

# How can you use disfluencies and still sound as a good speaker?

Helena Moniz<sup>1,2</sup>, Ana Isabel Mata<sup>2</sup>, Isabel Trancoso<sup>1,3</sup>, M. Céu Viana<sup>2</sup>

<sup>1</sup> L<sup>2</sup>F, INESC-ID, Spoken Language Systems Laboratory, Lisbon, Portugal; <sup>2</sup> CLUL / FLUL, University of Lisbon, Lisbon, Portugal; <sup>3</sup> IST, Technical University of Lisbon, Lisbon, Portugal  
{helenam, imt}@l2f.inesc-id.pt, aim@fl.ul.pt, mcv@clul.ul.pt

## Abstract

This paper explores the results of a previous experiment concerning listeners' ratings of different types of (dis)fluencies and extends the analysis of such phenomena to a corpus of university lectures. Results suggest that, although not all disfluency types are equally tolerated by listeners, such differences may be overridden by an adequate control of tonal scaling and pause length, at least.

**Index Terms:** disfluencies, prosody, fluency ratings.

## 1. Introduction

Previous studies on spontaneous speech phenomena for various languages have provided evidence for a view of disfluencies as regular linguistic devices used to achieve a better synchronization between interlocutors (*e.g.*, [1], [2], [3]). Our preliminary studies on disfluencies [4] suggested that in European Portuguese (EP) segmental prolongations (PRLs) occur more frequently than filled pauses (FPs), and are better rated by listeners. Contrarily to what has been observed for other languages, *e.g.* [3], these two classes of events are both used to signal upcoming delays and to gain time before syntactic complex units, as instances of a same device occurring in complementary distribution. The present work aims at extending our analysis to an enlarged corpus in order to verify the consistency of these trends, and at further exploring results concerning listeners ratings in order to better understand the prosodic constraints at play.

## 2. Data

The working corpus, of about 12h, encompasses different types of spontaneous and prepared oral presentations at high-school and university levels. The full corpus was manually annotated for disfluencies, following [5], and fluency ratings were provided by 3 independent annotators for a subset of about 4 hours (2h high school, 2h University). Sentence like units were coded by annotators for ease of expression, as felicitous or infelicitous.

## 3. Results

In a listening test with extracts of the high school corpus, 80% agreement was found between the annotators' ratings and the average ratings of 40 listeners, using a 5-point scale to judge felicitous and infelicitous moments in speaker performance. When only average answers  $\Rightarrow 4$  were considered felicitous, three different sets of disfluency phenomena clearly emerge, which are associated with different acceptability rates, as figure 1 shows: (1) FPs and PRLs; (2) Substitutions and deletions; (3) fragments, repetitions and complex D/F sequences. While PRLs are even better rated than FPs, repetitions (REPs), which are also commonly viewed as manifestations of planning load, are strongly penalized.

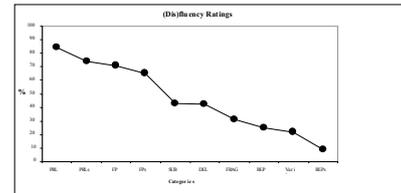


Figure 1: Percentage of felicitous moments by category.

Regarding tonal scaling, PRLs and FPs occurring in felicitous moments are regularly scaled relatively to their adjacent constituents. This is not the case for REPs and FPs occurring in infelicitous moments. In the university lectures, however, the speaker with the highest percentage of felicitous moments presents the lowest frequency rate for FPs and extensively uses single repetitions instead. Those are regularly scaled and their previous term most often elongated. As for single FPs and PRLs, the presence of a silent pause preceding the repetition appears to be crucial. Its removal or drastic reduction induces strongly negative judgments. Identical effects may be observed when silent or filled pauses occur within prosodic words or even phonological phrases, in the absence of a PRL.

## 4. Conclusions and future work

The fact that in our corpus, more than 80% of FPs and PRLs are followed by silent pauses of a reasonable length, supports the view that their presence may effectively be used by listeners as a cue to an upcoming delay. It is not surprising, then that the absence of such a pause is strongly penalized as contradictory/misleading information. This does not explain, however, why silent and/or filled pauses are infelicitous in some contexts. A possible explanation, in agreement with the fact that PRLs and FPs are in complementary distribution, is that PRLs violate less prosodic constraints whenever a break occurs within a minimal intonational phrase.

## 5. Acknowledgements

This work was funded by Prime National Project TECNOVOZ number 03/165.

## 6. References

- [1] Levelt, W., *Speaking*, Cambridge, Massachusetts, MIT Press, 1989.
- [2] Swerts, M., "Filled pauses as markers of discourse structure", *Journal of Pragmatics*, 30, pp. 485-496, 1998.
- [3] Clark, H. and Fox Tree, J. "Using uh and um in spontaneous speaking", *Cognition*, 84, pp. 73-111, 2002.
- [4] Moniz, H., Mata, A. I. and Viana, M. C., "On filled pauses and prolongations in European Portuguese", *Proc. Interspeech'07*, Antwerp, Belgium, pp. 2645-2648, 2007.
- [5] Shriberg, E., *Preliminaries to a Theory of Speech Disfluencies*, PhD, University of California, 1994.