On Filled-Pauses and Prolongations in European Portuguese

Helena Moniz [1], *Ana Isabel Mata* [2], *M. Céu Viana* [2]

[1] L2F INESC-ID /CLUL, University of Lisbon, Lisbon, Portugal
[2] CLUL/FLUL, University of Lisbon, Lisbon, Portugal

helenam@l2f.inesc-id.pt, aim@fl.ul.pt, mcv@clul.ul.pt

Abstract

This paper reports preliminary results from a study of disfluencies in European Portuguese, based on a corpus of prepared (non-scripted) and spontaneous oral presentations in high school context. We will focus on the contextual distribution and temporal patterns of filled pauses and segmental prolongations, as well as on the way those are rated by listeners.

Results suggest that filled pauses and segmental prolongations behave alike, have similar functions and may be considered in complementary distribution, obeying general syntactic and prosodic constraints.

Index Terms spontaneous speech, disfluencies, prosody.

1. Introduction

Although filled pauses, as well as other phenomena characteristic of spontaneous speech, were already considered in the early 80s, as devices used by speakers *to produce more error-free, high-quality speech ([1], p.150),* this fact is not yet widely accepted. There is strong empirical evidence, however, that speakers use (dis)fluencies (D/Fs) in similar ways across languages and that those play a fundamental role in the structuring of spontaneous speech, as they are used to achieve a better synchronization between interlocutors, by announcing upcoming topic changes, delays related to planning load or preparedness problems, as well as speaker's intentions to take/give the floor or to revise/abandon an expression he/she had already presented (see [2], [3], [4], [5], [6], [7] and references therein).

Spontaneous speech appears, thus, to have its own rules and devices and the information available in D/Fs can help listeners to compensate for different types of delay or misleading information potentially affecting comprehension and resulting from the fact that speaking takes place in real time and in a variety of communicative contexts and conditions ([8], [3], [9]). There is common agreement that D/Fs are accompanied by important modifications both at the segmental and prosodic levels and that speakers and listeners use such cues systematically and meaningfully. They appear, thus, as linguistic universal devices, that, as other similar devices, are regulated by language specific constraints and under the speaker control ([2], [3], [4], [5], [10], [11], [12].

The work reported in [13] and [14] clearly shows that the fluency of L2 speakers may be considerably improved if students attain a better comprehension of the form and contextual distribution of D/Fs (in particular filled pauses and segmental prolongations) in the target language, and are provided with adequate speaking/listening training to develop prosodic skills, mainly in what phrasing and pitch control are concerned.

Similar claims could be made in terms of first language teaching (L1). Although, in many countries, namely Portugal, the development of speaking skills is included in secondary school curricula, D/Fs (in particular filled pauses and segmental prolongations) are most often explicitly ruled out by teachers, as errors.

The goal of the present work is to carry out a more detailed descriptive study of filled pauses (FPs) and segmental prolongations (PLs) in what respects their contextual distribution across prosodic and syntactic units, as well as their durational characteristics, in order to gather a better understanding of the way they may interact in European Portuguese (EP).

The paper is organized as follows. Section 2 briefly describes the corpus and the main annotation criteria we adopted. The next section presents overall D/F's frequency ratings, as well as their distribution for two different tasks: prepared and spontaneous oral presentations. Section 3 presents our results concerning the contextual distribution of filled pauses and segmental prolongations, as well as some examples illustrating their phonological behaviour, both at the segmental and the prosodic levels. Before concluding, we briefly describe an experiment carried out aiming at validating the annotator's judgements and at gathering a better insight concerning the way different (dis)fluency types are rated by listeners.

2. Data

The corpus used in this study was extracted from the CPE-FACES corpus collected by [15]. It is constituted by ten oral presentations by one female teacher of Portuguese as L1 and four of her students (two male and two female). Five of these are prepared non-scripted oral presentations about a book they have read, according to specific programmatic guidelines. As for the five spontaneous ones, they were unexpectedly asked to briefly tell a pleasant personal experience. As in both cases questions could be asked, this excerpt, initially of two hours, ten minutes and three seconds, corresponds to 11,851 words (9,708 and 2,143 for the prepared and spontaneous presentations respectively), after overlapping voices, laughs and applauses were suppressed.

The D/F's annotation scheme closely followed [5]. Additional annotation tiers were added, containing information concerning the syntactic and prosodic context of the D/F(s), as well as those of all the silent pauses in the corpus.

3. Overall distribution of D/Fs

A total of 1569 D/Fs were observed, which results in a rate of 13.24 disfluencies per 100 words. This rate is somewhat higher than those previously observed for English, reported in [6], but similar to the ones observed for Swedish by [4].

In table 1, figures are given for the relative frequency rates of the different types of D/Fs observed in the corpus. The left column gives the total number of counts for each D/F type, independently of the fact the D/F occurs as a single isolated event or is combined with other D/F types in complex sequences; the right column accounts for single occurrences or combinations of the same D/F type. Truncations were not counted as separate events, as they always co-occur with other D/Fs. A total of 129 truncations were observed, 5.48% of which occur with repetitions and the remaining 2.74% with deleted and substituted words.

Table 1. *Overall distribution of disfluencies per type: Prolongations (PRL), Filled pauses (FP), Repetitions (REP), Substitutions (SUB), Deletions (DEL), Editing terms/expressions (ED,) and Insertions (INS).*

D/Fs	Total Obs.	Obs. same type
PRL	497 / 31.7%	288 / 45.9%
REP	485 / 30.9%	94 / 15.0%
FP	274 / 17.5 %	198 / 31.6%
SUB	177 / 11.3%	23 / 3.7%
DEL	112 / 7.1%	24 / 3.8%
ED	20 / 1.3%	0 / 0%
INS	4 / 0.2%	0 / 0%
Total	**1569 / 100%**	**627 / 100%**

A comparison between the two columns shows that prolongations present the highest frequency rates in both types of counts. They may also occur as single isolated events, far more often than filled pauses. This is contrary to overall figures that are given for other languages in general, or observed for similar corpora (e.g. for the EP corpus of university lectures by [16] or the French classroom corpus by [17]. It is possible this inversion reflects the fact that filled pauses tend to be stigmatized in the Portuguese high school context.

Although the quantity of speech materials considerably differ from one task to the other, our results basically agree with previous observations made for other languages regarding task and speaker dependent effects in the use and relative rate of D/Fs. (Dis)fluency rates are 2.8% lower on average in the spontaneous presentations. On the other hand, insertions, deletions and editing expressions become residual (2, 6 and 0 cases only, respectively). The relative frequency rates for repetitions are much higher in the spontaneous than in the prepared presentations (38.1% vs 29.4%) and the rate of prolongations and filled pauses decreases (32.7% vs 26.3% and 18.0% vs 15.3%, respectively). These differences in total counts are mainly due, to the number of times entire intonational phrases are repeated. For both tasks and for all subjects, however, FPs and PRLs differ from all the other D/Fs types, as they often occur as single, isolated events, while repetitions, substitutions, truncations, insertions and editing expressions tend to combine with each other forming complex sequences.

Even though it has been noted by [6], [17], [18] and [19] that filled pauses and prolongations may be considered as acoustically and functionally similar, this fact is not always accepted. In [4], the most extensive study we could find on prolongations, they are considered similar in form, as both are vocalizations that rely on durational cues only, but they appear to differ in their contextual distribution, in the phonotactic constraints they obey, as well as in their functionality. In their influential work, [3] claim that while filled pauses are consistently used to signal upcoming delays, prolongations reflect ongoing delays. Moreover, the latter may be viewed as a general phonological process applied to parts of words, whereas languages appear to have at least two FP contrastive forms to signal different degrees of upcoming delays (e.g. uh and um for English). The fact that those forms obey language specific phonotactic constraints and may also be lengthened constitute a strong piece of evidence for considering them as legitimate (English) words. This claim was (at least partially) verified for languages so different from English, as Mandarin and Japanese (e.g. [19], [20], [21]). The work described in the next section aimed at testing such proposals for EP.

3.1.The form of FPs and PRLs

For European Portuguese, little empirical work on spontaneous speech has been carried out so far. During the transliteration of the CPE-FACES corpus, [15] found basically three distinct forms for FPs: (i) an elongated central vowel only; (ii) a nasal murmur only; and (iii) a central vowel followed by a nasal murmur. She proposed they should be spelled as *aa*, *mm* and *aam*, respectively, as the quality of the central vowel most often coincides with the one of unstressed /a/. This schwa-like quality ([ɐː] or [əː]), was confirmed by [22] as the only effectively present in that same corpus and also by [16] for the university lectures one.

Although a schwa-like quality ([ɐː] or [əː]), appears to be the most commonly used, in a quick survey of other speech corpora available for EP, we have found, however, some speakers consistently using the neutral vowel [ɨː] instead, and others both [ɨː] and [ɐː], sometimes in the same sentence, depending on the quality of the previous word last vowel. Our point here is not to acknowledge that FP vocalizations may be built around central vowels and speakers may differ in their preferences, but that FPs do not appear to behave as other words in the language. In EP, [ɨ] and [ɐ] correspond to reduced forms of different vowels in unstressed position (/i/, /e/, /ɛ/ vs /a/, respectively) and words homophonous to *aa* (the preposition *a* or the feminine determinant *a*) do not undergo this type of contextual variation. The same appears to hold for prolongations. The lengthening of words ending in a coronal fricative, for instance, could be obtained by prolonging the entire rhyme and/or the fricative only. Most of the time, however, the neutral vowel [ɨː] is appended to achieve the desired effect. Contrarily to regular *sandhi* phenomena generally observed within as well as across word boundaries, the final fricative is never realized as [z], but as [ʒ].

Similarly to FP forms, single occurrences of PRLs between stretches of fluent speech, may be preceded and followed by silent pauses, as they most often occur on function words with a CV or V structure. Even though they are not always central, the vowels of such syllables, may be as long as the ones observed for FPs, and PRLs followed by clitic *mm* may be almost identical in form to *aam* instances. We can, thus, question ourselves if there is in fact a separate category constituted by a nasal murmur only, or if this nasal murmur simply serves as a means for further elongating both FPs and PRLs. Although, at least in some contexts, a long nasal murmur only, instead of *aam,* appears not to hurt the sensibility of native speakers, all single *mm* instances we could find in available corpora were always associated with functions different from those generally assigned to FPs, such as the expression of doubt, agreement or denial.

3.2. Contextual distribution

As first shown for EP in [22], and further confirmed in subsequent work by [16], (i) *aam* generally occurs at major intonational phrase boundaries, (ii) *aa* is the most likely form at minor intonational phrase boundaries, even though it may occur in practically all contexts, as it is the only form used by two of the speakers; (iii) *mm*, as mentioned above, is always cliticized onto prior elongated words.

Such locations generally correspond to different planning loads and numerous studies have shown that the difficulty of the task has direct reflections not only on how often and for how long speakers do pause but also on how often they need to signal a delay in speaking to the listener. Probably because PRLs affect first words of lower level constituents both at the syntactic and at the prosodic level, they are often viewed as related with difficulties in lexical search or with difficulties in the pronunciation of upcoming words. Such difficulties are observed, however, mostly in cases of stuttering and, at least in our corpus, PRLs occur at locations with identical potential planning loads as FPs.

Table 2. *Relative frequency rates for FPs and PRLs at constituent, clause and sentence initial boundaries.*

D/Fs	Const.B	ClauseB	Sent.B
aam	7.5%	15.0%	77.5%
aa	30.3%	25.7%	44.0%
mm	7.7%	92.3%	0%
PRLs	34.2%	56.5%	9.3%

Table 2 shows that FPs, in particular *aam* are clearly preferred sentence initially, but not *mm*. PRLs are more likely at internal clause boundaries, and their rate compares well with that of *aa* at the constituent level. The observations above may suggest a different syntactic and prosodic distribution but not necessarily a difference in planning effort. Moreover, the fact that a prolongation is implemented on the first word of a syntactic constituent does not necessarily entail the absence of an upcoming complex structure. A name, for instance, may have an embedded clause or prepositional phrase.

Table 3. *Frequency rates for FPs and PRLs relatively to the complexity of following syntactic structure*

D/Fs	Complex	Simple
aam	92.0%	7.5%
aa	88.8%	11.2%
mm	100.0%	0.0%
PRLs	94.8%	5.2%
Total	92.5%	7.5%

In table 3, we counted as complex all such cases. The observed distribution support the view that both FPs and PRLs occur most often when the upcoming discourse unit is syntactically complex (92.5% of the cases) and are rare at the beginning of simple sentences (7.5% only).

3.3. Durational features

Temporal characteristics have been used as a strong argument by [3] for postulating two different contrastive FP forms for English associated with upcoming delays, as well as for distinguishing them from PRLs. In order to verify if both types of D/Fs behave in fact differently, we calculated their mean values, as well as that of the silent pauses preceding and/or flowing them and tested for the significance of such differences. We found significant differences (p<0.001 either with Tahmane's and Tukey's *post-hoc* tests) between *aa* and *aam* as well as between simple prolongations and prolongations with a *mm* attachment, but not between *aa* and simple prolongations nor between those with *mm* attachments and *aam* (both *post-hoc* tests N.S.), the latter pair allowing for a much higher gain in time. Similar observations were made for the silent pauses preceding or following them.

Table 4. *Mean durations (in ms) of FPs and PRLs and of the preceding and following silent pauses (P/SP and F/SP, respectively).*

D/F	P/SP	D/Fs	F/SP
aam	800	655	616
aa	603	378	166
mm	651	585	744
PRLs	416	392	277

3.4.(Dis)fluency ratings

In order to validate the annotator judgments concerning fluency rates, a set of passages of about 12 seconds each was selected, containing different types of single as well as of complex D/F events.

Twenty high-school teachers of Portuguese as L1 and 20 speech engineers participated in the final experiment. They were told their help was needed to identify felicitous and infelicitous moments concerning ease of expression, during oral presentation by students and their teacher of Portuguese. For that purpose, they should try to guess the type of moment the passage had been extracted from, according to the following scale: (1) very bad; (2) infelicitous; (3) acceptable; (4) good or (5) excellent/very good.

The results agreed in 80% of the cases with those previously provided by the annotator and no significant difference was found between teachers and engineers. Listeners differ from the annotator in that they clearly rejected all passages with complex D/F sequences, and rated filled pauses and prolongations better, with a clear preference for the latter.

Not all instances of both filled pause and prolongations, however, were judged fluent. Filled pauses with ascending or descending F0 contours were strongly penalized. Good rates were only given for those presenting a stationary F0 contour and at intonational phrase boundaries, where they behave mostly as parentheticals, and do not disturb F0 global trends. Within intonational phrases, they were not well tolerated. In that location, prolongations were clearly preferred provided they do not break a phonological phrase. Contrary to filled pauses, the better rated prolongations occur on coordinative and completive conjunctions and show important F0 excursions, similar to those expected at the offset or onset of intonational phrases.

4. Conclusions

The observed trends concerning the distribution and duration of FPs, as well as the results of the previous experiment, may be viewed as manifestations of planning effort at different levels of the prosodic structure, at least partly confirming the observations of [3], [7] and [10].

Although further work on larger corpora is needed in order to get a better insight into the specific behavior of filled pauses and prolongations, our current results clearly suggest that these two classes of events occur in complementary distribution and are used as a device to both sustain fluency and gain time before syntactic complex units. Despite the early stage of our work in this area, the informal experiment we conducted has led us to a better understanding of how listener can score the (dis)fluent phenomena and to account for some of the prosodic characteristics that may influence the listeners' judgements.

5. Acknowledgements

This work was partially funded by PRIME National Project TECNOVOZ number 03/165, by the FCT project POSC/PLP/58697/2004 and by CLUL's FCT pluri-annual funding 2007.

6. References

[1] Heike, A., "A Content-Processing View of Hesitation Phenomena", Language and Speech 24: 147-160, 1981.

[2] Clark, H. H. and Wasow, T., "Repeating Words in Spontaneous Speech", Cognitive Psychology 37: 201-242, 1998.

[3] Clark, H. and Fox Tree, J., "Using *uh* and *um* in Spontaneous Speaking", Cognition 84: 73–111, 2002.

[4] Eklund, R., Disfluency in Swedish Human-Human and Human-Machine Travel Booking Dialogues, PhD Diss., Institute of Technology, Linköping University, 2004.

[5] Shriberg, E., Preliminaries to a Theory of Speech Disfluency, PhD Diss, Univ. of California, 1994.

[6] Shriberg, E., "To *Errrr* is Human: Ecology and Acoustics of Speech Disfluencies", JIPA, 31: 153-169, 2001.

[7] Swerts, M., "Filled Pauses as Markers of Discourse Structure", J. Pragmatics 30: 485-496, 1998.

[8] Brennan, S. and Schober, M., "How Listeners Compensate for Disfluencies in Spontaneous Speech", J. Memory and Language, 44: 274-296, 2001.

[9] O'Connell, D. and Kowal, S., "Uh and Um Revisited: Are they Interjections for Signaling Delay?"Jour. Psycholinguistic Research, 34: 555-576, 2005.

[10] Eklund, R. and Shriberg, E., "Crosslinguistic Disfluency Modeling: a Comparative Analysis of Swedish and American English Human-human and Human-machine Dialogs", Proc. ICSLP, Sydney, 2631-2634, 1998.

[11] Lickley, R, Detecting Disfluencies in Spontaneous Speech, PhD. Diss., Univ. of Edinburgh, 1994.

[12] Nakatani, C. and Hirschberg, J., "A Corpus-based Study of Repair Cues in Spontaneous Speech", JASA, 95: 1603-1616, 1994.

[13] Rose, R., The Communicative Value of Filled Pauses in Spontaneous Speech, M.A. Diss., Univ. of Birmingham, 1998.

[14] Wennerstrom, A., "The Role of Intonation in Second Language Fluency", Riggenbach, H. (Ed), Perspectives on Fluency, Univ. of Michigan Press, 102-127, 2000.

[15] Mata, A. I., Para o Estudo da Entoação em Fala Espontânea e Preparada no Português Europeu, PhD. Diss., Univ. of Lisbon, 1999.

[16] Trancoso, I., Nunes, R., Neves, L., Viana, C., Moniz, H., Mata, A. I. and Caseiro, D., "Automatic Speech Recognition of Classroom Lectures". Proc. Interspeech'2006, Pittsburgh, 2006.

[17] Candea, M., Contribution à l'Etude des Pauses Silencieuses et des Phénomènes dits « d'Hésitation » en Français Oral Spontané – Etude sur un corpus de récit en classe de Français, PhD. Diss., Université de Paris III – Sorbonne Nouvelle, 2000.

[18] Campione, E. and Véronis, J., "Pauses and Hesitations in French Spontaneous Speech", Proc. DISS'05, Aix-en-Provence, 42-46, 2006.

[19] Den, Y., "Some Strategies in Prolonging Speech Segments in Spontaneous Japanese", Proc. DISS'03, Göteborg, 87-90, 2003

[20] Watanabe, M, "The constituent complexity and types of fillers in Japanese", Proc. 15th ICPhS, Barcelona, 2473-2476, 2003.

[21] Zhao, Y., Jurafsky, D., "A Preliminary Study of Mandarin Filled Pauses", Proc. DISS' 05, Aix-en-Provence, 179-182, 2005.

[22] Moniz, H., Contributo para a Caracterização dos Mecanismos de (Dis)Fluência em Português Europeu, M.A. Diss., Univ. of Lisbon 2006.