Categorical Perception of intonational contrasts in European Portuguese

Isabel Falé[1,2], *Isabel Hub Faria*[1]

[1]Onset – CEL, Lab. Psicoling, DLGR, FLUL
[2]Universidade Aberta
Lisboa, Portugal
imsfale@netcabo.pt; ihfaria@fl.ul.pt

Abstract

European Portuguese (EP) intonational contrast between statement and question contours was tested on a Categorical Perception based paradigm. From two natural sentences one produced by a male speaker and another by a female, one multi-step *continuum* from each sentence was created, from declarative to question contour, through acoustic manipulation (PSOLA) and submitted to 20 EP listeners that performed two tasks: an identification and a discrimination task.

For the identification test, subjects had to categorize each presented stimulus. In addition to response data, reaction times of the identification task were also collected.

For the discrimination test, subjects were presented with an AX discrimination task and had to decide whether the stimuli in each pair were equal or different. Experimental design and procedures were developed with *E-Prime*.

Identification results confirmed that the contrast is indeed categorical. However, identification reaction times measurements point to continuous rather than categorical perception. The absence of a consistent peak of discrimination in the crossover between categories supports the continuous perception view.

1. Introduction

In European Portuguese, like in many other languages [1, 2], there is an intonational distinction between statements and questions [3, 4]. Although we can find several differences along the sentences, this intonational contrast is mainly due to the movements of F0 at its end [5]. Statements usually end with an abrupt fall in intonation that starts at the vowel that precedes the last stressed vowel (PRUVT) and continues on the last stressed vowel (UVT), where relevant micro-prosodic movements can be found, till the end of the sentence. Questions are said to end up with a strong F0 rise movement that begins in the last stressed vowel (UVT) and follows to the last phonetic segment (either a vowel or a voiced consonant). Production studies and phonological analysis classify this contrast as categorical, but that assumption has never been tested from a Categorical Perception (CP) view.

In order to study the categorical nature of this intonational distinction, we developed an experimental procedure based on the Categorical Perception paradigm [6]. CP paradigm, primarily developed to test the perception of segmental features, involves two kinds of tests: an identification test, where stimuli have to be classified in categories, and a discrimination test, where pairs of stimuli have to be assessed as identical or different. It is assumed that perception is categorical if the peak of discrimination coincides with the category boundary revealed by the identification test. Subjects are believed to better distinguish differences between categories than within categories.

Several studies have already adopted the CP paradigm to the study of intonation contrasts, mainly boundary tones contrasts [7, 8, 9], with encouraging results. It has been suggested that the use of reaction times measures, along with CP paradigm, could help to decide whether speech perception effects are categorical or continuous: if an increase in response's reaction time corresponds to an increase of the ambiguity of the stimulus, perception is rather continuous than categorical [10]. Reported data from discrimination tasks [7, 9] have shown results, which are not in accordance with the classical CP definition, suggesting that, although this paradigm can be used to test intonational contrasts, it should be modified to better explain these data. So, we developed two different strategies, one for identification and another for discrimination. The first strategy was to collect reaction-time measurements of identification responses and the second one was to present a discrimination task, that instead of just confirming or not identification results could shed light to some other questions such as: do people equally distinguish pairs of stimuli at all frequency ranges? Are discrimination patterns maintained when frequency differences between pairs of stimuli increase? How far pairs of stimuli must be, in order to be consistently understood as different?

2. Method

2.1. Stimuli

In a previous perception test [11], sentences '*O dador gostava da rapariga.'* and *'O dador gostava da rapariga?'*, produced by an European Portuguese male speaker, had been identified, by a sample of forty European Portuguese native speakers, as a declarative (92.5%) and as a question (97.5%), respectively. The intonation patterns of both sentences were acoustically analysed. Semitone averages were calculated for each observation point in the sentence (AT, VT, VT1, PRUVT, UVT and F). A new sentence, based on the declarative one, was generated with the average values using the PSOLA [12] resynthesis, available in Praat speech analysis and resynthesis software [13].

Further, both the last stressed vowel and the final vowel of the sentences were also manipulated. The last stressed vowel was maintained with a steady and flat F0, which means that all micro-prosodic variation, usual in this vowel, was kept aside. For the pitch of the final vowel, our main variable, a 15-step *continuum* was created, with equidistant points of one

semitone (st) along the semitone scale, from 2 to 16 semitones (see figure 1). From the end of the last stressed vowel till the beginning of the last vowel, pitch values were completed through phonetic interpolation provided by PSOLA resynthesis.

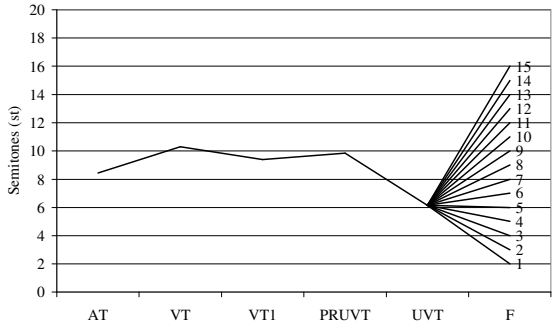

Figure 1*: Sentence Group A - 15-step continuum diagram*

Similar procedures were adopted for sentences '*O nadador gostava de águas calmas.'* and *'O nadador gostava de águas calmas?',* produced by an European Portuguese female native speaker. Both sentences had scored 95% in the previous perception test, being identified as a statement and as a question, respectively. An 11-step *continuum* was created with equidistant points of one semitone (st) along the semitone scale, from 3 to 14 semitones (see figure 2). The different number of steps of the *continua* is related to F point tonal extension produced by each speaker in these specific sentences.

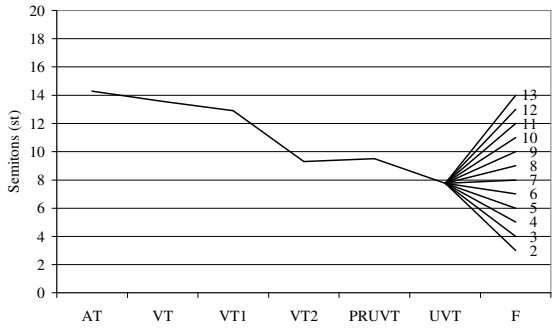

Figure 2*: Sentence Group B - 11-step continuum diagram*

Discrimination experiments failure has been reported with long stimuli due to auditory memory limitations [10]. In order to avoid this kind of problem, we have decided to shorten our stimuli to the part where variation occurs. We have also transformed these shortened sentences in *hummed* ones, using Praat, in order to prevent stimuli oddness. These cut and *hummed* sentences were paired in AB and BA orders, with an interval of 500 milliseconds between each sentence. A schema was built to select the possible combinations of pairs of stimuli that could be the most informative ones for our study purposes.

2.2. Tasks and experimental procedures

Subjects performed two tasks: an identification task and a discrimination task.

Identification task

For the identification test, a two-forced choice task was presented to the subjects, who had to identify each step of the *continuum* as a statement (declarative) or as a question. Each stimulus was repeated eight times in random order.

Subjects were instructed to listen to each stimulus and to proceed immediately to its classification (statement or question) by pressing a computer key as quickly as possible.

In order to collect reaction time data, the experimental procedure was developed in *E-Prime* [14]. Stimuli were auditorily presented through the computer, via headphones. Responses and the corresponding reaction times were registered through the computer keyboard. All technical tests were previously performed to guarantee adequate data collection quality by the computer.

Discrimination task

In an AX discrimination test, subjects had to decide if the pair of stimuli they had listened to was either equal or different stimuli. Stimuli were presented in AB e BA orders, in which A corresponds to the stimulus with the lowest pitch in the pair. The stimulus set was presented in random order.

Twenty pairs of stimuli with varying semitone differences (from 1 to 14) for Sentence Group A and eighteen pairs of stimuli (ranging from 1 to 11 st) for Sentence Group B, according to the discrimination schema, were repeated five times to subject evaluation.

Subjects

Twenty European Portuguese native speakers (10 female), aged between 27 and 44, with no history of hearing or language deficits or disorders, participated in the experiments. All, except one, had a graduate degree. Both experiments were run individually in one session.

3. Results

3.1. Identification test

3.1.1. Sentence Group A

Classification

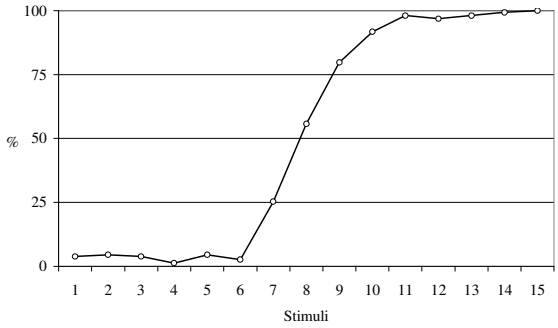

Figure 3*: Group A - 'Question' percent responses as a function of stimulus step number (averaged over subjects)*

Averaged classification results for all subjects on the identification task (figure 3) show a clear S-shaped curve. A full crossover from less than 20% to more than 80% is reached in three steps. These results suggest the presence of two

distinct categories: the range between stimuli 1 to 6 corresponds clearly to the 'declarative' category, while the range between stimuli 10 to 15 to the 'question' category. As we can see (figure 3), identification values within each category are high and stable. The shift from one category to another occurs in the range of stimulus 7 to 9.

Reaction Times

Averaged reaction time (RT) responses for all subjects on the identification task (figure 4) show a clear increase, especially on stimuli 7, 8 and 9. In all, except these stimuli, RT values are under 2300 milliseconds (msec), stimuli 12 to 15 recruiting the lowest RT measurements, below 2104 msec. RT values for stimuli 7 to 9 occur between 2400 and 2600 msec. RT values include stimulus duration.

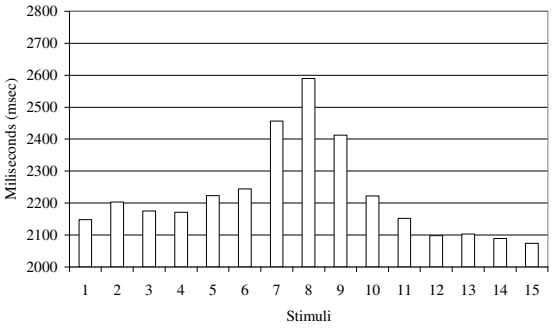

Figure 4*: Group A -Reaction time responses as a function of stimulus step number (values are averaged over subjects)*

The ANOVA (one factor) analysis of the reaction time results points as well to significant statistic differences (F= 26.341, p= 0.000) between stimuli. The *pos-hoc* Tukey test shows that stimuli 7, 8 and 9 are statistically different from the rest (p= 0.005).

3.1.2. Sentence Group B

Classification

Classification results for this group of sentences also exhibit a S-Shaped curve. A full crossover from less than 10% to more than 80%, with 4 steps, is observed.

Two distinct categories emerge from data: 'declarative' category lies between stimuli 2 to 8 and 'question' category between stimuli 11 to 13. Category shift takes place in the range of stimuli 9 to 10.

Reaction Times

Averaged reaction time (RT) results for all subjects on the identification task (figure 5) reveal an increase, specifically on stimuli 8, 9, 10 and 11. Like in the Sentence Group A, in all except the referred stimuli, RT average values are under 2300 milliseconds. The lowest average RT measurements can be found in stimuli 2, 3, 4, 5 and 6, which achieve values less than 2200 msec. RT values for stimuli 8 to 11 range between 2300 and 2400 msec.

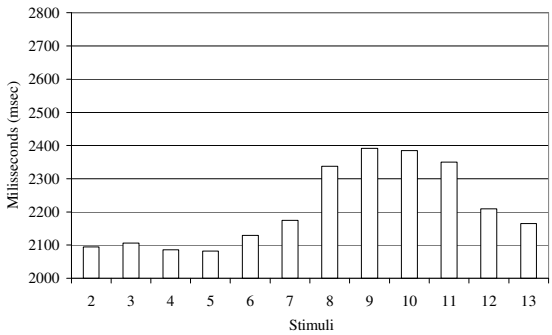

Figure 5*: Group B - Reaction time responses as a function of stimulus step number (values are averaged over subjects)*

The ANOVA (one factor) analysis confirms the presence of significant statistic differences (F= 12.803, p= 0.000) between stimuli RT values. The *pos-hoc* Tukey test opposes two groups of stimuli: one group is constituted by stimuli 2, 3, 4, 5, 6 and 7 and the other one by stimuli 8, 9, 10 and 11 (p= 0.005).

3.2. Discrimination test

A persistent order of presentation effect reported in previous studies [7, 9] was also present in our data: pairs of stimuli with BA order were worse discriminated than those with AB order.

As expected, pairs of stimuli with greater semitones differences between each member are better recognized. In averaged terms, AB ordered stimuli with a difference of 3 semitones are consistently perceived (that means that it is detected by, at least, 50% of our sample) in Sentence Group A, while in Sentence Group B this result is reached with a difference of 2 semitones.

Also, there is no asymmetry related to pitch location of pairs of stimuli. Indeed, we didn't find discrimination differences in pairs of stimuli placed in different points of the pitch range of the *continuum*. In particular, our results do not show a peak of discrimination in the crossover of categories revealed by the identification test.

4. Discussion and Conclusions

We strongly believe that the results from the identification task support our main hypothesis that statement and question contours are two distinct intonational categories in EP. Subject responses show a clear change between categories that takes place between the 7 to 9 stimuli interval in Sentence Group A and between the 9 to 10 stimuli interval in Sentence Group B. These data are supported by reaction-time results that present an increase in the answers to stimulus 7 to 9 in Sentence Group A and also in stimulus 9 to 10 of Sentence Group B, corresponding to the extra-time subjects needed to decide when a categorically ambiguous stimulus belonged to one of the categories. There was also an increase in RT values for stimulus 8 and 11 in Sentence Group B

In contrast, unambiguous stimuli reached high levels of identification associated with lower reaction times.

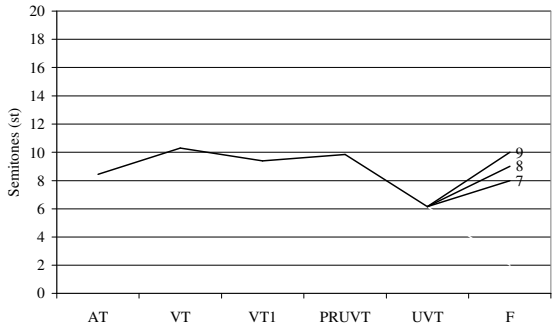

Figure 6*: Sentence Group A - Ambiguous stimuli diagram*

Ambiguous stimuli 7 to 9 (figure 6) from Sentence Group A ended with rises that were not yet perceived as questions and that were no longer declaratives. The same was observed in Sentence Group B with ambiguous stimuli 9 to 10.

This result suggests that the amplitude and location of the rise in the pitch range is determinant for the percept of the category 'question'. It seems that it is not enough to end up in a rise to be a 'question'. In fact, final rise must be, at least, higher than 2 st (counting from the last stressed vowel) to be understood as a clear 'question'. Phonological analyses that propose the presence of a high tone associated with the final boundary (H%) in EP are reinforced.

On the other hand, the last stimulus of the *continuum* categorized as a clear 'declarative' was stimulus 6 in Sentence Group A and stimulus 8 in Sentence Group B both ending with a slight rise (1 st) from the last stressed vowel. There seems to be no need for a final lowering after the fall from PRUVT to UVT for sentences to be perceived as declaratives. Data shows that stimuli with final falls achieve the same level of identification as 'declarative' than stimuli with flat or with a little rise. Whenever the final movement rises 2 st from the last stressed vowel, it becomes ambiguous between a 'declarative' and a 'question' category. These data are consistent with phonological analyses that argue against the presence of a low tone associated with final boundary (L%) in EP declarative neutral sentences.

As stated before, our results have shown no discrimination peak in the crossover of categories. In CP definition terms, this is enough to reject categorical perception. However, previous identical studies [9] already mention the absence of a peak of discrimination suggesting the need of a CP paradigm revision in order to better describe intonational contrasts study.

In the developed discrimination task, we only had a few pairs of stimuli under classical CP paradigm circumstances, i.e. pairs of stimuli that differ from one another by just one step (in this case, 1 st). Discrimination values for these pairs were very low, under 26%, either on pairs that occur in the same category or on pairs that occur in the crossover of the categories. Once again these values are quite different, much lower, from those reported in other studies. Two things cannot be dismissed from our analysis: the first one is that 1 st difference is rather difficult to detect in this context; and the second one is the fact that *hummed* stimuli may have an effect of lowering the discrimination values. In fact, consistent discrimination (50%) is only reached in pairs of stimuli with 2 or 3 st of difference. Also discrimination is not homogeneous within each group of stimuli pairs, which according to CP would be expected if perception was continuous. The analysis of reaction times associated with the identification data allows an explanation of continuous perception for this intonational contrast. Ambiguous stimuli cause gradual increase of reaction times that reflects the extra-time processing needs for the answer: the stronger the ambiguity, the higher the reaction-time values. We believe this is an evidence of a continuous perception model of intonational contrasts.

Our analysis was based on averaged results over subjects. However, we know that in psychometric studies individual variation is to be taken with caution. Therefore, a detailed analysis of individual behavior will follow this study.

5. References

[1] Ladd, R., *Intonational phonology*. Cambridge University Press, Cambridge, 1996.

[2] Hirst, D., Di Cristo, A., *Intonation systems. A survey of Twenty Languages*. Cambridge University Press, Cambridge, 1998.

[3] Viana, M.C., *Para a síntese da entoação do português*. PhD Dissertation, University of Lisbon, 1987.

[4] Frota, S., "Nuclear falls and rises in european portuguese: a phonological analysis of declarative and question intonation", *Probus*, vol. 14 (1), 113-146, 2002.

[5] Mata, A.I., Pereira, I., "Flexibilidade de contornos entoacionais em sequências de natureza interrogativa: percepção e interpretação", *Actas do VII Encontro Nacional da Associação Portuguesa de Linguística*, 211-224, Lisboa, 1991.

[6] Liberman, A.M., Harris, K.S., Hoffman, H.S. and Griffith, B.C., "The discrimination of speech sounds within and across phoneme boundaries", *Journal of Experimental Psychology* 61, 379-388, 1957.

[7] Remijsen, B., van Heuven, V., "Categorical pitch dimensions in dutch: diagnostic test", *Proceedings of the ICPhS*, 1865-1868, San Francisco, 1999.

[8] Ladd, R., Morton, R., "The perception of intonational emphasis: continuous or categorical?", *Journal of Phonetics* 25, 313-342, 1999.

[9] Schneider, K., Linftert, B., "Categorical perception of boundary tones in German". *Proceedings of the 15th ICPhS*, Barcelona, 1-4, 2003.

[10] Massaro, D., *Speech perception by ear and eye: a paradigm for psychological inquiry*, Hillsdale, New Jersey, London, 1987.

[11] Falé, I., *Reconhecimento de Padrões Entoacionais em Português Europeu*. PhD Dissertation, University of Lisbon, 2005.

[12] Moulines, E., Charpentier, F., "Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones". *Speech Communication* 9, 453-467, 1990.

[13] Boersma, P., Weenink, D., *Praat: doing phonetics by computer* (version 4.2.21), Computer Program, 2004.

[14] Schneider, W., Eschman, A., Zuccolotto, A., *E-Prime reference guide*. Psychology Software Tools, Pittsburgh, 2002.