# Time course of Intonation Processing in European Portuguese: a Gating paradigm approach

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

[1]Onset-CEL, Lab. Psicolinguística, Universidade de Lisboa; [2]Universidade Aberta

**ABSTRACT**

EP listeners seem to pay attention to phonetic detail specificities of the speech signal, which allow them to start building their internal representation of the intonation contour very early in the sentence.
Using the Gating Paradigm methodology (Grosjean 1980, 1996), 20 European Portuguese (EP) native speakers were presented with auditory speech stimuli gated in specific sentence locations, which they had to classify within a category and after it to rate their own answer confidence level.
The influence and weight of lexical cues, especially *wh*-words, for early sentence type recognition was considered.
Results showed that the lexical cues power for sentence type processing is stronger than the prosodic ones. Sentences starting with *wh*-words, whatever semantic or syntactic function or prosodic information they have tend to be classified as questions. The interplay of all cues seems to occur later in the sentence, after more information becomes available.

## 1. INTRODUCTION

Prediction is a very important issue in speech processing and it has effects on verbal communication accuracy: it prevents silent gaps on speech interaction and speeds speech processing, diminishing the interpretation hypotheses and leaving the listener free to perform other on-going cognitive processes. For prediction to occur, linguistic information must be available early in the speech signal so that processing may start immediately.

Until recently in European Portuguese (EP), most relevant intonation movements were believed to occur locally, at the end of the sentences [1, 2, 3]. This would imply that EP listeners would have to wait till the end of an utterance in order to interpret it as a sentence type. We would also expect to find frequent silent gaps within verbal interaction, corresponding to the sentence intonation processing which would only begin after sentence end. In this study, as well as in some previous ones we will show that this not the case [4].

Recent studies have been reporting the importance of phonetic detail for speech perception and processing. These studies argue that listeners process phonetic detail [5] available in the speech signal to start building the internal representation of the linguistic entities. This assumption is in accordance with the prediction phenomenon.

Our main hypothesis is that EP speech signal has early prosodic information that provides listeners with enough data to access and recognize sentence type before its end.

Another hypothesis is related to the possible lexical cues interference in the course of speech processing. In a previous EP perception study in this language [4], lexical cues didn't show to be as important as researchers think they are. In fact, subjects didn't achieve better identification and classification scores when, for instance, sentences started with what was thought to be strong lexical cues to specific interpretation. So, in order to get a better understanding of the way lexical cues interact with prosodic ones, we have studied the interplay between these cues, assuming that if lexical cues restrict speech processing then listeners will only need to have access to lexical data to be able to interpret and to classify sentences within categories.

## 2. METHOD

For hypothesis testing, we have chosen the linguistic intonation contrast between declaratives and interrogative sentences in EP reported by [1, 2, 3] using the *Gating Paradigm* methodology [6, 7], which was initially developed for word recognition.

In this experimental paradigm, target words are cut into smaller pieces of increasing duration. For instance, words can be gated with a fix interval of 30 msec: the first piece will have 30 msec, the second 60 msec, the third 90 msec and so on till the end of the stimulus. These gated stimuli are presented to subjects in a sequential (increasing duration) order and they have to guess which the word is. Just after that, subjects have to rate their confidence level on their own answer to the identification or guessing task.

Previous works within this paradigm have selected the temporal criterion for stimulus gating. However, most of these studies work at lexical word level. This criterion was transposed to bigger units of grammatical analysis [7].

Based on the assumption that stressed vowels in EP are strong informative prosodic units in an intonation contour [1], we developed a phonetic criterion to gate our sentences instead of the temporal one. Sentences were gated at the end of a stressed vowel within a word in the sentence.

The identification point of a sentence is a function of two independent measures: *isolation point (IP)* and *recognition point (RP)*. *Isolation point* refers to the location on the stimulus where correct identification is achieved and maintained over fragments of the same sentence till its final fragment. Correct identification, however, does not inform us on how confident subjects are on their answers. The rating of the answers confidence level by subjects will determine the *recognition point*. *Recognition point* is then reached when a stimulus is first rated as 'sure' and this rating is not changed till the end of the sentence.

### 2.1. Stimuli

For early prosodic identification hypothesis research, five pairs of stimuli (10 sentences) with variable duration were gated in diverse locations. In each pair of stimuli (A, B, C, D, E) there were two segmental identical sentences, both produced by the same speaker that differed in their intonation contour: one had question contour and the other a statement one. Both sentences in the pair were cut in the same locations. (See Table 1, for an example of a sentence gating).

| Sti | Gate | Sentence Fragment |
|---|---|---|
| 1 | FSV | Os golfi |
| 2 | SV1 | Os golfinhos cinzen |
| 3 | SV2 | Os golfinhos cinzentos anima |
| 4 | SV3 | Os golfinhos cinzentos animaram o jogador |
| 5 | SV4 | Os golfinhos cinzentos animaram o jogador e o adi |
| 6 | LST=SF | Os golfinhos cinzentos animaram o jogador e o adido naval |

**Table 1.** *Example of sentence gating*

For lexical cues hypothesis testing, sentences and sentence fragments beginning with *wh*-words (9 sentences in three sentence groups), either inducing question or statement interpretation, were selected. Part of these sentences introduced true *wh*-questions and another part was constituted by fragments initiated by *wh*-words that, although phonetically equal to true *wh*-words, have different semantic and syntactic functions and diverse prosodic features that we expected to be enough to block immediate question interpretations (See Table 2, for an example of the sentences used to test this hypothesis).

| Sentences – stimuli | Types |
|---|---|
| [O Pedro viu] quem (*who)* animou os golfinhos. | I |
| Quem *(Who)* animou os golfinhos [fez um bom trabalho]. | II |
| Quem *(Who)* animou os golfinhos? | III |

**Table 2.** *Example of stimuli with wh-words*

Type I stimuli refer to sentence fragments beginning with *wh*-words that were extracted from medial position in the original statement sentence. Type II stimuli were extracted from statement sentences beginning with a *wh*-word. Type III stimuli were true questions starting with *wh*-words.

### 2.2. Tasks and experimental procedures

Sentence fragments with increasing duration determined by stressed vowels location were presented sequentially to subjects that had to perform two tasks. First, they had to identify the fragment with a sentence type in a two-choice forced task, where possible answers were ‘statement’ and ‘question’. Secondly, they had to rate their confidence level for the given answer on fragment identification. A simple two-scale was available for this purpose: answers could be rated either as ‘sure’ or as ‘unsure’.

Subjects were instructed to listen to each stimulus and to proceed to its classification (statement or question) by pressing a computer key. Just after it, they had to rate the confidence level of their own answer.

The experimental procedure was developed in *E-Prime* [9]. Stimuli were auditory presented through the computer, via headphones. Responses were registered through the computer keyboard. All technical tests were previously performed to guarantee adequate data collection quality by the computer.

**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. Experiment was run individually in one session.

## 3. RESULTS

### 3.1. Early prosodic identification hypothesis

*3.1.1. Isolation Point*

Graphs in figures 1-5 present *IP* results’ for the 5 pairs of stimuli for all subjects. In graphs, black bars are for statement sentences and white bars to questions. Gates legend is the following: FSV - first stressed vowel; SVx - stressed vowel x; LSV - last stressed vowel; and SF - sentence final.

In pairs A, B, D and E, the last stressed vowel occurs in the final syllable of the sentence, so the last stimulus gate for these sentences was also its ending. In pair C, the last stressed vowel occurs in the penultimate syllable of the sentence, so there was an extra stimulus gate to reach the end: SF.

In statements, *IP* is achieved in the very first stressed vowel (FSV) for all pairs (see figures 1-5). By contrast, in questions, *IP* tends to occur later in the sentence, near its end. The exception goes to the question in pair C which is 'isolated' sooner (figure 3).

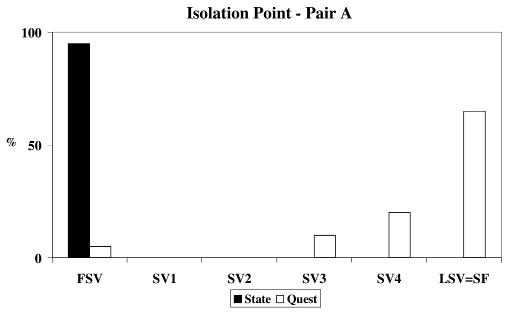


**Figure 1.** *Percent results for IP in pair A (all subjects included)*

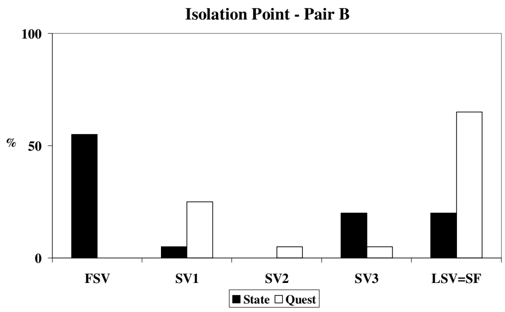


**Figure 2.** *Percent results for IP in pair B (all subjects included)*

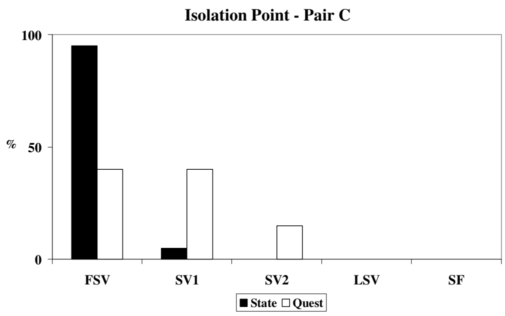


**Figure 3.** *Percent results for IP in pair C (all subjects included)*

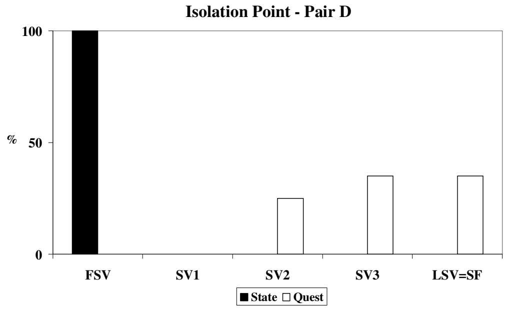


**Figure 4.** *Percent results for IP in pair D (all subjects included)*

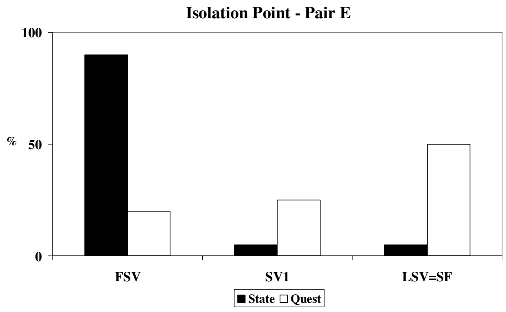


**Figure 5.** *Percent results for IP in pair E (all subjects included)*

*3.1.2. Recognition Point*

These results are more disperse than the *IP* ones (figure 6). In general, *RP* occurs later than *IP* (Table 3) which is explained by the different processing nature of these tasks. The processing time for decision making in the *IP* task is shorter than for the *RP* task because *RP* requires an explicit memory recalling that takes longer.

There is, however, an exception in the pair B question, where *RP* appears sooner than *IP*. Subjects tend to wait for more intonation data availability in order to decide whether they are sure or not about their own identification answer.

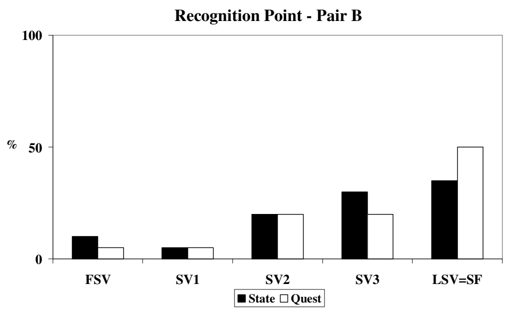


**Figure 6.** *Percent results for RP in pair B (all subjects included)*

There is no systematic relation between the distribution of *IP* and *RP* results. However, there is a tendency for statements to be recognized earlier (Table

3). At first we thought that this tendency could be a *default* answer effect but we observed that this was not what happened in other circumstances, for instance, when the sentences begin with *wh*-words. So, we didn't consider it to be a *default* answer.

Locations for the *IP* and *RP* on Table 3 were the result of the sum of percent data: whenever the percent data sum reached at least 50% (above chance level), we have considered that the *IP* or the *RP* of the sentence had been achieved and registered the gate in the sentence location where that happened.

| | Statement | | Question | |
|---|---|---|---|---|
| | ***IP*** | ***RP*** | ***IP*** | ***RP*** |
| **A** | FSV | SV1 | LSV=SF | LSV=SF |
| **B** | FSV | SV3 | LSV=SF | SV3 |
| **C** | FSV | SV2 | SV1 | SV2 |
| **D** | FSV | SV2 | SV3 | LSV=SF |
| **E** | FSV | SV1 | LSV=SF | LSV=SF |

**Table 3.** *Location results for IP and RP. Shadowed cells refer to sentence end (all subjects included)*

### 3.2. Lexical cues hypothesis

*3.2.1. Isolation Point*

*IP* results' for the 3 groups of sentences are available in graphs of figures 7-9. In graphs black bars represent Type I stimuli, white bars Type II and gray bars Type III.

In Type I and Type III stimuli from Sentence Group 1 (figure 7), the *IP* is located in the First Stressed Vowel (FSV).

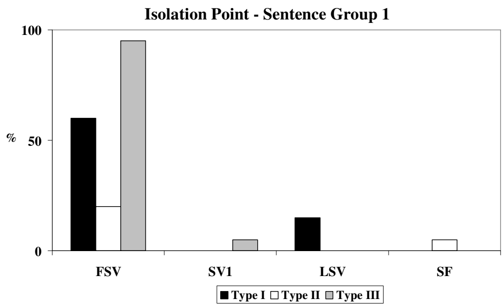


**Figure 7.** *Percent results for answer 'question' IP in Sentence Group 1 (all subjects included)*

In Type I, Type II and Type III stimuli from Sentence Group 2 (figure 8), the *IP* is also located in the FSV.

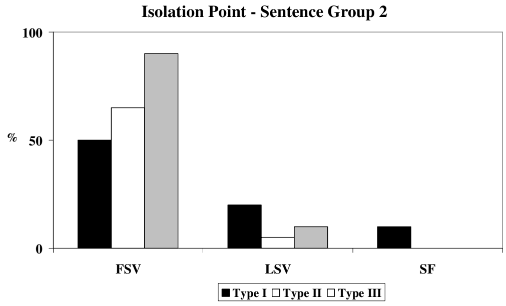


**Figure 8.** *Percent results for answer 'question' IP in Sentence Group 2 (all subjects included)*

The location for *IP* in Sentence Group 3 (figure 9) is only available for Type III stimuli and it relies on the FSV.

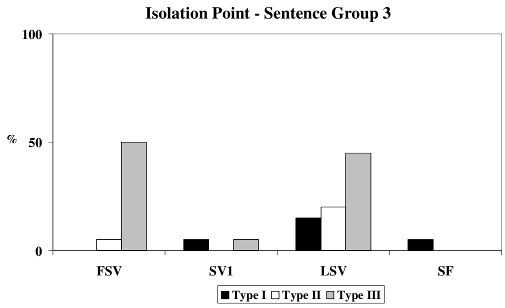


**Figure 9.** *Percent results for answer 'question' IP in Sentence Group 3 (all subjects included)*

*3.2.2. Recognition Point*

Like in the analysis of the first hypothesis, the results for *RP* are more diverse than *IP* ones. However, they are clearer because most of the *RP* results available are located in the Last Stressed Vowel (LSV), except for Sentence Group 1 (figure 10) in which they occur in the Stressed Vowel 1 (SV1).

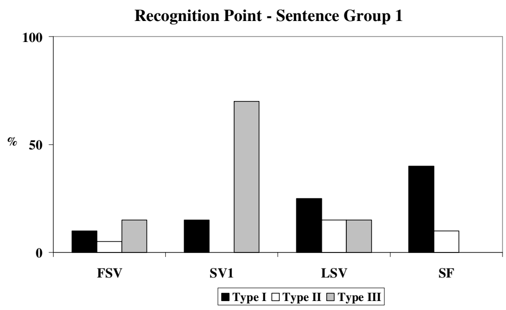


**Figure 10.** *Percent results for answer 'question' RP in Sentence Group 1 (all subjects included)*

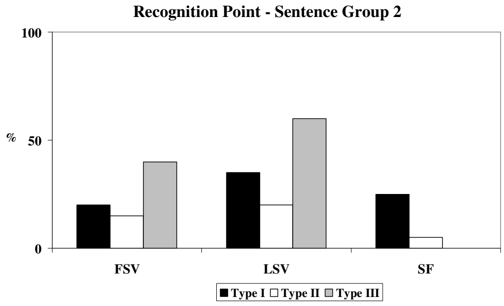


**Figure 11.** *Percent results for answer 'question' RP in Sentence Group 2 (all subjects included)*

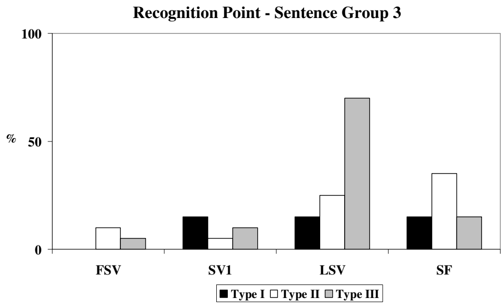


**Figure 12.** *Percent results for answer 'question' RP in Sentence Group 3 (all subjects included)*

We didn't detect any symmetry between *IP* and *RP* results as can be seen in Table 4.

| | Type I | | Type II | | Type III | |
|---|---|---|---|---|---|---|
| | ***IP*** | ***RP*** | ***IP*** | ***RP*** | ***IP*** | ***RP*** |
| **SG1** | FSV | LSV | --- | --- | FSV | SV1 |
| **SG2** | FSV | LSV | FSV | --- | FSV | LSV |
| **SG3** | --- | --- | --- | SF | FSV | LSV |

**Table 4.** *Location results for IP and RP 'question' answer. Shadowed cells refer to sentence beginning (all subjects included)*

## 4. DISCUSSION AND CONCLUSIONS

Statements are identified at the first stressed vowel in the sentence while questions (except the question in pair C) are identified by the last or the penultimate stressed vowel of the sentence.

A detailed acoustic analysis of the fundamental frequency of the question in pair C shows that its intonation contour presents specificities that facilitate its earlier identification when compared with the questions in the other pairs: an initial high and steady F0 plateau that points to a high pitch range, usual in questions. The availability of this early prosodic cue in the sentence allows for sooner identification.

As expected, subjects take longer to be confident on their own answers: recognition points for every sentence occur later than isolation ones, except for the question in pair B.

The fact that listeners are able to identify declarative sentence type by the end of the first stressed vowel of the sentence means that prosodic cues are available in the speech signal at this early location. These prosodic cues can be, for instance, the pitch peak of the sentence that allows subjects to place the boundaries of pitch range that is a global phenomenon. However, question results refer to the need of waiting for sentence end to identify interrogative sentence type. As we know, it is a final local rising movement that usually distinguishes questions from other sentence types. Therefore, both global and local data play an important role in intonation processing.

Both *IP* and *RP* results' reveal that EP listeners start to build their internal representation of intonation very early in the sentence.

Results also showed that the lexical cues power for sentence type processing is stronger than the prosodic ones because sentences starting with *wh*-words, whatever semantic or syntactic function or prosodic information they have tend to be classified as questions at first (see *IP* results in table 4). However, Sentence Group 3 stimuli results do not show the same behavior. We believe that this may be due to the different strength of interrogativity of the words. Probably, the *wh*-word that starts these stimuli is less 'marked' for question interpretation and, at this stage, the prosodic cues are enough to reject the classification as question. So, it seems that the presence of *wh*-words that have a heavy question interpretation allows for a question classification even without a proper question intonation in the sentence. By the contrary, if the lexical strength for interrogativity is poor, like in Sentence Group 3 stimuli, prosodic data became stronger.
The interplay of all cues seems to occur later in the sentence, after more information becomes available.

## 5. REFERENCES

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

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

[3] 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.

[4] Falé, I., *Percepção e Reconhecimento da Informação Entoacional em Português Europeu*. PhD Dissertation, University of Lisbon, 2006.

[5] Hawkins, S., "Contribution of fine phonetic detail to speech understanding", *Proceedings of the 15th ICPhS*, Barcelona, 293-296, 2003.

[6] Grosjean, F., "Spoken word recognition processes and the gating paradigm", *Perception and Psychophysics* 28 (4), 267-283, 1980.
[7] Grosjean, F., Hirt, C., "Using prosody to predict the end of sentences in English and French: normal and brain-damaged subjects", *Language and Cognitive Processes* 11 (1/2), 107-134, 1996.
[8] Grosjean, F., "Gating", *Language and Cognitive Processes* 11 (6), 597-604, 1996.
[9] Schneider, W., Eschman, A., Zuccolotto, A., *E-Prime reference guide*. Psychology Software Tools, Pittsburgh, 2002.
[10] Falé, I., Faria, I.H., "A glimpse of the time course of intonation processing", *Proceedings of the 9th European Conference on Speech Communication and Technology*. Lisboa, 2377-2380, 2005.