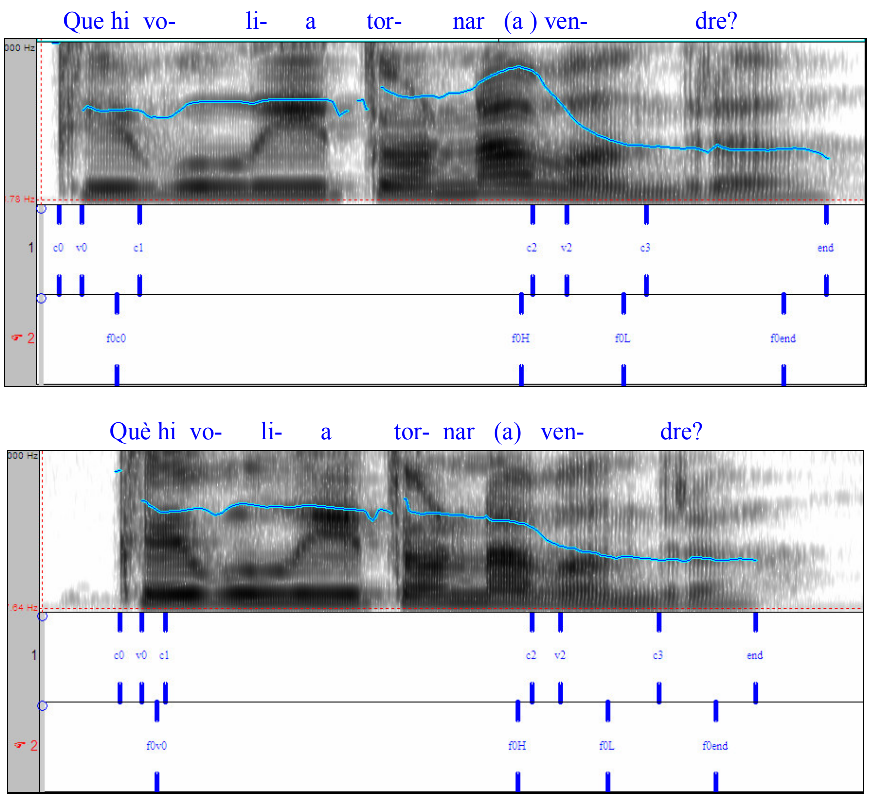


Figure 3. Waveform and fundamental frequency contour of the yes-no question *Que hi volia tornar a vendre?* ('Did s/he want to sell there again?'), upper panel, and wh-question *Què hi volia tornar a vendre?* ('What did s/he want to sell there again?'), bottom panel. Landmarks for time measurements are marked in tier 1 and landmarks for pitch measurements in tier 2.

These are the criteria that were followed in order to determine these measurements:

-*F0v0*. As there was clear peak in the case of neither the accented interrogative particle nor the unaccented interrogative particle, the point for F0 measurement was located at the end of the vowel in order to avoid the rise in F0 due to the presence of the voiceless obstruent [k] (Figure 4). Observe that when the interrogative particles were followed by clitics *en* and *hi,* interrogative particle and clitic together made up only one closed syllable. In these cases, the final time landmark of the interrogative particle was located at the end of the syllable, not at the end of the vowel of the interrogative particle (see Figure 5).

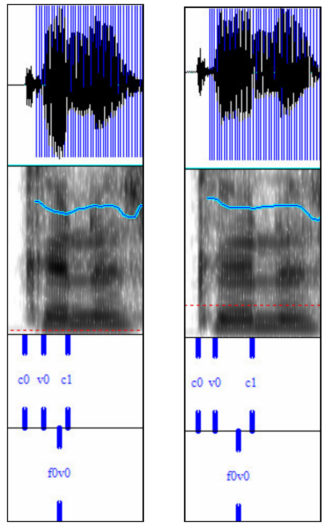
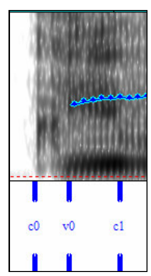


Figure 4. Time and pitch landmarks for the unaccented (left panel) and accented (right panel) interrogative particles for the utterances *Que n'esperava vendre?* ('Did s/he expect to sell any of it?') (left panel) and *Què n'esperava vendre?* ('What did s/he sell of it? ') (right panel).



[k əj]

Figure 5. Time landmarks of the interrogative particle of the utterance *Que hi va veure?* ('Did s/he see?').

-*F1v0 and F2v0*. The first and second formant frequencies were extracted automatically using Praat software. By choosing the "Show formants" command and "Get first/second formant", the Praat software extracted the mean F1 and F2 in the vowel stretch of the interrogative particles.

-*F0H*. The peak did not generally present problems of location. The landmark was situated just before the fall. Sometimes a plateau was found. When this happened, the H was placed at the end of the plateau, immediately before the contour started to fall.

-*F0L*. In the minima, the L was located just after the fall, at the point when F0 falling velocity stabilized, after the elbow. In yes-no questions it was easier to determine the L point because there was a larger slope than in wh-questions and the change in F0 velocity was more noticeable. On the other hand, in wh-questions the fall was more progressive and the elbow was not so clear. In these cases, the L point was located just after the inflection at the point when the F0 ceased to fall in a well-defined manner.

Automatic procedures have been used by Pierrehumbert and Beckman (1988), D'Imperio (2000), Frota (2002) and Welby (2006) to determine L points in a more consistent manner. This procedure consists of the use of several scripts to extract fundamental frequency, time information and the position of the elbows. D'Imperio (2000) describes the procedure used to calculate the position of the elbow: "[…] an automatic procedure was employed by which two straight lines were fitted to the f0 segment going from "b" to F0max [two reference points]. The parameters of the two linear models were estimated by means of conventional linear least-squares methods. To estimate the elbow position, i.e., the intersection of the two fitted lines, two linear regressions were computed for each possible elbow location […]". Though automatic procedures have not been used here, since the L points were not the focus of this study, future research comparing the detection of L points in falling accents by hand and automatically is planned.

2.1.3. Speakers and experimental procedure

Five speakers of Majorcan Catalan, two females (MC and MV) and three males (AB, AC and FF) read three repetitions of the three blocks of utterances. Utterances were ordered within each block so that minimal pairs were never grouped together. In order to facilitate the reading and interpretation of the utterances by the speakers, each test question was answered verbally by the author of this study (see Appendix for full list of questions and answers). Speakers were asked to read the utterances at a normal speech rate and avoid marked readings. This was necessary because when they realized that there were pairs of utterances that were exactly identical except for the particle *que/què,* speakers sometimes tried to unnaturally mark the difference between both particles. In such cases, subjects were asked to repeat the utterance and try to avoid the marked reading. Speaker AC had a low F0, with the result that some contours were not perfectly obtained. This made it difficult for Praat to collect L points. As a consequence, seventeen utterances by this subject were eliminated from the database. The final database under analysis thus contained a total of 433 utterances.

Speakers were recorded on a TASCAM TEAC Professional Division digital audio tape recorder and a SHURE PG81 microphone. The recorded utterances were digitized at a 44100 sample rate using Goldwave software.

2.1.4. Statistical analyses and graphing methods

Non-parametric tests were used because general assumptions regarding the data for using parametric tests were not fulfilled, that is: 1) the variables in our data set violate the normality assumption; and 2) the sample size is small (5 subjects).

The following variables were analyzed to see how they interacted with sentence type (yes-no question or wh-question) and/or utterance length (1-3 pitch accents): 1) duration, F0, F1 and F2 of the vowel of the interrogative particle *que*/*què;* 2) F0 peak and F0 minima within nuclear accent H+L*; 3) duration of the nuclear stressed syllable.

The following statistical tests were used (see Blalock, 1979 for an explanation of non-parametric statistics):

–In order to test how sentence type affects the variables listed above, a Wilcoxon matched pairs signed rank test was carried out. It is a non-parametric alternative to the paired Student's t-test for the case of two related samples or repeated measurements on a single sample.

–To test if utterance length affects these variables within the same utterance type, a Friedman test was used. It is a multiple testing technique similar to parametric ANOVA.

Box plots were used as a graphing method because they provide a visual summary of many important aspects of the distribution. Each one of the horizontal lines represents an important number related to the data set. The first and last lines (lowermost and uppermost) are drawn at the lowest and highest data values. The three lines that make up the box are drawn 25%, 50%, and 75% of the way through the data, and together are called a five-number summary (Figure 6).

The middle line represents the median, that is, the middle value of the distribution. The lowermost line is the twenty-fifth percentile of the data, called the first or lower quartile. The lower quartile is the median of the first 50% of the data. The uppermost line is the seventy-fifth percentile of the data, that is, the third or upper quartile. It is the median of the last 50% of the data. Therefore, the box contains the middle half of the scores.

The interquartile range (IQR) is defined as the difference between the third quartile and the first quartile. Outliers with values between 1.5 x IQR and 3 x IQR are represented by "o". Outliers with values greater than 3 x IQR are represented by "*".

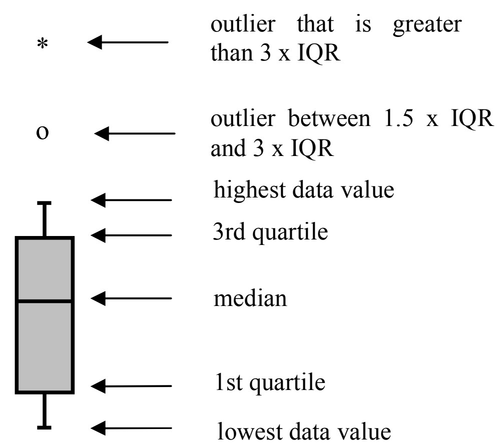


Figure 6. Schematic representation of a box plot with the five-number summary.

2.2. Results

2.2.1. Phonetic realization of particles *que*/*què*

2.2.1.1. Fundamental Frequency (F0)

As noted above, since no peak could be visually identified either in the case of the accented interrogative particle (wh-questions) or in the unaccented interrogative particle (yes-no questions), the point for F0 measurement was located at the end of the vowel in order to avoid the rise in F0 due to the presence of the voiceless obstruent [k]. Figure 7 shows the box plots of F0 for the vowel [ə] of the particle *que*/*què* in yes-no questions (grey boxes) and wh-questions (white boxes) for the five speakers that participated in the experiment. Observe that there is an important difference between median F0 points of the interrogative particle between yes-no questions and wh-questions for the five speakers separately. In yes-no questions the interrogative particle has a lower F0 median and in wh-questions, a higher F0 median. This difference in the F0 median for the vowel of the particle *que*/*què* is 20 Hz for speaker AB, 30 Hz for speaker AC, 16 Hz for speaker FF, 14 Hz for speaker MC and 29 Hz for speaker MV.

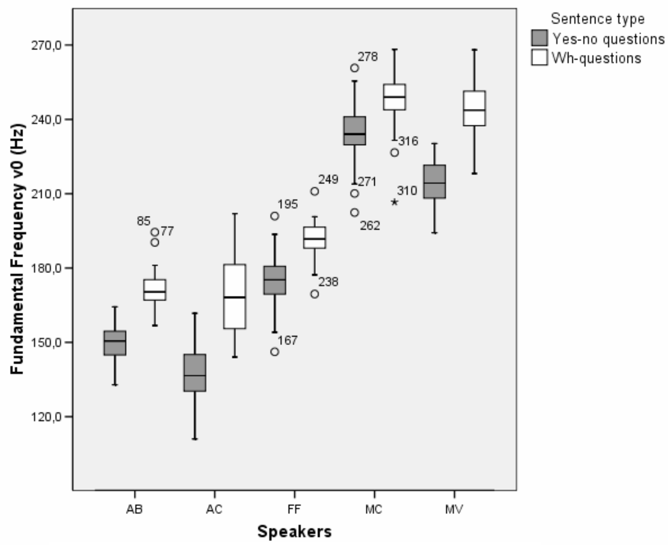
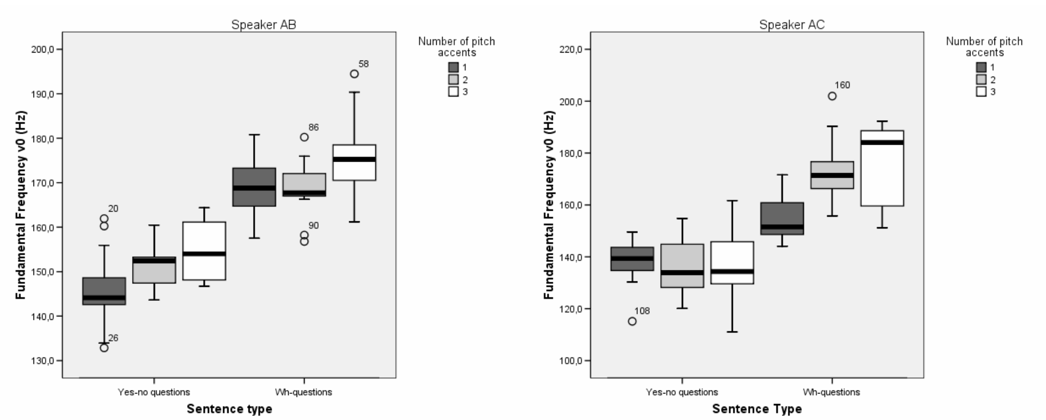


Figure 7. Box plots of F0 for the vowel [ə] of the interrogative particles *que*, yes-no questions (grey boxes) and *què*, wh-questions (white boxes) for all speakers.

The five graphs in Figure 8 show the box plots of F0 for the vowel [ə] of the particle *que*/*què* in yes-no questions (left-hand boxes) and wh-questions (right-hand boxes) for each of the five speakers, broken down into different utterance lengths (1-3 pitch accents).



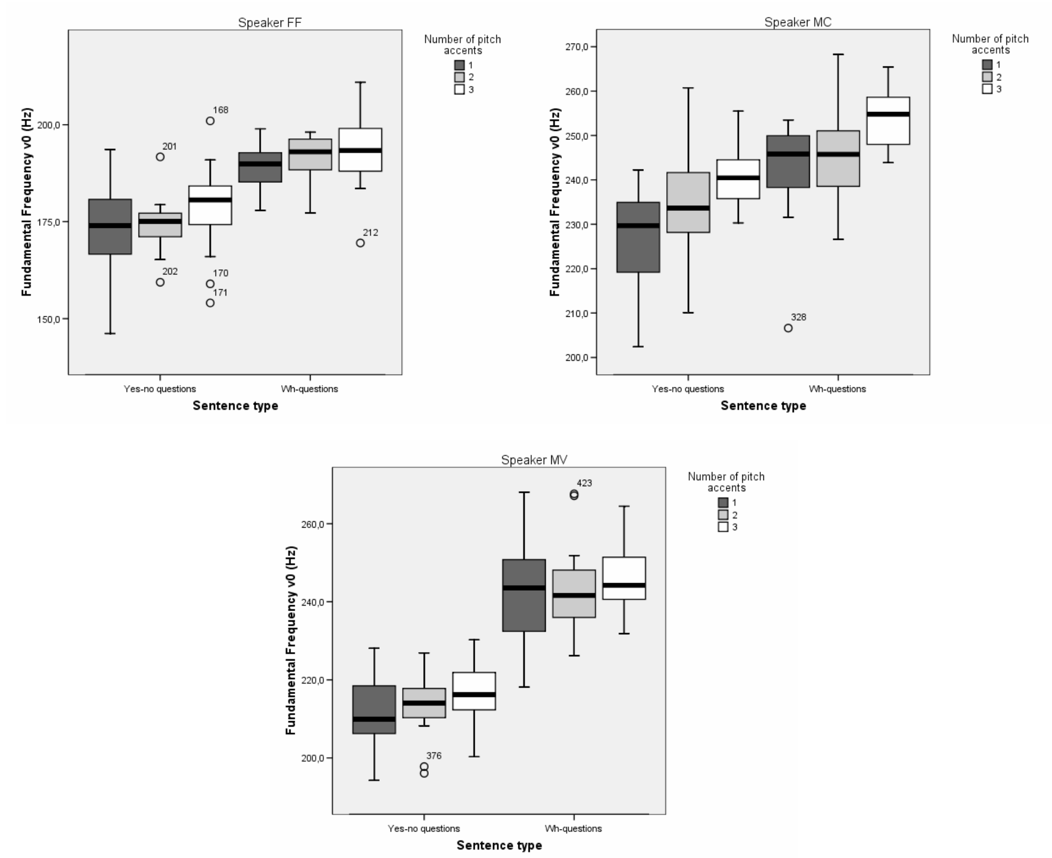


Figure 8. Box plots of F0 for the vowel [ə] of the interrogative particles *que*, yes-no questions (left-hand boxes) and *què*, wh-questions (right-hand boxes) in utterances of different lengths (1 pitch accent in dark grey, 2 pitch accents in light grey and 3 pitch accents in white) for each speaker.

Results of the Wilcoxon matched pairs signed rank test and Friedman test (Table II) revealed highly significant effects of sentence type on F0 of the vowel of the interrogative particle (p=0.000) for all the speakers. Moreover, speakers AB, AC and MC showed a significant tendency in wh-questions to anticipate the length of the utterance by beginning higher if the utterance was longer (presumably to avoid the pitch falling much lower) and lower if the utterance was shorter. This anticipation occurs also in yes-no questions, though for only two speakers (AB and MC). Evidence for this idea was presented by Hart (1979), Cooper and Sorensen (1981) and Rialland (2001), who report an influence of phrasal length on the scaling of the initial F0 peaks. This production mechanism in which the speaker would take into account the length of the utterance on producing the initial peak is called global preplanning, as opposed to local preplanning, in which peaks would be constant in a given position regardless of utterance length.

F0 vo	Sentence type	Sentence length	
		Yes-no questions	Wh-questions
AB	p=0.000*	p=0.002*	p=0.041*
AC	p=0.000*	P=0.607	p=0.016*
FF	p=0.000*	P=0.165	p=0.395
MC	p=0.000*	p=0.002*	p=0.002*
MV	p=0.000*	P=0.627	p=0.819

Table II. P-values of the Wilcoxon matched pairs signed rank test (testing the effect of sentence type on F0 of the vowel of the interrogative particle) and Friedman test (testing the effect of utterance length on F0v0) for each speaker. *Statistically significant at 0.05 level.

2.2.1.2. Duration

Figure 9 shows the box plots of duration of the vowel of the particle *que*/*què* in yes-no questions (grey boxes) and wh-questions (white boxes) for the five speakers that participated in the experiment. The graph shows that there is sometimes an effect of duration depending on whether the interrogative particle is accented or unaccented. Thus, when the interrogative particle is accented the vowel tends to be longer, and shorter when it is unaccented. This effect seems to be optional, however, since only for speakers AC, FF and MC is this lengthening effect significant at p=0.000. From the results of Astruc and Prieto (2006) and Prieto and Ortega-Llebaria (2006), it can be said that this lengthening effect may be related to the presence or absence of stress, as opposed to accent. Thus, stressed syllables would be longer than unstressed syllables. Remember that in our data we had a 2-way contrast, that is, [+stress] [+accent] vs. [-stress] [-accent]. Consequently, it was not possible to differentiate the effects of stress from those of accent.

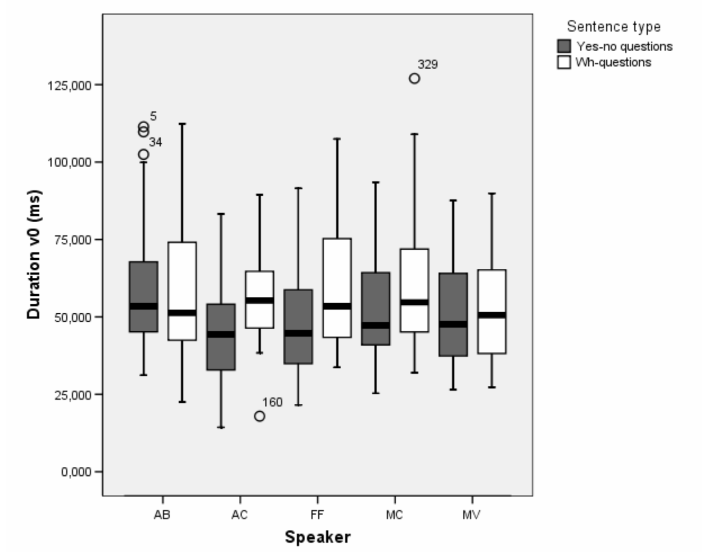


Figure 9. Box plots of duration of the vowel [ə] of the interrogative particles *que*, yes-no questions (grey boxes) and *què*, wh-questions (white boxes) for all speakers.

2.2.1.3. Vowel quality (F1/F2)

The two graphs in Figure 10 show the box plots of F1 (left panel) and F2 (right panel) for the vowel of the particle *que*/*què* in yes-no questions (grey boxes) and wh-questions (white boxes) for all the speakers. The graphs reveal that, depending on whether they are accented or unaccented, vowels display a difference in quality. This quality difference manifests itself in an increment of F1 median for the accented particle (wh-question), while F2 values are generally more similar. On the basis of these results, one can conclude that the difference in quality between the vowel of accented and unaccented particles could be due to a different opening of the vocal tract, with the tract more fully open in the case of the accented/stressed vowel than in the case of the unaccented/unstressed vowel.

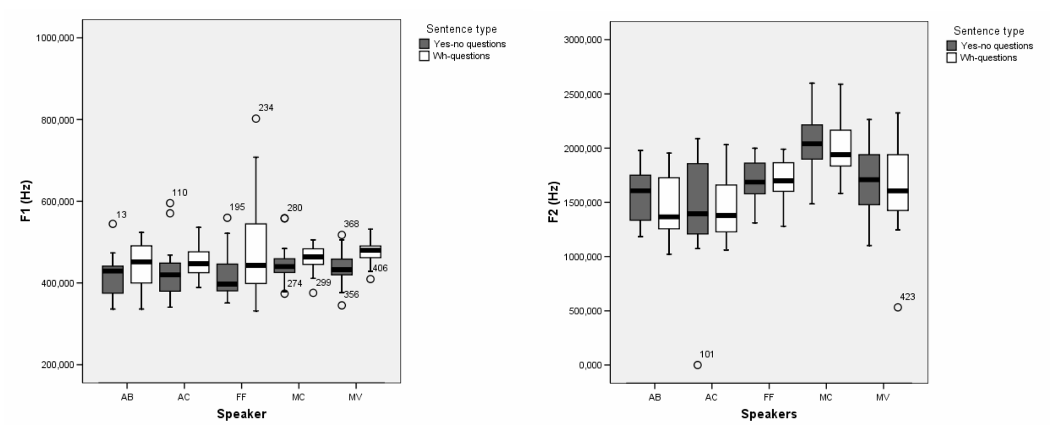


Figure 10. Box plots of F1 (left panel) and F2 (right panel) of the vowel [ə] of the interrogative particles *que*, yes-no questions (grey boxes) and *què*, wh-questions (white boxes) for all speakers.

Results from the Wilcoxon matched pairs signed rank test (Table III) for each speaker individually confirm a significant effect of sentence type on the F1 of the vowels of the interrogative particle for all speakers except speaker AC, but not on the more consistent F2.

F1/F2v0	**Sentence type F1**	**Sentence type F2**
AB	p=0.004*	p=0.286
AC	p=0.052	p=0.518
FF	p=0.001*	p=0.791
MC	p=0.000*	p=0.623
MV	p=0.000*	p=0.875

Table III. P-values of the Wilcoxon matched pairs signed rank test (testing the effect of sentence type on F1/F2 of the vowel of the interrogative particle) for each speaker. *Statistically significant at 0.05 level.

2.2.2. Phonetic realization of the nuclear pitch accent

2.2.2.1. Start of the fall or H leading tone

Figure 11 shows the box plots of F0 at the start of the fall within the falling nuclear accent H+L* in utterances of different lengths (1 to 3 pitch accents) for yes-no questions (left-hand boxes) and wh-questions (right-hand boxes). Note that although there is evidence that utterance length affects the H pitch height value, the H pitch value of yes-no questions is always higher than that of wh-questions.

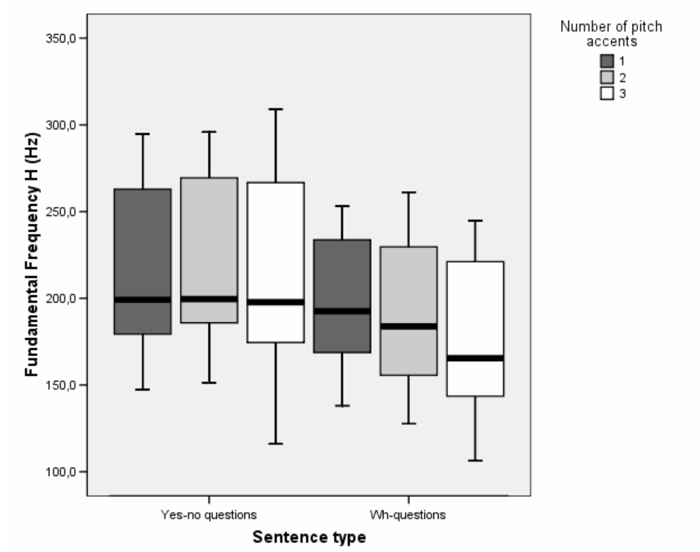


Figure 11. Box plots of H F0 within the nuclear accent H+L* for yes-no questions (left-hand boxes) and wh-questions (right-hand boxes) in utterances of different lengths (1 pitch accent in dark grey, 2 pitch accents in light grey and 3 pitch accents in white) for all speakers.

The five graphs in Figure 12 that plot the results for individual subjects suggest that this utterance-length effect is greater for wh-questions (right-hand boxes) than for yes-no questions (left-hand boxes), since this effect is constant for wh-questions over all subjects but not for yes-no questions. In wh-questions, on the basis of evidence for preplanning for three of the five speakers, I claim that this interaction between length and H pitch height could be due to the presence of downstep. However, as the goal of this study was not to study the behaviour of prenuclear pitch accents, pitch measurements related to the height of these accents were not taken. Such measurements would enable us to know whether every pitch accent had a lower pitch than the preceding pitch accent. Thus, the only evidence for downstep that we have is: 1) the effect of utterance length on pitch value of the leading tone of the nuclear pitch accent, it becoming much lower as the number of pitch accents increases and, 2) the raising of the starting point proportionately to the utterance length for three of the five speakers in wh-questions and for two speakers in yes-no questions.

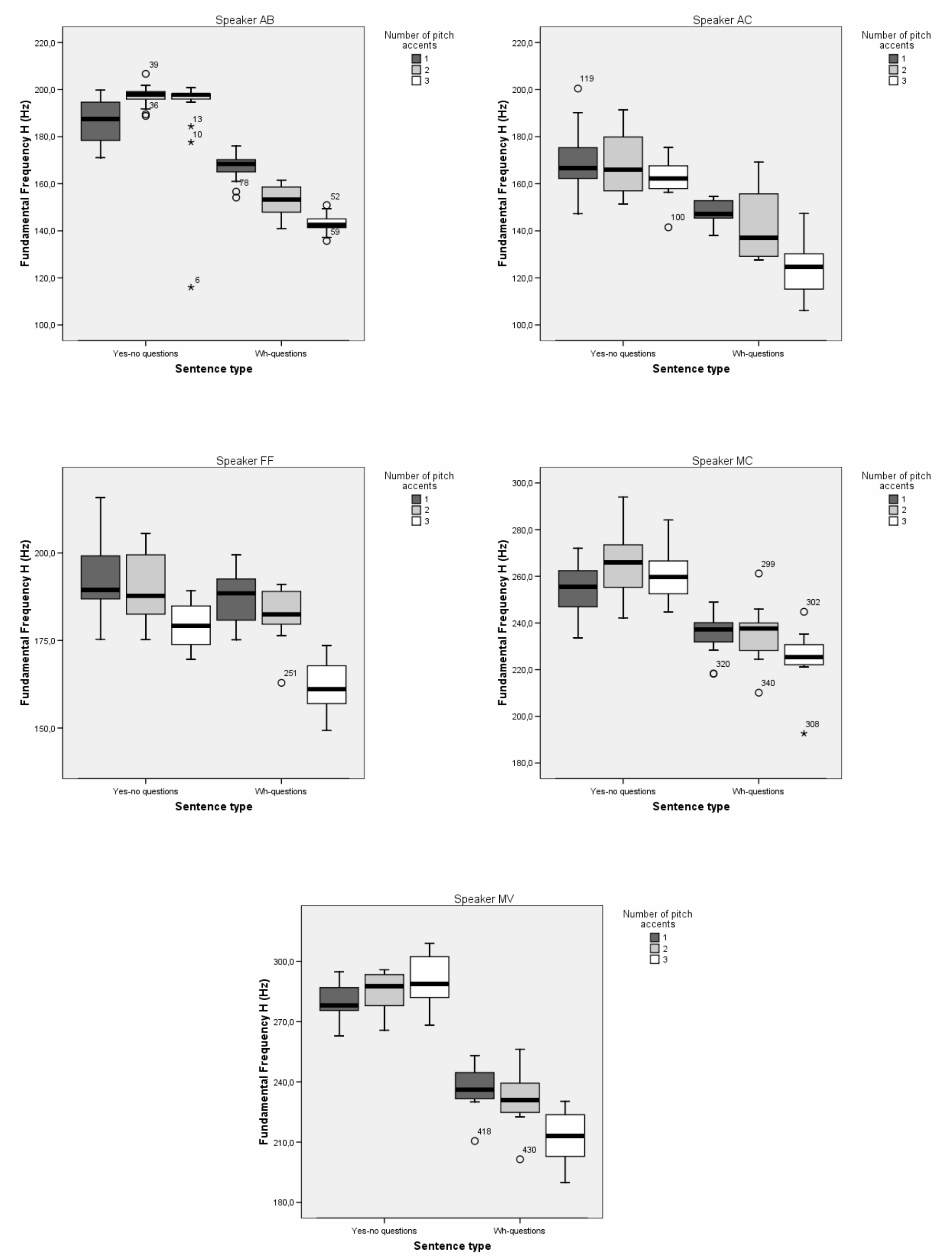


Figure 12. Box plots of H F0 within the nuclear accent H+L* for yes-no questions (left-hand boxes) and wh-questions (right-hand boxes) in utterances of different lengths (1 pitch accent in dark grey, 2 pitch accents in light grey and 3 pitch accents in white) for each speaker.

Table IV shows the results of the Wilcoxon matched pairs signed rank and Friedman tests and confirms that 1) there is a large effect of sentence type on the height of the value of the start of the fall (p=0.000 for all speakers) and, 2) there is an utterance length effect on H pitch value that is statistically significant for all speakers in wh-questions but only for one speaker (FF) in yes-no questions.

F0 start of the fall	Sentence type	Utterance length	
		Yes-no questions	Wh-questions
AB	p=0.000*	p=0.074	p=0.000*
AC	p=0.000*	p=0.257	p=0.005*
FF	p=0.000*	p=0.000*	p=0.000*
MC	p=0.000*	p=0.091	p=0.038
MV	p=0.000*	p=0.057	p=0.000*

Table IV. P-values of the Wilcoxon matched pairs signed rank test (testing the effect of sentence type on H F0 value within the nuclear accent) and Friedman test (testing the effect of utterance length on H F0 value) for each speaker. *Statistically significant at 0.05 level.

Our results clearly show two points, in particular: firstly, there is an unmistakable presence of a super-high variant of the high leading tone within the nuclear accent H+L* in yes-no questions; and secondly, the pitch height of the leading tone in yes-no questions is less affected by utterance length for all speakers except FF.

2.2.2.2. End of the fall or L tone

Figure 13 shows the box plots of F0 for the end of the fall within the nuclear accent H+L* in utterances of different lengths (1 to 3 pitch accents) for yes-no questions (left-hand boxes) and wh-questions (right-hand boxes). It seems that there is a small effect of sentence type on the L F0 value, with the value being higher for yes-no questions and lower for wh-questions, but this tendency is not constant across the subjects, as we can see in the five graphs in Figure 14. Nor is there present a constant utterance-length effect in the L F0 value. Nevertheless, it seems that utterance length affects the L F0 of wh-questions more than that of yes-no questions: the longer the utterance, the lower the L F0—with the exception of speaker MC, who does not follow this pattern.

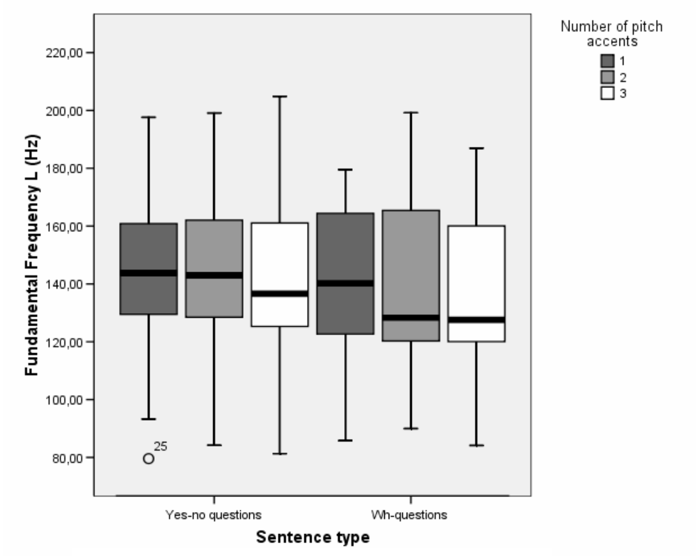
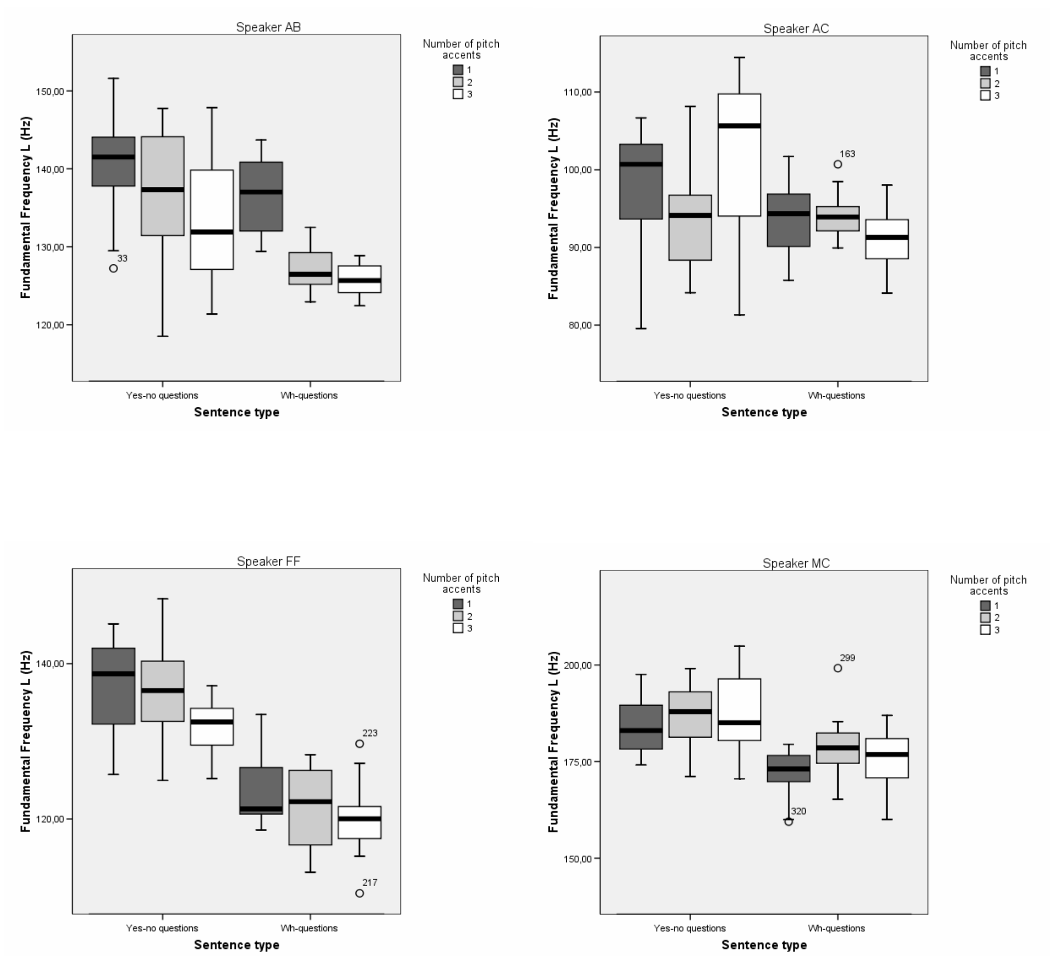


Figure 13. Box plots of L F0 within the nuclear accent H+L* for yes-no questions (left-hand boxes) and wh-questions (right-hand boxes) in utterances of different lengths (1 pitch accent in dark grey, 2 pitch accents in light grey and 3 pitch accents in white) for all speakers.



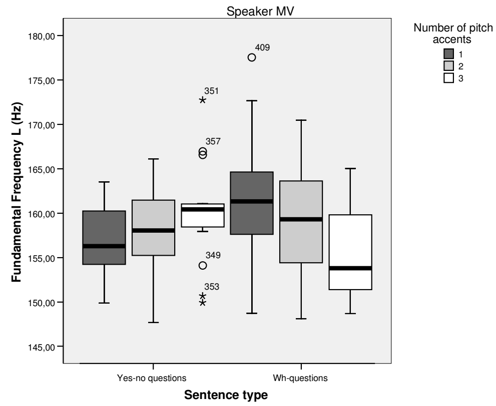


Figure 14. Box plots of L F0 within the nuclear accent H+L* for yes-no questions (left-hand boxes) and wh-questions (right-hand boxes) in utterances of different lengths (1 pitch accent in dark grey, 2 pitch accents in light grey and 3 pitch accents in white) for each speaker.

Table V shows the results of the Wilcoxon matched pairs signed rank test and reveals that 1) there is a statistically significant effect of sentence type on the L F0 value for all the speakers except speaker MV, and 2) there is an utterance length effect on the L F0 value that is significant only for speaker AC in yes-no questions and for speakers AB, MC and MV in wh-questions.

F0 end of the fall	Sentence type	Utterance length	
		Yes-no questions	Wh-questions
AB	p=0.000*	p=0.127	p=0.000*
AC	p=0.025*	p=0.017*	p=0.459
FF	p=0.000*	p=0.165	p=0.257
MC	p=0.000*	p=0.420	p=0.031*
MV	p=0.879	p=0.549	p=0.005*

Table V. P-values of the Wilcoxon matched pairs signed rank test (testing the effect of sentence type on the L F0 value within the nuclear accent) and Friedman test (testing the effect of utterance length on the L F0 value) for each speaker. *Statistically significant at 0.05 level.

If we compare these results with the results given in 2.2.2.1, we note two things. First, the L point is less affected by sentence type than the H point. Secondly, generally speaking it can be said that the effect of utterance length on the L F0 point in yes-no questions is nearly null, and in wh-questions this effect is not as constant across speakers and as pronounced as it is on the H F0 point.

2.2.2.3. Duration of the nuclear stressed syllable

Figure 15 shows the box plots of duration of the nuclear stressed syllable for yes-no questions (left-hand boxes) and wh-questions (right-hand boxes) for all speakers. The graph reveals no difference in duration between the two sentence-types, except for speaker AB. During the recording, this speaker was the one most given to an unnaturally marked reading. This difference in duration may be a result of that.

Results from the Wilcoxon matched pairs signed rank test confirmed that there is no difference in the duration of the nuclear stressed syllable between sentence types for all speakers except AB (p=0.035).

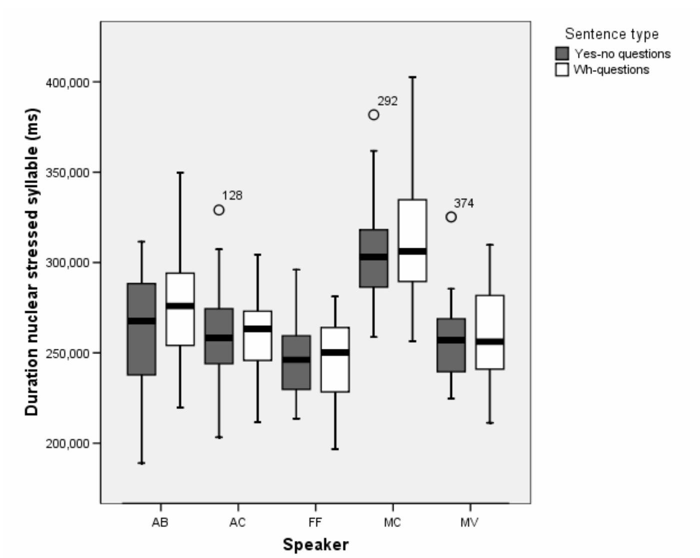


Figure 15. Box plots of duration of the nuclear stressed syllable in yes-no questions (grey boxes) and wh-questions (white boxes) for all speakers.

2.3. Discussion

This production experiment shows interesting results relating to the different acoustic properties that characterize the particles *que* and *què* that introduce yes-no questions and wh-questions respectively in Majorcan Catalan and the falling nuclear accent that characterizes these two different types of questions. Regarding interrogative particles, measures of F0, vowel quality and duration were taken for the vowel of the particles. The results revealed differences in F0 and vowel quality (F1) and optional differences in duration between the vowels of the stressed/accented particle *què* ('what') and unstressed/unaccented particle *que* ('that'). Thus, the stressed/accented interrogative particle has a higher mean pitch than the stressed/unaccented particle. As emphasized above, it is not possible to distinguish the effects of stress from those of accent purely on the basis of this test material. Therefore, it is on the basis of other studies (Astruc and Prieto, 2006; Prieto and Ortega-Llebaria, 2006) that we can claim that this difference in pitch is due to the presence of accent. In wh-questions evidence was found for significant anticipatory raising for three of the five speakers and in yes-no questions for two of the five speakers. Anticipatory raising occurs because speakers anticipate the length of the utterance by beginning higher when the utterance is longer and lower if the utterance is shorter. This allows us to see that a speaker would plan the height of the initial F0 peak (in wh-questions) but also the height of the utterance-initial F0 values (in yes-no questions) depending on the utterance length or number of pitch accents in the utterance. This raises the question of whether speakers plan F0 contours at a phrase level (i.e. utterance length is a determining factor in initial pitch height) or at a more local level (they proceed accent-by-accent). Our results follow previous findings in reporting that speakers begin more or less high depending on the utterance length (Hart J., 1979; Cooper and Sorensen, 1981; Rialland, 2001). By contrast, other researchers have reported no utterance length effect on peak values (Thorsen, 1982; Liberman and Pierrehumbert, 1984; van den Berg et al., 1992; Prieto et al., 1996). These contradictory results can be applied to tonal languages, in which evidence has been found for correlation between utterance length and height of the initial peak in Snider (1998) and Rialland (2001), but not in Lindau (1986), Connell (2003) and Connell (2004). However, the fact that interaction of utterance length on initial F0 values is statistically significant for only three of the five speakers in wh-questions and for two speakers in yes-no questions shows that global preplanning has to be interpreted as soft preplanning (Liberman and Pierrehumbert, 1984), that is, an option that speakers may use but that is not essential as a production mechanism. In Liberman and Pierrehumbert's terms, hard preplanning stands for "processing that is an essential part of intending to say something and that normally needs to be accomplished before executing that intention", while soft preplanning means "the sorts of preparation that a speaker may freely choose to make, out of rational calculation, ritual observance, or any other cause, and that might well be omitted for a linguistically equivalent utterance under other circumstances". This optional character that seems to be involved with soft preplanning, expressed also in speaker variation within languages in Prieto et al. (2006), is what might have caused the contradictory results that mentioned above.

Optional differences in duration between the vowel of stressed/accented and unstressed/unaccented interrogative particles were found. That is, the vowel is longer when the interrogative particle is accented and shorter when it is unaccented. This effect is statistically significant for only three of the five speakers. On the basis of other studies (Astruc and Prieto, 2006; Prieto and Ortega-Llebaria, 2006), we can infer that this difference in duration is triggered by the presence or absence of stress.

An increment in F1 value but not in F2 value was found in the vowel of the accented interrogative particle. This difference in F1 seems to be an acoustic correlate of stress differences and seems to be caused by a difference in the degree of openness of the vocal tract, it being more fully open in the case of the stressed vowel than in the case of the unstressed vowel. Evidence for differences in vowel quality brought about by varying openness of the vowel tract as an articulatory correlate of stress is presented in Beckman and Edwards (1991).

Effects of sentence type and utterance length on pitch height of the leading tone H of the nuclear accent H+L* are significant across speakers. Values of the F0 peak are higher for yes-no questions than for wh-questions. This causes us to consider the unmistakable presence of a super-high variant of the high leading tone within the nuclear accent H+L* in yes-no questions. The interaction between length and H value of the nuclear accent, with the F0 H value being much lower as the number of pitch accents increases, confirms what the results regarding the effect of utterance length on the F0 of the vowel of the interrogative particles suggested, namely, the existence of downstep. This interaction between length and H is higher in wh-questions than in yes-no questions. It is not only the presence of the super-high variant of the leading tone H but also a sort of blocking of the downtrend that could be interpreted as contrast enhancement between yes-no questions and wh-questions in Majorcan Catalan. These strategies would facilitate the perception of L tone in yes-no questions; otherwise the L tone might be misinterpreted from a downstepped or plain H.

By contrast, the L point is less affected by sentence type than the H point. That is what would be expected if the purpose of the upstepped H leading tone is to clarify and to disambiguate yes-no questions from wh-questions. The effect of utterance length on the L F0 point is small and not significant across all speakers and sentence types; thus, it is nearly null in yes-no questions and lower than the H F0 point in wh-questions.

Finally, our results revealed no difference in duration between the two sentence-types in terms of the nuclear stressed syllable. Only one subject produced longer nuclear syllables in yes-no question than in wh-questions. However, this speaker displayed a tendency to read very markedly with a very unnatural vocalization. This suggests that the results for this speaker should be interpreted with caution.

3. Perception experiment

This chapter reports a perception experiment based on the Categorical Perception (CP) paradigm (CP) which tests whether listeners make categorical linguistic use of F0 scaling differences in the height of the pretonic of the falling nuclear accent in perceiving yes-no questions as opposed to wh-questions in Majorcan Catalan. The application of the CP paradigm has been carried out bearing in mind the clear difference in degree of categorical perception that seems to exist between segmental and intonational contrasts and the conclusions which previous studies have reached about the suitability of applying the CP paradigm to intonational differences as well as alternative or additional measurement tools. Furthermore, there have been considered the two criteria that the results of the application of the CP paradigm should fulfill in order to be considered categorical, namely, that identification proportions should predict discrimination accuracy and that discrimination peaks should correspond to the location of the category boundaries determined by identification. Thus, in the identification task, in addition to response frequencies mean Reaction Times (RT) will be measured. The obtained and predicted discrimination functions will be compared through statistical analyses. Results will be analyzed according to the subjects' gender and musical background in order to see if these factors play any intervening role.

3.1. Method

There have been several different experimental approaches to address the question of whether variation in pitch range or pitch alignment is a phonologic or a phonetic contrast (Gussenhoven, 2004). The first method, the imitation task, was applied to investigate the contrast between L*+H and H*+L in American English (Pierrehumbert and Steele, 1989). Subjects were asked to reproduce stimuli from a continuum in which the timing of the peak was varied systematically. The results showed the inability of the subjects to do so. Instead, subjects grouped their productions within one of the two intonational categories. The contrast was therefore discrete. Dilley and Brown (to appear) also used an imitation study to investigate the importance of relative F0 level differences for phonological distinctions. Participants heard synthetic versions of the sentence *Some lemonade* where the F0 contours across *le*- and *mon*- were replaced with flat F0 patterns, thus removing within-syllable cues to F0 extremum type and timing. The F0 levels of *le*- and *mon*- were manipulated towards one another. Participants produced F0 peaks and valleys on syllables which had flat F0 patterns in the stimuli. They concluded that the relative F0 levels of successive syllables are more important to intonational phonological representations than F0 shape attributes on individual syllables. Redi (2003) applied this method to a synthetic speech continuum in which a F0 maximum and minimum were shifted in 25 ms through a target two-syllable sequence with a WS or a SW stress pattern. She concluded that there is a categorical effect for this particular acoustic continuum and claimed the distinction between the H+L* and H* accents proposed by the AM model. The second method, the pitch-range task, was adopted to address the difference between the high and the low rises in Dutch (Gussenhoven and Rietveld, 2000). Listeners had to rate stimuli with different pitch ranges on a number of semantic scales whose meaning varied with pitch range. The results showed that the high and the low rises present discretely different contours. The third method, the semantic task, is closely related to the fourth one. According to Post (2000), there are three different ways in which the subject can interpret an intonation contour: s/he can rate it on a semantic scale (Uldall, 1964), either as a yes/no decision about the appropriateness of a given meaning (Bartels and Kingston, 1994), or as a judgment of acceptability in a particular context (Caspers, 1998). According to Post (2000), this method does not really ensure that a given contrast is discrete, since gradient variation can also be meaningful.

In the fourth method, the Categorical Perception paradigm, subjects perform two tasks: an identification task and a discrimination task. In the identification task, participants listen to randomly ordered stimuli constructed from a continuum and judge which of two categories each stimulus represents. In the discrimination task, the participants listen to the stimuli again, but this time they are asked to identify the test stimulus in terms of a paired reference stimulus. In the pair-wise AX task, the participants hear the test stimulus and a single reference stimulus, and decide whether the two stimuli are two instances of the same stimulus or are different stimuli. The patterns of results expected are shown in Figure 1. The functions of responses to the identification task have an S-shape, i.e. an abrupt shift from one category to the other rather than a gradual transition. Figure 1 also shows the idealized discrimination function (dashed line). If perception is categorical, discrimination is easier when it straddles the boundary between the categories than within categories (i.e. between stimuli that have been identified as members of 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). Discrimination results are predicted through a formula taken originally from Liberman et al. (1957) and the boundary between categories in sigmoid response curves are calculated by probit analysis (Finney, 1971).

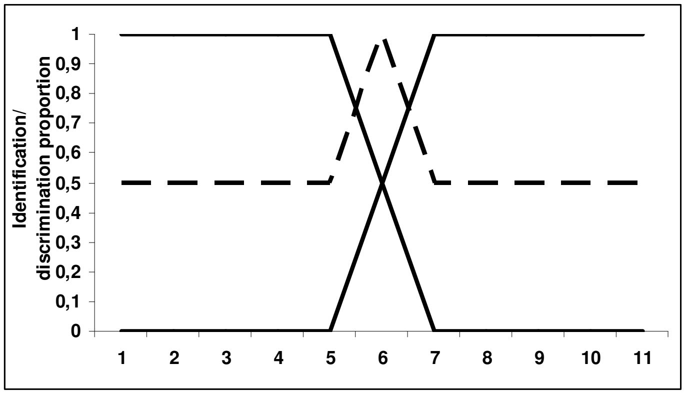


Figure 1. Idealized identification and discrimination functions.

3.1.1. Stimuli

One token of the yes-no question *Que l'hi duries?* ('Would you take it to him/her?') and one token of the wh-question *Què li duries?* ('What would you take him/her?') were produced by a 24-year-old native female speaker of Majorcan Catalan. The two tokens are homophonic at the segmental level. In the yes-no question token the leading tone was 263 Hz while in the wh-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 with the onset of the syllable DU, H with the offset of the vowel of the syllable DU, and L2 with the offset of the vowel of the syllable RI. From these two base tokens, ten stimuli were created by means of PSOLA synthesis. Five stimuli were created by shifting the peak downwards from the yes-no question token in four steps of 15 Hz each (Figure 2, left panel) and another five stimuli by shifting the peak upwards from the wh-question token in like fashion (Figure 2, right panel).

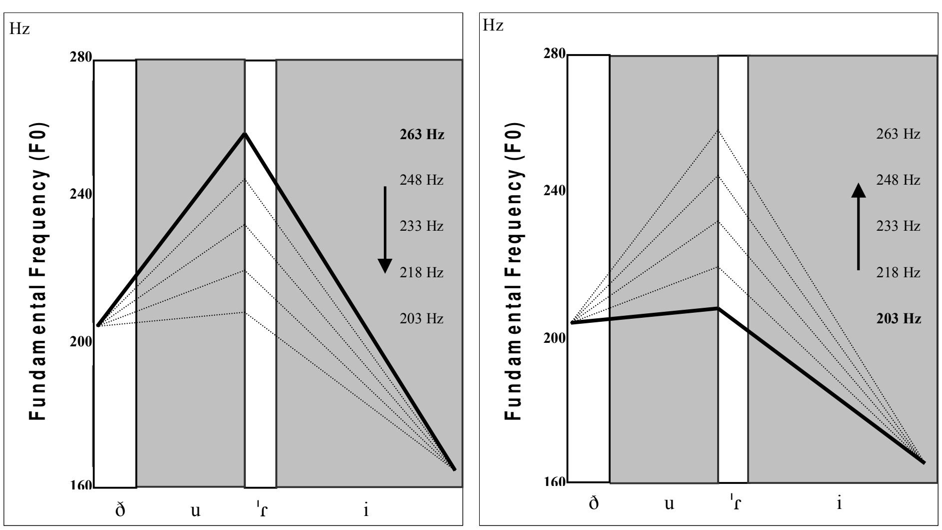


Figure 2. Schematized creation of the stimuli from yes-no question base stimulus (left panel) and wh-question base stimulus (right panel).

3.1.2. Tasks and experimental procedures

Subjects were seated at a laptop in a quiet room and the stimuli were played back through headphones. The perception test was played by means of PERCEVAL, software for performing computerized auditory and visual perception experiments, which also records the RTs. As we were interested in RT measurements, the listeners were instructed to maintain 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 blocks. The test lasted approximately 30 minutes.

3.1.2.1. Identification task

The materials for the identification task consisted of 5 repetitions of each of the 10 stimuli (five stimuli from the yes-no question base stimulus and five stimuli from the wh-question base stimulus). These 50 stimuli (10 stimuli x 5 repetitions) were presented in blocks of 10 in random order. The first block constituted the practice block. 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 wh-question *Què li duries* ('What would you bring him/her?'), they were to press the "A" key (for *Això*= "That").

3.1.2.2. 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 (four from the yes-no question base stimulus and four from the wh-questions base stimulus): step 1 + step 2, step 2 + step 3, etc. Eight pairs of stimuli were created in BA order (four from yes-no question base stimulus and four from wh-questions base stimulus): step 2 + 1, step 3 + 2, etc. Additionally, 10 control AA pairs were created which contained two identical stimuli: step 1 + step 1, step 2 + step 2, etc. Two repetitions of these stimuli were randomized. The practice block before the test session was based on the yes-no question base stimulus for half the listeners and on the wh-question base stimulus for the other half. Thus each subject heard a total of 65 stimuli (4 AB stimuli + 4 BA stimuli + 5 AA stimuli x 2 base stimuli x 2 repetitions + one practice block of 13 stimuli—see Table I). Subjects were asked to decide whether they heard the pair of stimuli as the same or different. If the two stimuli sounded the same, they were to press the "I" key on the keyboard (*igual*—same) and if the stimuli sounded different, they were to press the "D" key (*diferent*—different). The interval between the two stimuli in each pair was 0.5 seconds. The order of the blocks was counterbalanced, that is, half of the listeners, whose practice block was based on the yes-no question base stimulus, started with the block based on the yes-no question base stimulus and went on to the block based on the wh-question base stimulus and alternated thus until they had listened to the four test blocks; the other half of the listeners heard a practice block based on the wh-question and thereafter heard the test blocks in the reverse order relative to the first group of subjects.

	Yes-no question base stimulus	Wh-question base stimulus
AB (1+2, 2+3, 3+4, 4+5)	4 pairs	4 pairs
BA (2+1, 3+2, 4+3, 5+4)	4 pairs	4 pairs
AA (1+1, 2+2, 3+3, 4+4, 5+5)	5 pairs	5 pairs
	13 pairs of stimuli	13 pairs of stimuli
	26 pairs of stimuli x 2 repetitions + 13 pairs of practice stimuli= 65 stimuli (presented in blocks of 13 in random order)	

Table I. Stimuli for the discrimination task.

3.1.3. Subjects

Forty-two native speakers of Majorcan Catalan (twenty-five females and seventeen males)[4], 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 was analyzed. 11 of these subjects had had basic musical training, 11 had had long-term formal training (they had a degree in music) and 10 had no musical training.[5]

[4] Since gender differences have been reported in production and perception experiments (Jensen and Carlin, 1981; Johnson et al., 1999; Rogers, 2003), we wanted to see if this factor played any intervening role.

[5] As some studies seem to show differences in tonal perception accuracy depending on the musical training of listeners (Glenn Schellenberg, 2002; Cummins et al., 2005), we wanted to compare our data with these previous findings.

3.1.4. Statistical analyses

We used non-parametric tests because general assumptions regarding the data for using parametric tests were not fulfilled, that is: 1) some variables in our data set, including the dependent variable, were measured in a nominal scale; 2) some variables in our data set, including the dependent variable, violated the normality assumption; and 3) the sample size was not very large (32 subjects).

The following statistical tests were used (see Blalock, 1979 for an explanation of non-parametric statistics):

–Wilcoxon matched pairs signed rank test. This test was used to compare the results between two groups, i.e. for adjacent stimuli in the identification function to determine the boundary shift. Note that probit analysis was not used to determine the boundary shift; instead, the Wilcoxon matched pairs signed rank test was used to determine whether adjacent stimuli were significantly different, on the assumption that the boundary between categories would be located between the two adjacent stimuli which were perceived to be most different.

–Mann-Whitney U test. This test was used, for instance, to compare the identification performance of men and women for the same stimulus.

–Friedman test. This test was used to compare broadly different functions, i.e. for the different functions (AA, AB and BA) in the discrimination task.

Since in most cases the same test was repeated in many subsamples, the Bonferroni correction was applied by adjusting the p-values multiplied by the number of contrasts performed. However, as the application of the Bonferroni correction did not change the significance of the results, the adjusted p-values have not been reported in this study.

3.2. Results

3.2.1. Identification results

3.2.1.1. General identification results

Figure 3 shows the proportion of yes-no question responses for the two continua created from the yes-no question (in black) and wh-question (in grey) base stimuli. As we can observe, the function presents an S-shape. The frequency range from 203 Hz to 218 Hz seems to correspond to the "wh-question" category and the range from 248 Hz to 263 Hz, to the "yes-no question" category. Thus, the transition between the two categories would correspond to the range between 218 Hz and 248 Hz.

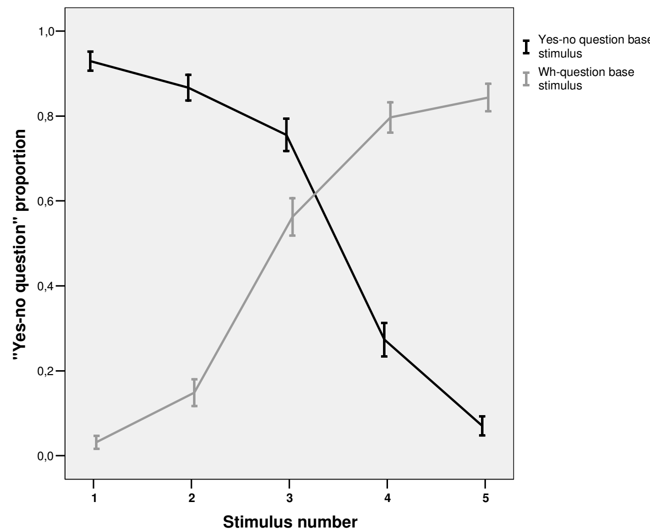


Figure 3. Identification function of the synthesized continuum ranging from 263 to 203 Hz (continuum from the yes-no question base stimulus) and from 203 to 263 Hz (continuum from the wh-question base stimulus). Error bars represent the standard error of the mean.

Table II reports the standard error of the mean for every stimulus. Stimulus 1 and stimulus 5 for both base stimuli have lower standard errors (shaded numbers). That 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 lies in the middle of the frontier region between the two categories, has a higher standard error, indicating less agreement among listeners. It is worth noting that the closer the stimuli are to the shift between the categories, the higher the standard error they present, (i.e. listeners hesitate when judging these stimuli).

	Stimulus 1	Stimulus 2	Stimulus 3	Stimulus 4	Stimulus 5
Yes-no question base stimulus	0.023	0.030	0.038	0.040	0.023
Wh-question base stimulus	0.015	0.032	0.044	0.036	0.032

Table II. Standard error of the mean for every stimulus from yes-no question base stimulus and wh-question base stimulus.

Table III shows the results of the Wilcoxon matched pairs signed rank test that tests whether there are significant differences between the responses to the adjacent stimuli. According to the CP paradigm, there should not be significant differences between the responses to stimuli pairs

1-2 and 4-5 because they form part of the same category, but there should be between stimulus 3 and the other stimuli, as stimulus 3 represents the boundary between categories. As can be seen in Table IV, the results are nearly what we would expect: stimulus 3 is significantly different from stimulus 2 and stimulus 4 because it is between the two categories, but adjacent stimuli 4-5 (created from the yes-no question base simulus) and 1-2 (created from the wh-question base stimulus) were perceived to be rather more different than would be expected (shaded p-values). Before trying to provide possible explanations, it is important to compare how men and women, and non-musicians, occasional musicians and qualified musicians performed the task.

	Stimuli 1-2	Stimuli 2-3	Stimuli 3-4	Stimuli 4-5
Yes-no question base stimulus	p= 0.046*	p= 0.023*	p= 0.000*	p= 0.000*
Wh-question base stimulus	p= 0.001*	p= 0.000*	p= 0.000*	p= 0.134

Table III. P-values of the Wilcoxon matched pairs signed rank test comparing adjacent stimuli for continuum created from yes-no question base stimulus and for continuum created from wh-question base stimulus. *Statistically significant at 0.05 level.

3.2.1.2. Gender differences

Figure 4 represents the men's performance of identification task and Figure 5 the women's performance. Though gender differences have been reported in production and perception experiments (Jensen and Carlin, 1981; Johnson et al., 1999; Rogers, 2003), it seems that there are no great significant differences in identification accuracy in our own results.

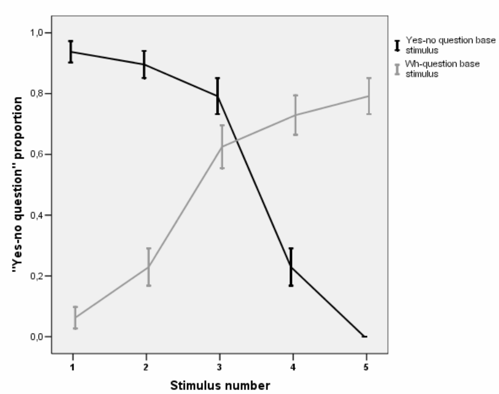


Figure 4. Male listeners' identification functions. Error bars represent the standard error of the mean.

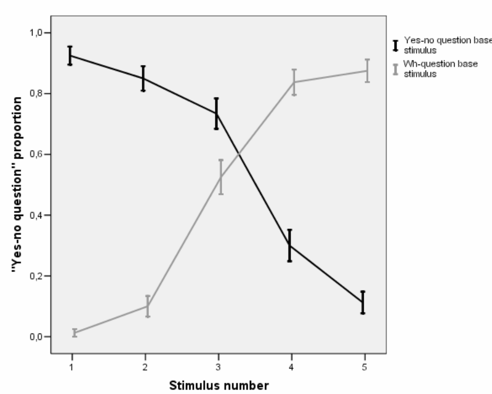


Figure 5. Female listeners' identification functions. Error bars represent the standard error of the mean.

The Mann-Whitney U test, which assesses the significance of the differences in two independent samples, reveals in general no significant differences in the identification of men and women for each stimulus (Table IV). Note, however, that for stimulus 5 created from yes-no question base stimulus, the difference between men and women is significant (shaded p-value).

	Stimulus 1	Stimulus 2	Stimulus 3	Stimulus 4	Stimulus 5
Yes-no question base stimulus	p= 0.790	p= 0.461	p= 0.265	p= 0.386	p= 0.016*
Wh-question base stimulus	p= 0.117	p= 0.047*	p= 0.271	p= 0.142	p= 0.211

Table IV. P-values of the Mann-Whitney U test comparing men and women identification performance for every stimulus created from yes-no question and wh-question stimuli. *Statistically significant at 0.05 level.

In spite of not finding sex differences regarding identification accuracy, we find a small difference with respect to variability between men and women. The results for women display slightly less variability than those for men. Table V plots the standard error of the mean for both men and women.

	Stimulus 1		Stimulus 2		Stimulus 3		Stimulus 4		Stimulus 5	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
Yes-no question base stimulus	0.035	0.030	0.045	0.040	0.059	0.050	0.061	0.052	0.000	0.036
Wh-question base stimulus	0.035	0.013	0.061	0.034	0.071	0.056	0.065	0.042	0.059	0.037

Table V. Standard error of the mean for every stimulus from yes-no question and wh-question stimuli, broken down by gender.

In table VI (for men) and table VII (for women) the results of the respective Wilcoxon matched pairs signed rank tests are presented. These results follow the same pattern as the general results, that is, although stimuli pairs 4-5 (created from yes-no question base stimulus) and 1-2 (created from the wh-question base stimulus) should be identified as forming part of the same category, there are significant differences (shaded values) between the perception of both stimuli sets for men and women.

MEN	Stimuli 1-2	Stimuli 2-3	Stimuli 3-4	Stimuli 4-5
Yes-no question base stimulus	p=0.317	p=0.248	p=0.000*	p=0.001*
Wh-question base stimulus	p=0.021*	p=0.000*	p=0.225	p=0.317

Table VI. P-values of the Wilcoxon matched pairs signed rank test comparing adjacent stimuli for continua created from yes-no question and wh-question base stimuli. Male subjects. *Statistically significant at 0.05 level.

WOMEN	Stimuli 1-2	Stimuli 2-3	Stimuli 3-4	Stimuli 4-5
Yes-no question base stimulus	p=0.083	p=0.050*	p=0.000*	p=0.003*
Wh-question base stimulus	p=0.008*	p=0.000*	p=0.000*	p=0.257

Table VII. P-values of the Wilcoxon matched pairs signed rank test comparing adjacent stimuli for continua created from yes-no question and wh-question base stimuli. Female subjects. *Statistically significant at 0.05 level.

3.2.1.3. Differences between musicians and non-musicians

Some studies have reported differences in the accuracy of perception depending on the degree of musical training of the subjects (Cummins et al., 2006; Glenn Schellenberg, 2002). Figures 6, 7, and 8 represent the results of the identification task according to whether the subjects were non-musicians, occasional musicians or qualified musicians.

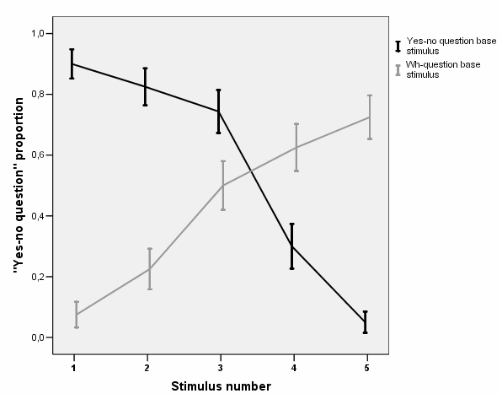


Figure 6. Identification functions for non-musician subjects. Error bars represent the standard error of the mean.

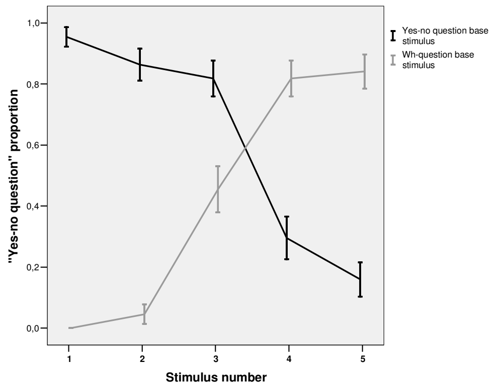


Figure 7. Identification function for occasional musician subjects. Error bars represent the standard error of the mean.

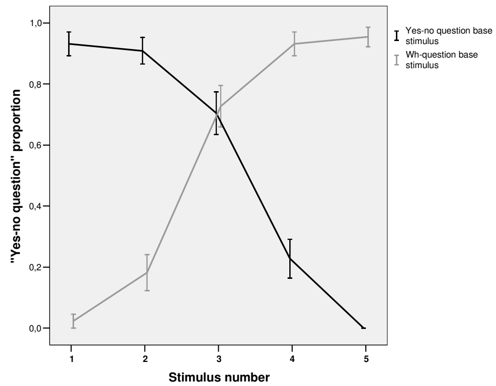


Figure 8. Identification function for musician subjects. Error bars represent the standard error of the mean.

As in the case of differences between men and women, the Mann-Whitney U test reveals no overall significant differences in identification among these three groups (Tables VIII, IX and X). Observe that, when these differences exist (shades values), they occur only in the continuum created from the wh-question base stimulus.

	Stimulus 1	Stimulus 2	Stimulus 3	Stimulus 4	Stimulus 5
	N-O	N-O	N-O	N-O	N-O
Yes-no question base stimulus	p= 0.321	p= 0.596	p= 0.976	p= 0.911	p= 0.117
Wh-question base stimulus	p= 0.063	p= 0.014*	p= 0.760	p= 0.042*	p= 0.286

Table VIII. P-values of the Mann-Whitney U test comparing the identification performance of non-musicians (N) and occasional musicians (O) for every stimulus created from yes-no question and wh-question base stimuli. *Statistically significant at 0.05 level.

	Stimulus 1	Stimulus 2	Stimulus 3	Stimulus 4	Stimulus 5
	O-Q	O-Q	O-Q	O-Q	O-Q
Yes-no question base stimulus	p= 0.610	p= 0.554	p= 0.540	p= 0.550	p= 0.007*
Wh-question base stimulus	p= 0.306	p= 0.037*	p= 0.016*	p= 0.128	p= 0.017*

Table IX. P-values of the Mann-Whitney U test comparing the identification performance of occasional (O) and qualified musicians (Q) for every stimulus created from yes-no question and wh-question base stimuli. *Statistically significant at 0.05 level.

	Stimulus 1	Stimulus 2	Stimulus 3	Stimulus 4	Stimulus 5
	Q-N	Q-N	Q-N	Q-N	Q-N
Yes-no question base stimulus	p= 0.623	p= 0.274	p= 0.574	p= 0.489	p= 0.140
Wh-question base stimulus	p= 0.274	p= 0.662	p= 0.040*	p= 0.001*	p= 0.001*

Table X. P-values of the Mann-Whitney U test comparing the identification performance of qualified musicians (Q) and non-musicians (N) for every stimulus created from yes-no question and wh-question base stimuli. *Statistically significant at 0.05 level.

Nevertheless, a slight difference was found in the standard error of the mean (Table XI) for every stimulus depending on whether subjects were non-musicians, occasional musicians or qualified musicians. As we can see, the more musical training subjects have, the less variability in their responses.

	Stimulus 1			Stimulus 2			Stimulus 3			Stimulus 4			Stimulus 5		
	NM	OM	QM	NM	OM	QM	NM	OM	QM	NM	OM	QM	NM	OM	QM
Yes-no question base stimulus	0.048	0.032	0.038	0.061	0.052	0.044	0.071	0.059	0.070	0.073	0.070	0.064	0.035	0.056	0
Wh-question base stimulus	0.042	0	0.023	0.067	0.032	0.059	0.080	0.076	0.068	0.078	0.059	0.038	0.071	0.056	0.032

Table XI. Standard error of the mean for every stimulus from yes-no question and wh-question stimuli for non-musicians (NM), occasional musicians (OM) and qualified musicians (QM).

Tables XII, XIII and XIV display the results of the Wilcoxon matched pairs signed rank test. As noted, according to the CP paradigm, there should be no significant differences between the responses to stimuli pairs 1-2 and 4-5 because they form part of the same category whereas there should be between stimulus 3 and the other stimuli, as stimulus 3 represents the boundary between categories. Generally speaking, it can be said that stimuli pairs 1-2 and 4-5 are not perceived as significantly different (except by qualified musician subjects) and when they are, the difference is less significant compared to the difference between stimuli 3 and the other stimuli.

On the other hand, the results of the Wilcoxon matched pairs singed rank test (Tables XII, XIII and XIV) suggest that the boundary between the two categories may be located at the range of values between 218 Hz and 233 Hz because stimuli located at these frequency values were the most differently perceived for both continua. Thus, stimulus 3 (233 Hz) is perceived more different from stimulus 4 (218 Hz) than stimulus 2 (248 Hz) when it is derived from a yes-no question base stimulus, and more different from stimulus 2 (218 Hz) than stimulus 4 (248 Hz) when it is derived from a wh-question base stimulus. This can be also seen but to a lesser extent in the general identification results (Table III).

NON-MUSICIANS	Stimuli 1-2	Stimuli 2-3	Stimuli 3-4	Stimuli 4-5
Yes-no question base stimulus	p= 0.083	p= 0.564	p= 0.000*	p= 0.004*
Wh-question base stimulus	p= 0.058	p= 0.002*	p= 0.132	p= 0.102

Table XII. Non-musician subjects' p-values from the Wilcoxon matched pairs signed rank test comparing adjacent stimuli for continua created from yes-no question and wh-question base stimuli. *Statistically significant at 0.05 level.

OCCASIONAL MUSICIANS	Stimuli 1-2	Stimuli 2-3	Stimuli 3-4	Stimuli 4-5
Yes-no question base stimulus	p= 0.157	p= 0.405	p= 0.000*	p= 0.109
Wh-question base stimulus	p= 0.157	p= 0.000*	p= 0.001*	p= 0.655

Table XIII. Occasional musician subjects' p-values from the Wilcoxon matched pairs signed rank test comparing adjacent stimuli for continua created from yes-no question and wh-question base stimuli. *Statistically significant at 0.05 level.

QUALIFIED MUSICIANS	Stimuli 1-2	Stimuli 2-3	Stimuli 3-4	Stimuli 4-5
Yes-no question base stimulus	p= 0.655	p= 0.013*	p= 0.000*	p= 0.002*
Wh-question base stimulus	p= 0.008*	p= 0.000*	p= 0.020*	p= 0.655

Table XIV. Qualified musician subjects' p-values from the Wilcoxon matched pairs signed rank test comparing adjacent stimuli for continua created from yes-no question and wh-question base stimuli. *Statistically significant at 0.05 level.

Note that when they are created from the yes-no question base stimulus, stimuli 3 and 4 are identified with a high rate of "yes-no question" responses (Figure 3). The results analyzed according to the subjects' gender and musical background also show that stimuli 3 and 4 follow the same tendency. When they are created from the yes-no question base stimulus, they are identified more often as “yes-no question” than would be expected, since stimulus 3 is given a proportion of 0.76 “yes-no question” judgments and stimulus 4 a proportion of 0.27 “yes-no question” judgments. By contrast, when these stimuli are created from the wh-question base stimulus, the results are closer to what we would expect, since stimulus 3 receives “yes-no question” and “wh-question” judgments in equal proportions, and stimulus 4 a proportion of around 0.8 “yes-no question” judgments (Figure 3). This phenomenon represents a general tendency that could be due to listeners using a supplementary cue when listening to ambiguous stimuli. Listeners might use the presence or absence of the accented interrogative particle in trying to identify these stimuli. Since in stimuli created from the yes-no question base stimulus there is no accented interrogative particle, subjects would tend to identify such stimuli more as “yes-no question” than “wh-question”. On the other hand, when they hear stimuli created from the wh-question base stimulus, as the interrogative particle is accented, this accentuation would act as a supplementary cue that would permit subjects to identify these stimuli more accurately.

3.2.2. Reaction Time

3.2.2.1. General RT measurements

Figure 9 plots the mean of RT measurements of the peak height continua created from the yes-no question (black bars) and wh-question (grey bars) base stimuli. As in Chen (2003), listeners are faster within categories and slower in the crossover. Observe that while there is a peak in RT measurement for the continuum created from the wh-question base stimulus, in the continuum created from the yes-no question base stimulus, RT measurements of stimuli 3 and 4 are balanced. This is probably related to the results of the identification task, where it has been hypothesized that ambiguous stimuli created from the yes-no question base stimulus could be more difficult to identify, since they lack a supplementary cue for that facilitates rapid identification. Thus, for stimuli created from the yes-no question base stimulus there is a difference of 114 ms between the mean RT for stimuli 3 and 2, and 27 ms between the mean RT for stimuli 3 and 4. On the other hand, for stimuli created from the wh-question base stimulus there is a difference of 349 ms between the RT mean for stimuli 3 and 2 and 312 ms between the RT mean for stimuli 3 and 4.

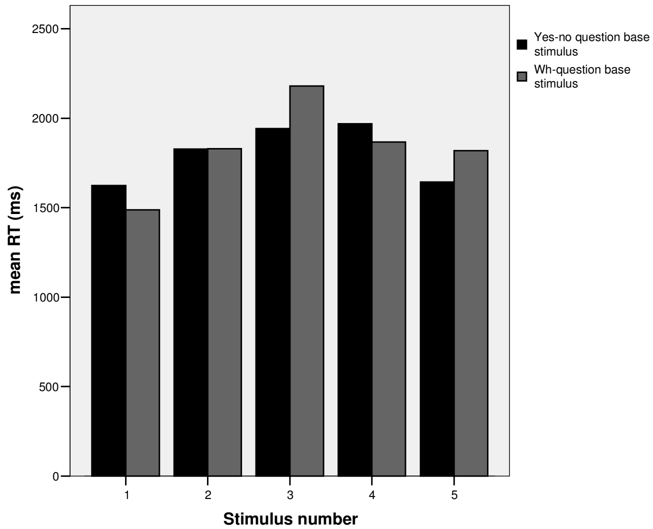


Figure 9. Mean RTs of the peak height for continua created from yes-no question and wh-question base stimuli.

Table XVI shows the results of the Wilcoxon matched pairs signed rank test comparing the difference between RT measurements for adjacent pairs. Judging by these results, for stimuli created from the yes-no question base stimulus, stimuli 2, 3 and 4 are more difficult to identify (all these stimuli require approximately the same time to be identified). On the other hand, for the continuum created from wh-question base stimulus, only one stimulus, stimulus 3, is difficult to identify (there are significant differences between the RT mean for stimulus 3 and adjacent stimuli).

	RT 1-2	RT 2-3	RT 3-4	RT 4-5
Yes-no question base stimulus	p= 0.056	p= 0.183	p= 0.211	p= 0.000*
Wh-question base stimulus	p= 0.000*	p= 0.008*	p= 0.024*	p= 0.071

Table XV. P-values of the Wilcoxon matched pairs signed rank test comparing mean RTs for adjacent stimulus pairs for continua created from the yes-no question and wh-question base stimuli. *Statistically significant at 0.05 level.

3.2.2.2. Gender differences

As we have seen, there do not appear to be great gender differences related to the accuracy of intonation perception; however, we do find gender differences in RT for every stimulus. We can see in Figures 10 and 11 that there are differences in identification speed between men and women, with women faster than men.

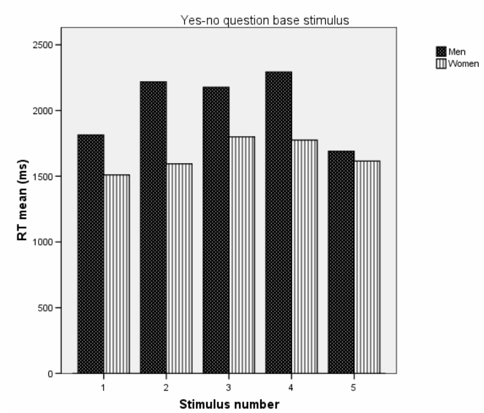


Figure 10. Mean RTs of the peak height for the continuum created from the yes-no question base stimulus for both male and female subjects.

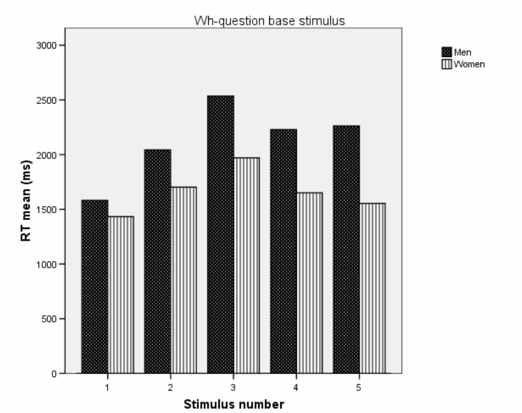


Figure 11. Mean RTs of the peak height for the continuum created from the wh-question base stimulus for male and female subjects.

In general, the Mann-Whitney U test confirms that these differences in the RT measurement between men and women are significant for every stimulus (Table XVI). This difference is more noticeable in the continuum created from the wh-question base stimulus than in the continuum based on the yes-no questions stimulus (Table XVI). Also observe that in the continuum created from the wh-question base stimulus (Figure 11), mean RTs follow a more consistent pattern.

	RT 1	RT 2	RT 3	RT 4	RT 5
Yes-no question base stimulus	p= 0.004*	p= 0.036*	p= 0.063	p= 0.000*	p= 0.227
Wh-question base stimulus	p= 0.239	p= 0.004*	p= 0.001*	p= 0.015*	p= 0.002*

Table XVI. P-values of the Mann-Whitney U test comparing mean RTs of men and women for every stimulus created from the yes-no question and wh-question base stimuli. *Statistically significant at 0.05 level.

3.2.2.3. Differences between musicians and non-musicians

Figures 12 and 13 display the RT measurements for every stimulus for non-musicians, occasional musicians and qualified musicians. Non-musicians are slower than musicians in reacting to all the stimuli except for stimulus 3 where, oddly, qualified musicians are the slowest in both continua. It seems that qualified musicians pay closer attention to ambiguous stimuli.

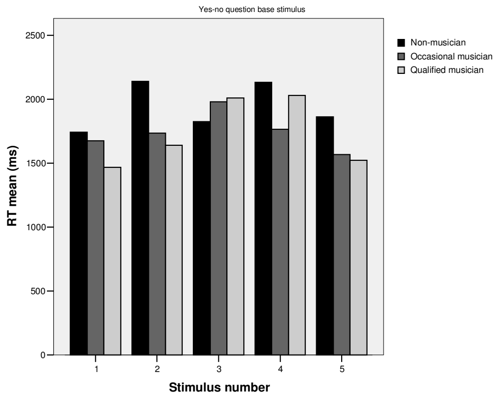


Figure 12. Mean RTs of the peak height for the continuum created from the yes-no question base stimulus for non-musicians, occasional musicians and qualified musicians.

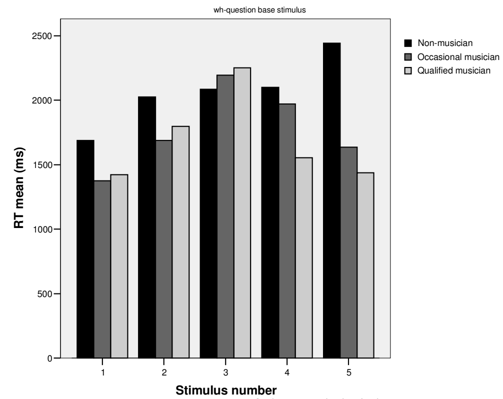


Figure 13. Mean RTs of the peak height for the continuum created from the wh-question base stimulus for non-musicians, occasional musicians and qualified musicians.

In particular, the Mann-Whitney U test reveals significant differences in the RT measurements between non-musician and musicians (whether occasional or qualified) for every stimulus (Tables XVII, XVIII and XIX) except stimulus 3. Interestingly, RT for stimulus 4 created from the yes-no question base stimulus again reveals special behavior. Similar to stimulus 3, qualified musicians required approximately the same time to identify this stimulus as non-musicians did (p=0.248).

	RT 1	RT 2	RT 3	RT 4	RT 5
	N-O	N-O	N-O	N-O	N-O
Yes-no question base stimulus	p= 0.041*	p= 0.062	p= 0.751	p= 0.057	p= 0.275
Wh-question base stimulus	p= 0.035*	p= 0.021*	p= 0.511	p= 0.914	p= 0.015*

Table XVII. P-values of the Mann-Whitney U test comparing mean RTs of non-musicians (N) and occasional musicians (O) for every stimulus created from the yes-no question and wh-question base stimuli. *Statistically significant at 0.05 level.

	RT 1	RT 2	RT 3	RT 4	RT 5
	O-Q	O-Q	O-Q	O-Q	O-Q
Yes-no question base stimulus	p= 0.276	p= 0.409	p= 0.638	p= 0.608	p= 0.664
Wh-question base stimulus	p= 0.809	p= 0.512	p= 0.499	p= 0.045*	p= 0.590

Table XVIII. P-values of the Mann-Whitney U test comparing mean RTs of occasional musicians (O) and qualified musicians (Q) for every stimulus created from the yes-no question and wh-question base stimuli. *Statistically significant at 0.05 level.

	RT 1	RT 2	RT 3	RT 4	RT 5
	Q-N	Q-N	Q-N	Q-N	Q-N
Yes-no question base stimulus	p= 0.069	p= 0.012*	p= 0.953	p= 0.248	p= 0.088
Wh-question base stimulus	p= 0.054	p= 0.141	p= 0.823	p= 0.067	p= 0.005*

Table XIX. P-values of the Mann-Whitney U test comparing mean RTs of qualified musicians (Q) and non-musicians (N) for every stimulus created from the yes-no question and wh-question base stimuli. *Statistically significant at 0.05 level.

Finally, if we break down our data by both gender and musical training, the results show that there is interaction between musical training and gender in the sense that only female musicians are significantly faster for the two continua than male musicians, which allows us to conclude that what is really relevant for RT in reacting to stimuli is musical training and not gender.

3.2.3. Discrimination results

The material for the discrimination task was made up of different pairs (AB pairs), different pairs but in reverse order (BA pairs) and same pairs (AA pairs) for each stimulus base. Figure 14 shows the proportion of AB hits ("different" responses to dissimilar pairs: 263 vs. 248 Hz, 248 vs. 233 Hz, etc.), BA hits ("different" responses to dissimilar pairs but in reverse order: 248 vs. 263 Hz, 233 vs. 248 Hz, etc.) and false alarms ("different" responses to identical pairs: 263 vs. 263 Hz, 248 vs. 248 Hz, etc.) corresponding to the continuum created from the yes-no question base stimulus (ranging from 263 to 203 Hz).

We find a discrimination peak at the pair 218 vs. 233 Hz (BA pair), which one would expect, given the functions yielded by the identification results. The most striking aspect of these results has to do with the BA hits. Note that the discrimination peak appears in BA pairs in which the second stimulus has a higher peak than the first. These results suggest that listeners have more trouble discriminating between pairs of stimuli presented in AB order (pairs in which the second stimulus has a lower peak than the first).

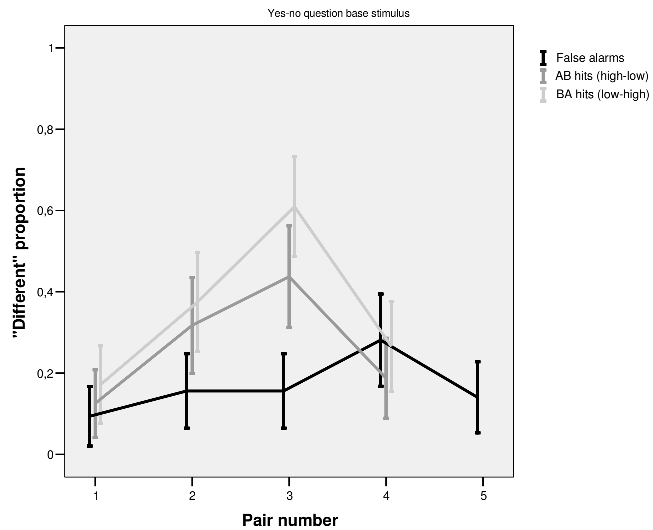


Figure 14. Proportion of "different" responses for pairs that were truly different (hits) and pairs that were actually identical (false alarms) corresponding to the synthesized continuum ranging from 263 to 203Hz (continuum from yes-no question base stimulus). Error bars represent the standard error of the mean.

Several statistical tests were performed. A Friedman test showed that the AA, AB and BA functions were significantly different overall (p= 0.000). A Wilcoxon matched pairs signed rank test compared the AA, AB and BA functions by pairs and displayed significant differences between them (Table XX). Table XXI shows the results of the Wilcoxon matched pairs signed rank tests comparing the responses to adjacent pairs for AB and BA functions. Table xxii shows the results of the Wilcoxon matched pairs signed rank tests comparing the responses to each pair in one function (AB) vs. the responses to the same pair in the other function (BA), for instance: pair AB1 (stimuli 1-2) vs. pair BA1 (stimuli 2-1), pair AB2 (stimuli 2-3) vs. pair BA2 (stimuli 3-2) and so on. From the results displayed in Table XI we can infer that the real discrimination peak is in the BA function because pair 3 (218-233 Hz), where the discrimination peak is located, is significantly different from both pair 2 and pair 4 (shaded values). There is also a significant difference (p= 0.041) between the hypothetical discrimination peaks of both AB and BA functions (see Table XXII, AB3-BA3, shaded value), which confirms the existence of order of presentation effects.

	AA-AB	AB-BA	BA-AA
Yes-no question base stimulus	p= 0.008*	p= 0.016*	p= 0.000*

Table XX. P-values of the Wilcoxon matched pairs signed rank test comparing AA, AB and BA functions. *Statistically significant at 0.05 level.

	AB1-AB2	AB2-AB3	AB3-AB4	BA1-BA2	BA2-BA3	BA3BA4
Yes-no question base stimulus	p= 0.014*	p= 0.209	p= 0.001*	p= 0.007*	p= 0.003*	p= 0.000*

Table XXI. P-values of the Wilcoxon matched pairs signed rank test comparing adjacent pairs for AB and BA functions. *Statistically significant at 0.05 level.

	AB1-BA1	AB2-BA2	AB3-BA3	AB4-BA4
Yes-no question base stimulus	p= 0.467	p= 0.371	p= 0.041*	p= 0.275

Table XXII. P-values of the Wilcoxon matched pairs signed rank test comparing each pair in one function (AB) vs. the same pair in the other function (BA). *Statistically significant at 0.05 level.

Results of the discrimination task using stimuli created from the wh-question base stimulus (corresponding to the continuum ranging from 203 to 263 Hz) are plotted in Figure 15, as the proportion of AB hits (203 vs. 218 Hz, 218 vs. 233 Hz, etc.), BA hits (218 vs. 203 Hz, 233 vs. 218 Hz, etc.) and false alarms (203 vs. 203 Hz, 218 vs. 218 Hz, etc.).

We can observe a major discrimination peak at AB pair 218 vs. 233 Hz (the response proportion ratio reaches nearly 0.7), which agrees with the results of the identification task (we found the shift between the categories around 233 Hz). These results confirm the findings shown in Figure 16, that is, it appears that subjects have trouble discriminating between stimuli when the direction of change in frequency is downwards.

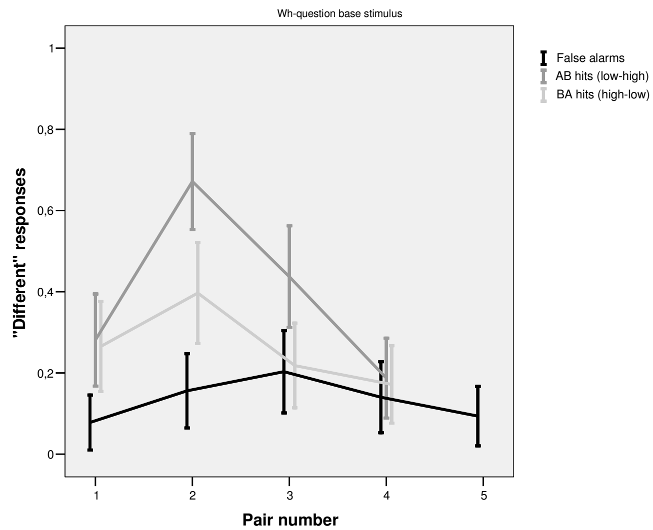


Figure 15. Proportion of "different" responses for pairs that were truly different (hits) and pairs that were actually identical (false alarms) corresponding to the synthesized continuum ranging from 203 to 263Hz (continuum from wh-question base stimulus). Error bars represent the standard error of the mean.

A Friedman test revealed that there are significant differences (p=0.000) between the AA, AB and BA functions. A Wilcoxon matched pairs signed rank test comparing AA, AB and BA functions in pairs showed significant differences between them (Table XXIII). Table XXIV shows the results of the Wilcoxon matched pairs signed rank test comparing the responses to adjacent pairs for AB and BA functions. Table xxv shows the results of the Wilcoxon matched pairs signed rank tests comparing the responses to each pair in one function (AB) vs. the responses to the same pair in the other function (BA). The discrimination peak is located again at the 218 vs. 233 Hz interval in the AB function (low-high order). This can be seen in the results shown in Table XXIV, where the differences between pair 2 and adjacent pairs in the AB function (shaded values) are significant. By contrast, differences between pair 2 and adjacent pairs in the BA function (underlined values) are not significant. Order of presentation effects are confirmed from the results shown in Table XXV, where there are significant differences (shaded values) between pairs 2 of the AB and BA functions but not between the other pairs.

	AA-AB	AB-BA	BA-AA
Wh-question base stimulus	p= 0.000*	p= 0.001*	p= 0.000*

Table XXIII. P-values of the Wilcoxon matched pairs signed rank test comparing AA, AB and BA functions. *Statistically significant at 0.05 level.

	AB1-AB2	AB2-AB3	AB3-AB4	BA1-BA2	BA2-BA3	BA3BA4
Wh-question base stimulus	p= 0.000*	p= 0.011*	p= 0.002*	p= 0.144	p= 0.041*	p= 0.513

Table XXIV. P-values of the Wilcoxon matched pairs signed rank test comparing adjacent pairs for AB and BA functions. *Statistically significant at 0.05 level.

	AB1-BA1	AB2-BA2	AB3-BA3	AB4-BA4
Wh-question base stimulus	p= 0.841	p= 0.003*	p= 0.006*	p= 0.796

Table XXV. P-values of the Wilcoxon matched pairs signed rank test comparing each pair in one function (AB) vs. the same pair in the other function (BA). *Statistically significant at 0.05 level.

3.2.4. Models of discrimination performance

3.2.4.1. Relating discrimination responses to identification responses

The classical assumption of categorical perception is that stimulus pairs will be distinguished only if they have been identified as members of different categories. 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. In order to determine whether discrimination can be predicted by identification proportions, two formulas for predicting discrimination were tested. The first formula was taken from Liberman et al. (1957), who used it to predict the results of an ABX discrimination task. However, Pollock and Pisoni (1971) showed that the same equation can also be used to predict performance in a same/different discrimination task:

(1) $P(disc_{12})= 0.5[1 + (p_1 - p_2)^2]$

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. In the words of 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 in studies on categorical perception by children such as Wolf (1973) and Brandt and Rosen (1980) that employed same/different discrimination tasks:

(2) Proportion discriminated= $(P_{1a} \times P_{2b}) + (P_{1b} \times P_{2a})$, 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 16 and 17 show the obtained and predicted discrimination functions as a result of the application of formula (1). The proportion of correct discrimination for each pair was calculated as in Francis and Ciocca (2003), which uses the same kind of task as in this study, that is, as the average of the proportion of a “different” response to different pairs and the proportion of a “same” response to same pairs. For example, the proportion correct for the 1-2 pair was the average of the proportion of “different” responses to the 1-2 and 2-1 pairs and the proportion of “same” responses to the 1-1 and 2-2 pairs.

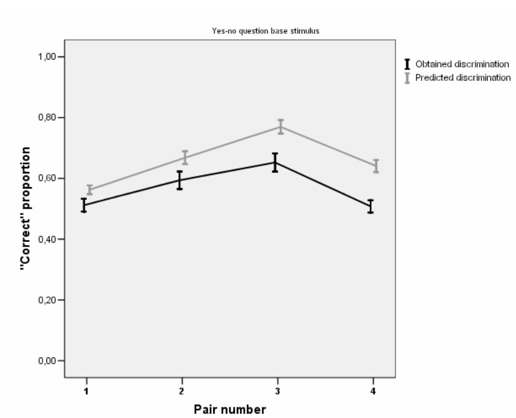


Figure 16. Predicted and obtained discrimination proportions for the continuum created from the yes-no question base stimulus (continuum ranging from 263 to 203 Hz). Error bars represent the standard error of the mean.

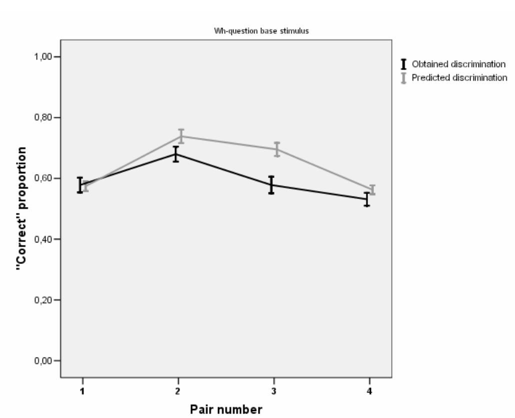


Figure 17. Predicted and obtained discrimination proportions for the continuum created from the wh-question base stimulus (continuum ranging from 203 to 263 Hz). Error bars represent the standard error of the mean.

A Wilcoxon matched pairs signed rank test was carried out in order to compare predicted and obtained discrimination proportions (Table XXVI) for pairs. 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 only in the case of stimuli created from the wh-question base stimulus (shaded values) are there no significant differences between predicted and real discrimination peaks. Thus, it was concluded that the use of this type of prediction formula is not appropriate for our data.

	Obtained 1 vs. predicted 1	Obtained 2 vs. predicted 2	Obtained 3 vs. predicted 3	Obtained 4 vs. predicted 4
Yes-no question base stimulus	p= 0.127	p= 0.077	p= 0.022*	p= 0.004*
Wh-question base stimulus	p= 0.802	p= 0.331	p= 0.003*	p= 0.174

Table XXVI. P-values of the Wilcoxon matched pairs signed rank test comparing predicted and obtained discrimination proportions. *Statistically significant at 0.05 level.

Figures 18 and 19 show the predicted and obtained discrimination functions as a result of the application of formula (2). The real discrimination values have been calculated as the proportion 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.

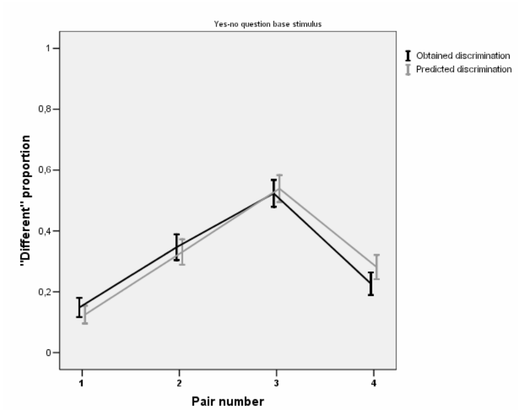


Figure 18. Predicted and obtained discrimination proportions for the continuum created from the yes-no question base stimulus (continuum ranging from 263 to 203 Hz). Error bars represent the standard error of the mean.

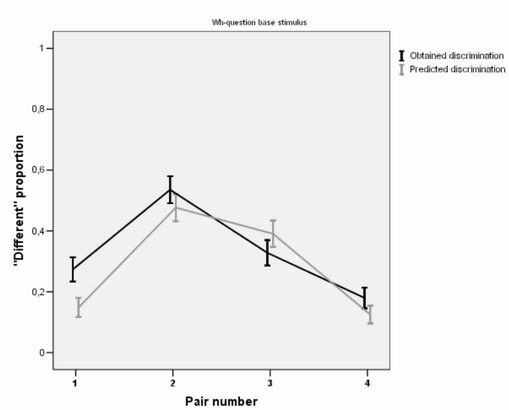


Figure 19. Predicted and obtained discrimination proportions of the continuum created from the wh-question base stimulus (continuum ranging from 203 to 263 Hz). Error bars represent the standard error of the mean.

The results of a Wilcoxon matched pairs signed rank test (Table XXVII) show that the differences between predicted versus obtained discrimination is significant only for pair 1 with stimuli created from the wh-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.

	Obtained 1 vs. predicted 1	Obtained 2 vs. predicted 2	Obtained 3 vs. predicted 3	Obtained 4 vs. predicted 4
Yes-no question base stimulus	p= 0.590	p= 0.796	p= 0.796	p= 0.307
Wh-question base stimulus	p= 0.009*	p= 0.354	p= 0.317	p= 0.223

Table XXVII. P-values of the Wilcoxon matched pairs signed rank test comparing predicted and obtained discrimination proportions. *Statistically significant at 0.05 level.

3.2.4.2. 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) denotes the distance between a signal and a signal + noise, and an experiment may be conceived of as a presentation of signals and non-signals to subjects, who try to detect only the signals and ignore the non-signals. D prime has been used in the discrimination literature for obtained and predicted discrimination by Best et al. (1981) and in addition to %correct/%different responses to different pairs by Francis and Ciocca (2003). In order to validate the results plotted in Figures 14 and 15 related to the presence of order of presentation effects, Signal Detection Theory was applied to our data. Figures 20 and 21 show the discrimination results presented as d' for each stimulus pair in each order of presentation. 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 and Creelman (1991), d' was calculated using roving methods[6] (using Table A5.4, pp. 338-354).

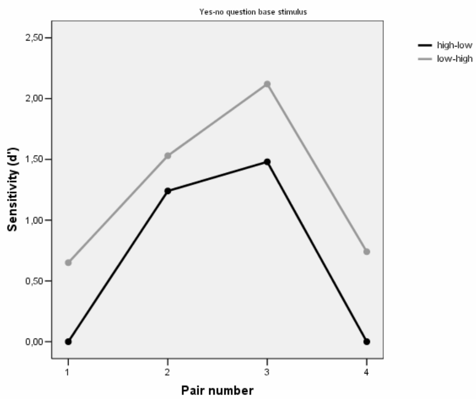


Figure 20. D' calculated for both orders of presentation of pairs from the continuum created from the yes-no question base stimulus.

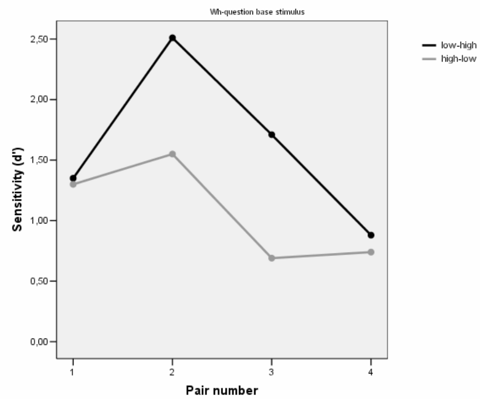


Figure 21. D' calculated for both orders of presentation of pairs from the continuum created from the wh- question base stimulus.

A Wilcoxon matched pairs signed rank test (Table XXVIII) was carried out on hit rate minus false alarm rate for pairs in different orders. Significant differences were found only for the second and third pairs (shaded values) for stimuli created from the wh-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. Nevertheless, at pair 3, which represents the discrimination peak, the difference is nearly significant (p= 0.064). Thus, the application of d prime in addition to %different responses to different/same pairs has confirmed the presence of order of presentation effects which are statistically significant only for pairs 2 and 3 for stimuli created from the wh-question base stimulus.

	AB1-BA1	AB2-BA2	AB3-BA3	AB4-BA4
Yes-no question base stimulus	p= 0.491	p= 0.405	p= 0.064	p= 0.225
Wh-question base stimulus	p= 0.922	p= 0.008*	p= 0.007*	p= 0.822

Table XXVIII. P-values of Wilcoxon matched pairs signed rank test performed on hit rate minus false alarm rate for pairs in different orders. *Statistically significant at 0.05 level.

[6] According to Macmillan and Creelman (1991), there are two different strategies 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 2 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).

3.3. Discussion

The present study appears to provide evidence that Majorcan Catalan listeners make categorical linguistic use of F0 scaling differences in perceiving yes-no questions as opposed to wh-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. We observed in Figure 3 that the presence of the accented wh-particle in the continuum created from the yes-no question base stimulus and the presence of the unaccented wh-particle in the continuum created from the wh-question base stimulus do not interfere in the categorical perception, that is, the identification functions appear unmistakably S-shaped with a proportion of identifications that goes from more than 0.8 to about 0.2 (in the case of the continuum ranging from 263 to 203 Hz) and from less than 0.2 to about 0.8 (in the case of the continuum ranging from 203 to 263 Hz) within 2 steps of the 5-step continuum. In other words, responses to the two continua mirror each other.

Evidence of this linguistic contrast is also provided by RT measurements. Thus, a mean RT peak could be observed in Figure 9 at the identification boundary for both continua. In addition, 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, the application of the Haskins formula and the more general formula taken from Godfrey et al. (1981) permitted us to predict the shape of the discrimination curve. From Figures 18, 19, 20 and 21 and from statistical analyses, it can be said that the formula used by Godfrey et al. (1981) and based simply on phonetic categorization, without assumptions about guessing, is the one that best fits our data. Because the results of a Wilcoxon matched pairs signed rank test (Table XXVIII) show that the differences between predicted and obtained discrimination are significant only for pair 1 with stimuli created from the wh-question base stimulus, it is concluded that this contrast is categorically perceived. However, for classic CP it is not enough that either discrimination can be predicted from identification results, or that an abrupt shift in the identification function is observed; this shift must correspond to the discrimination peak. By comparing from statistical analyses how different adjacent stimuli were responded to, it has been observed that, although there is a shift from one response to the other in the range of values between 218 to 248 Hz, the threshold between the "yes-no question" and the "wh-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.

Thus, on the basis of this evidence, we may safely conclude that a strong perceptual cue that Majorcan listeners use when distinguishing yes-no questions from wh-questions is the difference in pitch height of the leading tone of the nuclear accent. 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 that would play a role in ambiguous stimuli. This is supported by two relevant characteristics of both identification and discrimination results: first, the high proportion of "yes-no question" responses at stimuli 3 and 4 for the continuum created from yes-no question base stimulus in identification results; second, the general tendency to have more sharply differentiated results in the continuum created from the wh-question base stimulus than in that created from the yes-no question base stimulus, not only in the identification task but also in the discrimination task. For the latter it was hypothesized that this tendency might be due to a question of bias. That is, faced with the dilemma of having to identify these stimuli as yes-no or wh-questions, most listeners decided to choose "yes-no question" responses. But note that if this was so, the same thing would have happened for stimuli from the wh-question base stimulus. For the former, in a preliminary first stage of the study, it was speculated that such a tendency could be because of the order in which listeners performed the discrimination task. After the practice block, the first and third blocks were made up of stimuli from the yes-no question base stimulus and the second and fourth of stimuli from the wh-question base stimulus. So the subjects had become better trained in the discrimination task by mid-end of the task, that is, by the second and third block (bear in mind that listeners had to judge five blocks: one practice block and four test blocks). This may explain why the function of results for the second and fourth blocks (both blocks made up of stimuli proceeding from the wh-question base stimulus) plots better results. Afterwards, the order of the blocks was counterbalanced, so the more sharply distinguished perception of the two categories cannot be due to that factor. For this reason, it seems plausible to suspect that the presence of an accented particle may act as a supplementary cue, a cue which would only be accessible in the continuum created from the wh-question base stimulus and would come into play in ambiguous stimuli. This would explain why stimuli 3 and 4 from the yes-no question base continuum have been identified with a high proportion of "yes-no question" choices and why results from the wh-question base continuum are always more sharply differentiated and, in terms of RT measurements, more consistent.

The impact of gender differences and differences involving the degree of musical training on intonational perception as well as mean RTs are confirmed. It seems that the pattern of results for the identification task is more consistent for women than for men and that women are faster than men in the classification task. In the same way, qualified musicians seem to have less variability in responses to the identification task than occasional musicians and non-musicians, and seem to be a little faster in making identification decisions except in the case of stimulus 3, which represents the boundary between yes-no question and wh-question categories. When the data on gender and musical training was compared, interaction was found between musical training and gender in the sense that only female musicians were significantly faster for the two continua than male musicians. Thus, what seems to be relevant for the time that subjects need to identify the stimuli is musical training and not gender.

Finally, it is worth trying to account for the discrimination asymmetries that our results report. 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 and Rietveld, 1988). F0 declination has been argued to be a universal and language-independent characteristic of speech production. Evidence for compensation of this declination effect has been provided for American English listeners (Pierrehumbert, 1979), Dutch listeners (Gussenhoven and Rietveld, 1988) and Cantonese listeners (Wong, 1999). According to Francis and Ciocca (2003), these asymmetries may be explained in terms of a compensation for a 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. However, there is no clear evidence that Majorcan Catalan listeners exhibit F0 declination in speech production or that they compensate perceptually for this expected declination. For this reason, further research on the perception of pitch in Majorcan Catalan is necessary to confirm that our results can be explained in these terms.[7]

[7] Additionally, it may be that the presence of the unstressed [u] and stressed [i] (both high vowels) in the stimuli used in this study ('Que l'hi d**u**r**i**es'?/'Què li d**uri**es?') may have enhanced the difference in perception of low-high order tokens, as it is known that there exists a correlation between vowel height and F0 (intrinsic pitch) and that this effect is more important in stressed that in unstressed syllables (Gussenhoven, 2004).

4. Conclusions

4.1. Summary of the findings

The goals of this study were 1) to investigate the acoustic properties of both interrogative particles *que* ('that') and *què* ('what') and the nature of the falling nuclear accent (H+L*) that characterizes both yes-no questions and wh-question in Majorcan Catalan interrogatives, 2) to test through the Categorical Perception paradigm (CP) whether listeners make categorical linguistic use of F0 scaling differences in distinguishing yes-no questions from wh-questions, and 3) to test if the application of the CP paradigm to scaling differences involves the presence of asymmetries in the discrimination of a particular change.

The production study of the acoustic properties of the stressed/accented particle *què* and the unstressed/unaccented particle *que* that introduce wh-questions and yes-no questions respectively reveals differences in pitch, vowel quality and optional differences in duration for the vowel of the interrogative particles. Thus, the vowel of the accented interrogative particle *què* is acoustically characterized by a higher pitch, a higher F1 value and an optional increment in duration compared to the vowel of the unaccented interrogative particle. In previous studies of stress and accent in Catalan (Astruc and Prieto, 2005; Prieto and Ortega-Llebaria, 2005), differences in vowel quality and duration have been interpreted as acoustic correlates of stress, and pitch differences as acoustic correlates of accent. Moreover, utterance length has an important effect on the pitch of the interrogative particles, especially of wh-questions, having a higher pitch as the length of the utterance increases. This has been taken as evidence for the presence of tonal preplanning. Nevertheless, the fact that there is variation between speakers shows that global preplanning cannot be taken in the strong sense, but rather as soft preplanning.

Results from the production experiment also showed that sentence type has a large effect on the pitch of the leading tone H within the nuclear accent H+L*, it being significantly higher for yes-no questions than for wh-questions. The H pitch value is also affected by utterance length, given that the pitch of the leading tone H becomes lower as the number of pitch accents in the utterances increases, especially in wh-questions. This effect has been interpreted as downstep implementation. However, this utterance length effect does not attenuate the sentence type effect; on the contrary, regardless of the different number of pitch accents, the difference between yes-no questions and wh-questions in terms of the height of the H point remains statistically significant. The fact that utterance length effects are smaller for yes-no questions than for wh-questions, since there is not a significant lowering for all speakers as the number of pitch accents increases, has been interpreted as a mechanism to enhance the inflection between the super-high variant of the H leading tone and the L tone that characterizes yes-no questions.

On the other hand, we found a small sentence type effect on the pitch of the L tone, and also a very small utterance length effect and not significant across speakers and sentences types.

Results from the perception experiment that followed the production experiment were interpreted taking into account all the peculiarities that it seems that can be derived from the previous studies in which the CP paradigm has been applied to intonational contrasts (see § 1.3.2., page 16 for a reminder of these peculiarities). Furthermore, there have been considered the two criteria that the results of the application of the CP paradigm should fulfill, namely, that identification proportions should predict discrimination accuracy and that discrimination peaks should correspond to the location of the category boundaries determined by identification. Having checked that the discrimination results can indeed be predicted from the identification proportions, that the discrimination peak lies within the crossover region between the categories, and that the RT measurement results agree with the findings of previous studies regarding the discreteness of the contrasts tested, we claim a categorical status for both the upstepped high tone and the "normal" high tone within the nuclear accent H+L* that characterizes yes-no questions and wh-questions in Majorcan Catalan. The perceived categoriality of this contrast permits us to say that pitch scaling is a very strong cue in distinguishing yes-no questions from wh-questions in this language. Nevertheless, this does not exclude the possibility that other cues such as the presence of an accent in the interrogative particle *què* play an important role. Interestingly, the results seem to suggest that the presence of the accented interrogative particle *què* would facilitate the identification of ambiguous stimuli, that is, any stimulus within the crossover area between the two categories.

Finally, as reported by previous studies, the application of the CP paradigm involved discrimination asymmetries that depended on the direction of change, it being easier to distinguish between the stimuli pair when the direction of change is upwards in terms of tonal level. These asymmetries were explained in terms of a compensation mechanism for an expected declination in F0 over the course of an utterance. This compensation mechanism would enhance the perception of the difference between two tokens when the second token has a higher pitch than the first.

4.2. Conclusions

In sum, our production experiment showed that speakers systematically produced the leading tone H of the nuclear accent H+L* with a significantly higher pitch in yes-no questions than in wh-questions. Even when the number of pitch accents in utterances is increased, this difference in pitch of the H tone between yes-no questions and wh-questions is still significant. Moreover, a difference in pitch scaling on the leading H tone seems to be key in causing listeners to change the perceived category from yes-no questions to wh-questions and vice versa, as was confirmed by the results of our perception experiment based on these CP paradigm. These results prove that pitch scaling on the H level has a phonological character in Majorcan Catalan interrogatives.

Appendix

Sentences used in the production experiment. The questions were read aloud by subjects. The answers were read off the script by the author of this study. Answers were provided in this way in order to facilitate the interpretation of the utterances by the speakers. Stressed syllables are marked in boldface.

BLOCK I

*Que hi va **veu**re?* ('Did s/he see?') *–Sí, hi havia llums* ('Yes, there were lights').

***Què** va **beu**re?* ('What did s/he drink?') *–Una cervesa, com sempre* ('A beer, as usual').

***Què** en va **ven**dre?* ('What did s/he sell of it?') *–Set quilos* ('Seven kilos').

*Que se va **mou**re?* ('Did s/he move?') *–No gens* ('Not at all').

*Que va **beu**re?* ('Did s/he drink?') *–No, mai no beu* ('No, s/he never drinks').

***Què** va **mol**dre?* ('What did s/he grind?') –Un paquet de cafè ('A packet of coffee').

*Que en va **ven**dre?* ('Did s/he sell any of it?') *–No* ('No').

***Què** va **mou**re?* ('What did s/he move?') *–Tots es mobles* ('All the furniture').

*Que va **mol**dre?* ('Did s/he grind?') *–No, perquè no li varen dur res* ('No, because nobody brought him/her anything').

***Què** hi va **veu**re?* ('What did s/he see?') *–Res* ('Nothing').

BLOCK II

***Què** vo**li**a **beu**re?* ('What did s/he want to drink?') *–Un tassonet de vi* ('A small glass of wine').

*Que hi vo**li**a **veu**re?* ('Did s/he want to see?') *–Ves!* ('Yes, indeed!').

***Què** li vo**li**a **mol**dre?* ('What did s/he want to grind for him?') *–Només un paquet de cafè* ('Just a little packet of coffee').

***Què** vo**li**a **mou**re?* ('What did s/he want to move?') *–Tots es mobles* ('All the furniture').

*Que vo**li**a **beu**re?* ('Did s/he want to drink?') *–Sí, com cada dissabte* ('Yes, just like every Saturday').

*Que l'hi vo**li**a **mol**dre?* ('Did s/he want to grind it for him?') *–No* ('No').

***Què** n'espe**ra**va **ven**dre?* ('What did s/he expect to sell of it?') *–Almanco tres quilos* ('At least three kilos').

*Que se vo**li**a **mou**re?* ('Did s/he want to move?') **–***Sí* ('Yes').

***Què** hi vo**li**a **veu**re?* ('What did s/he want to see there?') *–Sa casa acabada* ('The house finished').

*Que n'espe**ra**va **ven**dre?* ('Did s/he expect to sell any of it?') *–Sí* ('Yes').

BLOCK III

*Que hi vo**li**a tor**nar** a **veu**re?* (‘Did s/he want to see again?’) *–Ves!* (‘Yes, indeed!’).

Què *vo**li**a tor**nar** a **beu**re?* (‘What did s/he want to drink again?’) *–Un tassonet de vi* (‘A small glass of wine’).

Què *tor**na**va ve**nir** a **mol**dre?* (‘What did s/he come to grind again?’) *–Només un paquet de cafè* (‘Just a packet of coffee’).

Què *vo**li**a tor**nar** a **mou**re?* (‘What did s/he want to move again?’) *–Tots es mobles* (‘All the furniture’).

*Que tor**na**va ve**nir** a **mol**dre?* (‘Did s/he come to grind again?’) *–No* (‘No’).

*Que vo**li**a tor**nar** a **beu**re?* (‘Did s/he want to drink again?’) *–Sí, com cada dissabte* (‘Yes, just like every Saturday’).

Què *hi vo**li**a tor**nar** a **ven**dre?* (‘What did s/he want to sell there again?’) –Cossiols (‘Flowerpots’).

*Que se vo**li**a tor**nar** a **mou**re?* (‘Did s/he want to move again?’) –Sí (‘Yes’).

*Que hi vo**li**a tor**nar** a **ven**dre?* (‘Did s/he want to sell there again?’) –Sí (‘Yes’).

Què *hi vo**li**a tor**nar** a **veu**re?* (‘What did s/he want to see there again?’) *–Sa casa acabada* (‘The house finished’).

References

Astruc, L., and Prieto, P. 2006. Acoustic Cues of Stress and Accent in Catalan. Paper presented at *Speech Prosody 2006*, 803-806, Dresden.

Bartels, C., and Kingston, J. 1994. Salient pitch cues in the perception of contrastive focus. In *Conference on Focus*, eds. P. Bosch and R. van der Sandt, 11-28: Focus & Natural Language Processing.

Beckman, M.E., and Edwards, J. 1991. Phonological Structure and Phonetic Form. Paper presented at *Laboratory Phonology III*, UCLA.

Best, C.T., Morrongiello, B., and 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.

Bonet, E., and Lloret, M.R. 1998. *Fonologia catalana*. Barcelona: Ariel.

Brandt, J., and 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. Paper presented at *One-Day Meeting for Young Speech Researchers*, University College, London.

Caspers, J. 1998. Who's next? The melodic marking of questions vs. continuation in Dutch. *Language and Speech* 41: 371-394.

Chen, Aoju. 2003. 8th European Conference on Speech Communication and Technology. Paper presented at *Eurospeech 2003*, 97-100, Geneva.

Connell, B. 2003. Pitch realization and the four tones of Mambila. In *Cross-Linguistic Studies of Tonal Phenomena*, ed. S. Kaji. Tokyo: Research Institute for the Languages and Cultures of Asia and Africa.

Connell, B. 2004. Tone, Utterance Length and F0 Scaling. In *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*. Beijing, China: Research Institute for the Language and Cultures of Asia and Africa.

Cooper, W.J., and Sorensen, J. 1981. *Fundamental Frequency in Sentence Production.* Heidelberg: Springer.

Cummins, F., Doherty, C., and Dilley, L. 2006. Phrase-Final Pitch Discrimination in English. Paper presented at *Speech Prosody*, 467-470, Dresden.

D'Imperio, M. 2000. The role of perception in defining tonal targets and their alignment, The Ohio State University.

D'Imperio, M., and House, D. 1997. Perception of Questions and Statements in Neapolitan Italian. Paper presented at *Eurospeech 97*, 251-254, Rhodes, Greece.

Dilley, L., and Brown, M. to appear. Effects of relative F0 level on F0 extrema in an imitation task. *Journal of Phonetics*.

Face, T.L. 2004. F0 Peak Height and the Perception of Sentence Type in Castilian Spanish: Brigham Young University and The Ohio State University.

Falé, I., and Hub Faria, I. 2006. Categorical Perception of intonational contrasts in European Portuguese. Paper presented at *Speech Prosody*, 69-72, Dresden.

Finney, D.J. 1971. *Probit analysis*. Cambridge: Cambridge University Press.

Francis, A., and 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., and Kei Chit Ng, B. 2003. On the (non)categorical perception of lexical tones. *Perception & Psychophysics* 65: 1029-1044.

Frota, S. 2002. Tonal association and target alignment in European Portuguese nuclear falls. Paper presented at *Laboratory Phonology, VII*.

Fujisaki, H., and Kawashima, T. 1971. A model of the mechanism for speech perception: Quantitative analyses of categorical effects in discrimination. *Annual Report of the Engineering Research Institute*: 59-68.

Gerrits, E., and Schouten, M.E.H. 2004. Categorical perception depends on the discrimination task. *Perception & Psychophysics* 66: 363-376.

Godfrey, John J., K., Syrdal-Laskym Ann, Millay, Kathleen K., and Knox, Carol M. 1981. Performance of Dyslexic Children on Speech Perception Tests. *Journal of Experimental Child Psychology* 32: 401-424.

Gussenhoven, C., and Rietveld, A. 2000. The behavior of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech* 43: 183-203.

Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hart, J. 't. 1979. Explorations in automatic stylization of F0 curves. *IPO Annual Progress Report* 14: 61-65.
Jensen, Marvin D., and 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., and 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. Paper presented at *11th International Congress of Phonetic Sciences*, 331-333, Tallinn.
Ladd, D.R., and Morton, R. 1997. The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics* 25: 313-342.
Liberman, A., Harris, K., Hoffman, H., and Griffith, B. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54: 358-368.
Liberman, M., and Pierrehumbert, J. 1984. Intonational Invariance under Changes in Pitch Range and Length. In *Language Sound Structure*, eds. M. Aronoff and R.T. Oehrl, 157-234. Cambridge: MIT Press.
Lindau, M. 1986. Testing a model of intonation in a tone language. *JASA* 80: 757-764.
Macmillan, N.A., and Creelman, C. Douglas. 1991. *Detection Theory: A User's Guide*. Cambridge: Cambridge University Press.
Macmillan, Neil A., Kaplan, Howard L., and Creelman, C. Douglas. 1977. The Psychophysics of Categorical Perception. *Psychological Review* 84: 452-471.
Mascaró i Pons, I. 1986. Introducció a l'entonació dialectal catalana. *Randa* 22: 5-38.
Moll, Francesc de B. 1971. *La lengua de la Baleares (III)*. Palma de Mallorca: Editorial Moll.
Payà, M., and 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.
Pierrehumbert, J, and Steele, S. 1989. Categories of tonal alignment in English. *Phonetica* 46: 181-196.
Pierrehumbert, J. B., and Beckman, M.E. 1988. *Japanese Tone Sructure*. Cambridge: MIT Press.
Pierrehumbert, J.B. 1980. *The Phonetics and phonology of English intonation*. New York: Garland Press.
Pisoni, D.B. 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Percept. Psychophys.* 13: 253-260.
Pisoni, D.B., and Tash, J. 1974. Reaction times to comparisons within and across phonetic categories. *Percept. Psychophys.* 15: 285-290.
Polka, L., and Werker, J.F. 1994. Developmental changes in perception of non-native vowe contrasts. *J. Exp. Psychol.: Human Percept. Perform.* 20: 421-435.
Polka, L., and Bohn, O. 2003. Asymmetries in vowel perception. *Speech Communication* 41: 221-231.
Pollock, I., and 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. 2001. L'entonació dialectal del català: el cas de les frases interrogatives absolutes. In *Actes del Novè Col·loqui de la North American Catalan Society (Barcelona 1998)*, eds. A. Bover, M.R. Lloret and M. Vidal-Tibbits, 347-377: Publicacions de l'Abadia de Montserrat.
Prieto, P. 2003. Scaling of H1 peaks in Spanish: evidence from five sentence types. Paper presented at *15th ICPhS*, Barcelona.
Prieto, P., D'Imperio, M., Elordieta, G., Frota, S., and Vigário, M. 2006. Evidence for *soft* preplanning in tonal production: Initial scaling in Romance. Paper presented at *Speech Prosody*, 803-806, Dresden.
Prieto, P., and Ortega-Llebaria, M. 2006. Stress and Accent in Catalan and Spanish: Patterns of duration, vowel quality, overall intensity, and spectral balance. Paper presented at *Speech Prosody*, Dresden.
Redi, Laura C. 2003. Categorical effects in production of pitch contours in English. Paper presented at *15th ICPhS*, 2921-2924, Barcelona.
Remijsen, B., and van Heuven, V. 1999. Gradient and categorical pitch dimensions in Dutch: Diagnostic test. In *Proceedings of the 14th International Congress of Phonetic Sciences*, 1865-1868.
Repp, B.H., and Crowder, R.G. 1990. Stimulus order effects in vowel discrimination. *Journal of Acoustical Society of America* 88: 2080-2090.
Rialland, A. 2001. Anticipatory raising in downstep realization: Evidence for preplanning in tone production. Paper presented at *Symposium Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Japanese Accentology, and Other Topics.*, 301-321, Tokyo, University of Foreign Studies.

Rogers, H. 2003. Male-female Phonetic Bibliography: http://ccat.sas.upenn.edu/~haroldfs/bibliogs/malefeml.html.

Rosch, E. 1975. Cognitive reference points. *Cognition* 7: 532-547.

Schneider, K., and Linftert, B. 2003. Categorical perception of boundary tones in German. Paper presented at *15th ICPhS*, 1-4, Barcelona.

Schouten, B., and van Hessen, A. 1992. Modelling phoneme perception: I. Categorical perception. *Journal of the Acoustic Society of America* 92: 1841-1855.

Snider, K. 1998. Tone and utterance length in Chumburung: An instrumental study. Paper presented at *28th Colloquium on African Languages and Linguistics*, Leiden.

Tartter, V.C. 1981. A comparison of the identification and discrimination of synthetic vowel and stop-sonsonant stimuli with various acoustic properties. *Journal of Phonetics* 9: 477-486.

Tversky, A., and Gati, I. 1978. Studies of similarity. In *Cognition and Categorization*, eds. E. Rosch and B.B. Lloyd. Hillsdale, NJ: Lawrence Erlbau.

Uldall, E. 1964. Dimensions of meaning in intonation. In *In honour of Daniel Jones: Papers contributes on the occasion of his eightieth birthday*, eds. D. Abercrombie, P. Fry, N. McCarthy and J. Trim Scott, 271-279. London: Longman.

Vanrell, M.M. 2003. Estudi sobre l'entonació dels parlars de Llucmajor, Campos i Porreres, Departament de Filologia Catalana, Universitat de les Illes Balears.

Welby, P. 2006. French intonational structure: Evidence from tonal alignment. *Journal of Phonetics* 34: 343-371.

Wolf, C.G. 1973. The perception of stop consonants by children. *Journal of Experimental Child Psychology* 8: 351-361.