Broad focus vs contrastive focus: Is there categorical perception in Pisa Italian?

Barbara Gili Fivela

University of Salento & CRIL – Lecce, Italy
barbara.gili@ateneo.unile.it

Abstract

The paper reports the results of a perception experiment run in order to test whether two pitch accents are categorically perceived. The pitch accents under investigation are found in nuclear position, in Pisa Italian broad focus and contrastive focus phrases.

The perception experiment is run within the Categorical Perception paradigm, i.e. by testing both the identification and the discrimination of stimuli that represent a continuum of variation from one pattern to the other. In particular, in order to ensure that the results do not depend on possible limits in the task used for the identification test, the same continuum of stimuli is proposed in three different tasks. The results of both the identification test - three tasks - and the discrimination test are then discussed with reference to the criticism found in the literature on the use of Categorical Perception paradigm for investigating intonation.

1. Introduction

It has been observed in the literature [5] that, according to some theories of intonational meaning, only categories are the linguist's concern, although intonation mainly deals with paralanguage (e.g., it conveys expressive and attitudinal information). That is, many studies only investigate categories of intonation, trying to distinguish them from paralinguistic modifications: the former are expected to be discrete - as linguistic elements are - and the latter are expected to be gradient. Nevertheless, teasing paralinguistic and linguistic information apart in intonation may not be simple, especially when the meaning or the function of patterns is taken into account.

One way to test the presence of categorical, rather than gradient, variations is the Categorical Perception (CP) paradigm [8]. Within this paradigm, both an identification and a discrimination test have to be performed. In the identification test, subjects hear stimuli that represent a continuum of gradual changes (for specific characteristics) between two different patterns, possibly representing two phonological categories: subjects are asked to assign stimuli to either category. In the discrimination test, subjects hear pairs of stimuli - adjacent in the continuum used for the identification task - and judge whether they are the same or they are different. The results of the two tests - the presence of an abrupt shift in the association of stimuli to either category, and the presence of a high peak in the discrimination rate for the stimuli across that shift – make it possible to see whether there is categorical perception.

The use of this paradigm for studying intonation has been giving heterogeneous results. For instance, it was used to demonstrate the existence of categorical perception of pitch height contrast in Majorcan Catalan wh- and yes-no questions [10]. On the other hand, in investigating low and high boundary tones in German, Schneider et al. [7] report that a plateau rather than a peak was found in the discrimination results. Some criticism on the use of CP paradigm for investigating intonation is then reported. However, a relevant example for the purpose of the present paper is described in [6], where the authors use the CP paradigm to test the categorical distinction of the pitch accents found in neutral statements and contradiction in English. In their experiment, the authors find no categorical perception, although subjects could interpret categorically stimuli at the extremes of the continuum.

In the present paper, the CP paradigm will be used to test the perception of pitch accents found in nuclear position in broad focus and in contrastive focus phrases in Pisa Italian (patterns that play functions similar to those considered for English in [6]). The test is particularly complex because of the differences in the phonetic characteristics of the two pitch accents under investigation. In fact, they differ in terms of alignment and scaling characteristics and, crucially, in terms of the presence/absence of a low target aligned with the beginning of the syllable [3]. In the experiment, alignment and scaling characteristics have been manipulated separately in order to find out their specific role in perception. The test, therefore, represents a challenge because of the complexity of the patterns, and one could argue that, in some cases, the perceptual test, at least through the CP paradigm alone, may not be fair enough.

A first perceptual study on the two Pisa Italian pitch accents was described in [2]. Some of the acoustic manipulations and results are summarized here, as part of the stimuli are used in this paper to perform different tasks in the identification test.

2. Materials and method

The experiments described below were carried out by asking Pisa Italian speakers to listen to acoustically manipulated stimuli. Stimuli were crated from source utterances whose characteristics will be specified separately for each test. In the source utterances, apart from the specific production and interpretation, the sentence was always *mangia il melone* 's/he eats the melon', produced by a female speaker of Pisa Italian who did not took part in the perception test. The speaker produced at least three repetitions of the sentence with the following interpretations [2]:

- Broad focus: the sentence was produced as an answer to the question *cosa succede?* 'what's going on?', i.e. the whole sentence contains new and relevant information;
- Contrastive focus: the sentence was produced by answering to the confirmation seeking question *mangia i lamponi?* 'does s/he eat raspberries?', eliciting a negative answer including the correct object, i.e. the melon.

Three repetitions of each type of utterance were labelled using PRAAT (Boersma and Weenink, University of

Amsterdaam) for the position of the peak, both the prenuclear and the nuclear one, for the position of the low target preceding the peak in the contrastive utterance, and for the low elbow following the peak. Tonal target fundamental frequency (F0) values and their positions with respect to syllable segment boundaries were measured. Their mean values have been considered to calculate the interval to be covered, in terms of hertz and milliseconds, in order to create a gradual shift from the characteristics of the final pitch accent found in the broad focus interpretation to those found in the accent realized in the contrastive interpretation, and viceversa.

2.1. Stimuli

One production of a broad focus utterance and one production of a contrastive focus interpretation were chosen as source utterances for the acoustic manipulation. Creating a gradual shift from a broad focus to a contrastive focus pitch accent was particularly problematic because of the number of differences among the two patterns: F0 inspection and detailed acoustic measurements of the materials described above showed that the two patterns differ for the syllable duration and both the alignment and scaling of the tonal targets (see figure 1, where the analysis in autosegmental-metrical terms is given: H+L* for the broad focus accent and [L+]H*+L for the contrastive focus accent [3]).

The nuclear syllable, in fact, is shorter in the broad than in the contrastive interpretation (219 vs 284ms); the high tone is aligned by the end of the prenuclear syllable in the broad accent and it is aligned (110ms) after the nuclear syllable left boundary in the contrastive accent; moreover, it is scaled lower in the former (184 vs 242Hz for the materials described above, although this is not always the case); furthermore, the following low tone is aligned within the nuclear syllable in the broad pattern and by the end of it in the contrastive one, and no major differences in its scaling are observed (145 vs 148 Hz). Crucially, the two patterns differ for the presence (in the contrastive accent) vs. the absence (in the broad focus accent) of a rise to peak whose starting point is aligned at the beginning of the nuclear syllable.

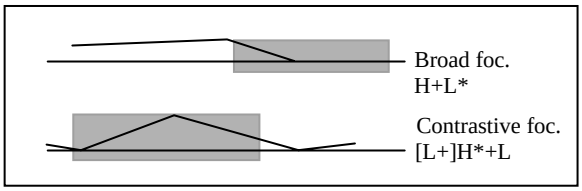

Figure 1: *Schematic representation of the pitch accents*.

Previous work on Pisa Italian [2] was carried out in order to test whether the identification of the contrastive accent could be affected by the scaling of the low target representing the starting point of the rise. The identification test with target stimuli where the height of the starting point of the raise was manipulated (the test was performed by 9 subjects of Pisa Italian who had to judge whether the stimulus could represent either a correction to a previous utterance or a neutral statement) showed that increasing the elbow height does not affect the judgements of the subjects and does not lead to the perception of a different category. Thus stimuli where the peak was the ending point of a straight interpolation from the prenuclear accent were interpreted as stimuli showing a low target aligned by the beginning of the nuclear syllable. On the basis of these results, stimuli for follow up experiments, for instance the one described here, were created by manipulating the alignment and scaling of the peak - i.e. the end point of a straight interpolation from the prenuclear accent - and following low, with no change in presence/absence of a low target aligned by the beginning of the nuclear syllable. Moreover, no changes in syllable duration were performed.

- Broad focus source

In the work described above [2], another series of stimuli was realized by manipulating the characteristics of one of the utterances corresponding to a broad focus statement, showing a nuclear pitch accent analysed as H+L*. The manipulations are summarized here, as the same stimuli have been used here in performing a different task (see section 2.2). The manipulations aimed at finding out whether a later alignment and/or a higher scaling of the peak could favour the contrastive interpretation, and how variations in alignment and scaling interact.

The source stimulus was manipulated as follows [2]:

1) The height of the peak – aligned before the nuclear syllable – was increased in 9 steps of 7.2 Hz each (for a total of 10 stimuli - the resynthesized natural stimulus plus the 9 manipulated ones); here only steps 2,4,6,8 and 9 will be shown: they are labelled as H2, H4 etc.;

2) The alignment of the peak (and of the following low target) was forward shifted in 5 steps of 22 ms each (their labels contain characters such as al1, al2). This was realized on the stimulus with a naturally scaled peak, and for each stimulus corresponding to two steps of variation in the peak scaling, i.e. the variations described in point 1 as H2, H4, H6, H8 plus H9 (for a total amount of 30 stimuli).

In comparison to stimuli used in [2], two extra steps in alignment have been added (al6 and al7 in the plots).

- Contrastive focus source

A series of stimuli was realized by manipulating the characteristics of one of the utterances corresponding to a contrastive focus statement, showing a nuclear pitch accent analysed as [L+]H*+L. The manipulations aimed at finding out whether an earlier alignment and/or a lower scaling of the peak could favour the broad focus interpretation, and how variations in alignment and scaling interact.

The source stimulus was manipulated as follows:

1) The height of the peak – aligned within the nuclear syllable – was lowered in 9 steps of 7.2 Hz each (for a total of 10 stimuli - the resynthesized natural stimulus plus the 9 manipulated ones);

2) The alignment of the peak (and of the following low target) was backward shifted in 5 steps of 22 ms each. This was realized on the stimulus with a naturally scaled peak, and for each stimulus corresponding to two steps of variation in the peak scaling, i.e. the variations described in point 1 as H2, H4, H6, H8 plus H9 (for a total amount of 30 stimuli).

The PSOLA resynthesis of stimuli was performed using PRAAT and then used for both tests described below.

2.2. Identification test

The stimuli have been presented to subjects, speakers of Pisa Italian, through the software Perceval (by LPL, Aix-en-Provence). Stimuli were presented three times in random order[1]. Three different tasks have been performed separately,

[1] The small number of stimuli repetitions was necessary in order for the test not to be too long and tiring for subjects. However, a test performed by presenting the stimuli for five times – the stimuli, moreover, where created by coherently manipulating both alignment and scaling characteristics - offers similar results to those discussed

and by keeping separated the stimuli created from different source utterances. The tasks were forced choices, and the two options were available by pressing two buttons on the keyboard.

1. Does it correct a preceding utterance?

Subjects listened twice to each stimulus, judging whether it could represent a correction to a previous utterance or just a neutral statement.

- Broad focus source: 9 subjects
- Contrastive focus source: 12 subjects

2. Question-answer context

Subjects listened to a question-answer sequence. The question was an out-of-the-blue question (thus a broad focus answer was expected to best suit it), and the answer was one of the manipulated stimuli. Subjects were asked to judge whether the answer suited the context, i.e. the question.

- Broad focus source: 11 subjects
- Contrastive focus source: 12 subjects

3. Is it a question?

This test was organized by exploiting a specific property of the contrastive accent, found in the dialogic exchange within dialogues [3]: speakers explicitly lacking information and uttering contrastive accents/patterns provoked a yes/no answer, as their statements were interpreted as a check of information by the interlocutor[2]. In the test, subjects were given the above mentioned context, they listened to each stimulus in isolation, and were asked to judge whether the utterance could be interpreted as a check of information and they could give a yes/no answer.

- Contrastive focus source: 13 subjects

2.3. Discrimination test

Discrimination tests are performed by having subjects hearing pairs of stimuli. In the case of studies involving tones, discrimination appears to be easier if the second stimulus in the pair is higher in pitch [6]. Thus, subjects listened to pairs of stimuli: AB, where B was either higher in pitch (or having a later peak), and AA pairs, as controls. Subjects were asked to judge whether the stimuli were the same or different.

- Broad focus source: 9 subjects
- Contrastive focus source: 9 subjects

3. Results

3.1. Identification test

When subjects perceive instances of two discrete sets, test results are expected to show an abrupt shift in the grouping of stimuli in one category or in the other [8].

Results of the identification test show that, independently from the task exploited for testing, subjects do not group stimuli showing an abrupt shift in the answers in favor of either one set or the other; nevertheless, at least for stimuli created from the contrastive source utterance, they appear to categorically interpret stimuli belonging to the extremes of the continuum for both alignment and scaling characteristics. This is in line with results on English [6]. Detailed results for the tasks exploited are given below.

1. Does it correct a preceding utterance?

- Broad focus and contrastive focus source

Results for the identification test performed with this task and stimuli from a broad focus source were discussed in [2], where it was shown that subjects do not show categorical perception behavior: their answers in favor of the contrastive interpretation (utterance suitable for correcting) ranged from about 0% to only 50%, depending on alignment and scaling characteristics. They gave more answers in favor of the contrastive interpretation for stimuli having a higher peak, but they did not even interpret categorically stimuli belonging to the extreme of the continuum (again, answers reached 50%).

In case of stimuli created from a contrastive focus source, coherently, nothing resembling categorical perception is found, although results are slightly more explicit and point to the existence of a source utterance effect. Figure 2 shows the mean of the subject answers in favour of the contrastive interpretation, calculated over the mean of all the three answers given by each subject to the single stimulus (value between 0 and 1). As the figure shows, subjects do not identify two discrete sets of stimuli depending on alignment steps, and they give a higher number of positive responses to stimuli having a higher peak. Nevertheless, no abrupt shift in the answers is found, depending neither on alignment nor on scaling. However, in this case, differently from what was observed above for broad focus source stimuli, subjects appear to perceive as different stimuli belonging to the extreme of the continuum for both alignment and scaling characteristics – see values highlighted by circles in figure 2. Notice that in the present study, in fact, only stimuli at the extreme of both the alignment and the scaling continuum were characterized by coherent correlates; other stimuli were always ambiguous at least in the case of one correlate, i.e. either alignment or scaling or both. Thus, in [6]'s terms, speakers categorically interpret the extremes of the continuum, but not through categorical perception.

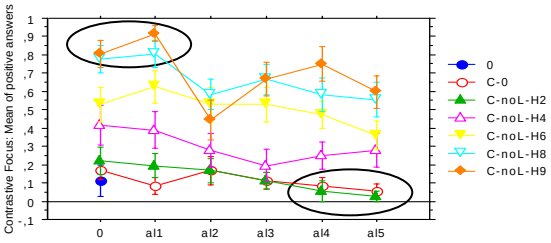

Figure 2: *Mean of all speaker answers in relation to stimuli manipulated as for the peak (and following low) alignment - x axis – and peak height – colors/forms.*

2. Question-answer context

- Broad focus and contrastive focus source

Figure 3 shows the mean of the subject answers in favor of a correct question-answer sequence, calculated over the mean of all the three answers given by each subject to the single stimulus (value between 0 and 1) – broad focus and contrastive focus source, upper and lower panel respectively. As the results show, subject answers do not change depending on the task: nothing resembling categorical perception is found, although answers are slightly more differentiated depending on pitch scaling for the contrastive focus stimuli –

---

the present paper [4]. This suggests that the results discussed here do not depend on the number of stimuli repetitions.

[2] The Map-task requires one speaker to give directions to the interlocutor for reconstructing a path. When the instruction-follower exploits a contrastive pattern/accent and does not aim at disagreeing with the instruction-giver, his utterance asks for confirmation and induces a yes/no answer; e.g.:

inf.giver: You have to go toward the lake
inf.follower: I have to go on the RIGHT.
inf.giver: Yes.

lower panel. As expected, subjects give fewer answers in favor of a correct question-answer sequence for stimuli showing a pitch height more coherent with the contrastive focus pitch accent characteristics (see answers for stimuli C-H8, C-H9). This points again to a source utterance effect.

3. Is it a question?

- Contrastive focus source

Although the idea was trying to focus on a difference such as the question-statement one, that in principle allows quite clear cut judgments, this specific task was particularly difficult because of the effort needed to keep in mind the right context where the contrastive pattern could be interpreted as a question – see 2.2. This may be one of the reasons why results show, again, that nothing resembling categorical perception is found nor a categorical interpretation of extremes is observed.

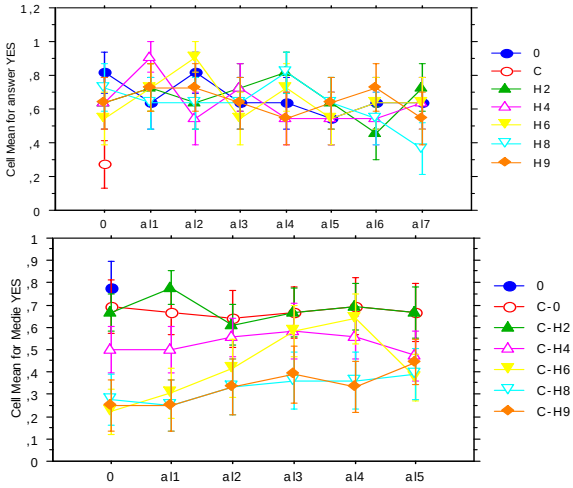

Figure 3: *Mean of all speaker answers in relation to stimuli manipulated as for the peak (and following low) alignment - x axis – and peak height – colors/forms. Broad focus source – upper panel – contrastive focus source – lower panel.*

3.2. Discrimination test

When the stimuli represent two different categories, test results are expected to show a greater distinction for pairs across perceptual boundaries [8]. Nevertheless, no discrimination is found, according to subject answers.

Figure 4, upper panel, shows the mean of all speaker answers in favor of 'different' when listening to pairs of stimuli differing for alignment and kept constant as for pitch scaling. There is no discrimination at all. Moreover, it has been proposed in the literature that reaction times are higher at perceptual boundaries [1]. When looking at reaction times– figure 4, lower panel – higher values are found for three pairs that may be considered as 'central' in the continuum, rather than at a more clear cut and abrupt perceptual boundary. This is similar to what observed by [7] for German.

4. Discussion and conclusions

The Categorical Perception paradigm was followed in order to check whether there is categorical perception of a broad focus and a contrastive focus pitch accent in Pisa Italian. Independently from the task exploited for the identification test, subjects did not show behaviors coherent with categorical perception; nevertheless, at least for stimuli created from the contrastive source, they appear to categorically interpret stimuli belonging to the extremes of the continuum for both alignment and scaling characteristics. This is in line with previous observations on English accents [6].

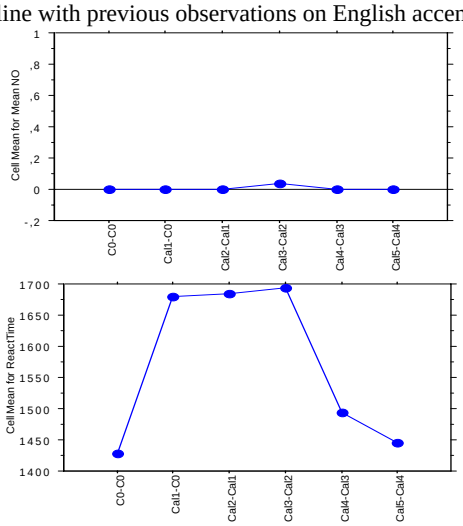

Figure 4: *Mean of all speaker answers in favor of 'different' – y-axis, upper panel – and mean of their reaction times – y-axis, lower panel.One AA and relevant AB pairs on the x-axis.*

Coherently, no discrimination is found, although reaction times could indicate a greater cognitive effort in judging some of the pairs (in line with [7]). However, apart from stimuli at the extreme of both the alignment and the scaling continuum, stimuli in this study were always ambiguous as for at least one correlate, i.e. either alignment or scaling or both. For this reason, the present study may be limited by the fact that, for instance, only one correlate at a time was changed. Nevertheless, limits could be connected to the Categorical Perception procedure for testing intonation, as already observed by others (cfr.[7]). In this respect, it looks particularly interesting to mention that subjects asked to imitate - along the lines of [9] - the same stimuli used for the present study, appear to be able to produce patterns belonging to discrete sets [4]. Therefore, the test may suggest that, in case of intonation patterns that are particularly complex to compare, the CP paradigm alone may not be fair enough.

5. References

[1] Chen A. 2003. Reaction Time as an Indicator of Discrete Intonational Contrasts in English. In *Prc. Eurospeech'03*,97-100.

[2] Gili Fivela, B., 2006. 'Scaling' e allineamento dei bersagli tonali: l'identificazione di due accenti discendenti. In *Proceedings of AISV*, Salerno, December 2005, 214-232

[3] Gili Fivela B. 2004. The phonetics and phonology of intonation: The case of Pisa Italian. PhD Diss., Scuola Normale Sup., Pisa.

[4] Gili Fivela B. In prep. Perceiving and imitating: Testing categories in intonation.

[5] Kohler K. 2006. Paradigms in Experimental Prosodic Analysis: From measurements to function. In *Methods in Empirical Prosody Research*, Sudhoff etal.(ed)W.deGruyter,Berlin,123-152

[6] Ladd B., Morton R. 1997. The perception of intonational emphasis:Continuous or categorical? *J.of Phonetics* 2, 313-342.

[7] Schneider K.,Linfert B.,Dogil G.,Möbius B., 2006. Phonetic grounding of prosodic categories. In Methods in *Empirical Prosody Research*,Sudhoff etal.(eds)W.deGruyter,Berlin,33-362.

[8] Repp B. 1984. Categorical perception: Issues, methods, and findings. In Lass N. (ed.) *Speech and Language: Advances in Basic Research and Practice*. Academic Press,10,243-335.

[9] Pierrehumbert J., Steele S.A. 1989. Categories of tonal alignment in English. *Phonetica* 46, 181-196

[10] Vanrell M. 2006. A scaling contrast in Majorcan Catalan interrogatives. In *Proc. of Speech Prosody*, Dresden, Germany.