Analysis of disfluencies in a corpus of university lectures

Helena Moniz[1], Fernando Batista[2], Ana Isabel Mata[3] & Isabel Trancoso[4]
[1] FLUL/CLUL, University of Lisbon & INESC-ID, Portugal.
[2] DCTI, ISCTE-IUL, Portugal.
[3] FLUL/CLUL, University of Lisbon.
[4] Department of Electronics Engineering, IST & INESC-ID, Portugal.

Abstract

This paper analyzes the prosodic properties of disfluencies and of their contexts in a corpus of university lectures. Results show that there is a general tendency to repair fluency by means of prosodic contrast marking strategies (pitch and energy increase), regardless of the specific disfluency type, but still there are degrees in the contrast made by certain types. As for tempo patterns, the region to repair is longer than the repair itself, and there is a strong trend manifested in lengthy silences between those regions. However, the monitoring of the lengthiness effect varies considerable amongst speakers and disfluency types.

Key words: Prosody, (dis)fluency, contrast, and university lectures.

Introduction

When targeting the prosodic analysis of the different regions of a disfluent sequence, one must be aware of two main strategies accounted for in the literature (Levelt & Cutler 1983): (i) a contrastive strategy between the *reparandum* (or region to repair) and the repair of fluency, manifested by pitch and energy increases at the onset of the repair; and (ii) a parallel prosodic strategy between the same areas, meaning, the repair mimicking the tonal patterns of the *reparandum*. Although Levelt & Cutler (1983) pointed out, in a map-task corpus, that contrastive marking strategies are associated with error correction categories (mostly substitutions) and parallelism with appropriateness categories (*e.g.*, repetitions and insertions), there is no one-to-one mapping between the strategy used when monitoring speech and the distinct disfluent category or even, in a larger scale, between the strategy and the domain itself (Savova 2003; Cole *et al.* 2005; *inter alia*).

In this paper we analyze the mapping between the disfluency and the fluency repair in order to verify (i) if there are contrast *vs.* parallelism strategies; (ii) which prosodic parameters configure these strategies; and (iii) if a variation of prosodic strategies may be related to proficiency degrees in the university lecture domain.

Proceedings of the International Conference of Experimental Linguistics ExLing 2012, 27-29 August 2012, Athens, Greece.

Data and methods

This study uses a subset of the Lectra corpus (Trancoso *et al.* 2008), with a total of 16 hours, 7 speakers, 110427 words, and 3.46% of disfluencies. Disfluencies were annotated accordingly to Shriberg (1994) and Eklund (2004). As Table 1 shows, filled pauses and complex sequences are the most frequent types, followed by repetitions; the selection of the disfluency types is speaker dependent (*e.g.*, filled pauses are the most frequent type for S5 and S7, and repetitions are quite frequent for S4 and S6).

Table 1. Distribution of disfluencies per speaker. "S" stands for speaker.

Type	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**Total**
Complex	115	100	116	124	124	247	227	**1053**
Deletions	13	20	66	74	6	91	35	**305**
Filled pauses	92	52	45	66	379	163	342	**1139**
Fragments	13	42	35	21	11	59	33	**214**
Prolongations	35	0	0	0	46	32	22	**135**
Repetitions	45	56	59	108	58	205	84	**615**
Substitutions	29	43	60	45	33	98	34	**342**
Total	**342**	**313**	**381**	**438**	**657**	**895**	**777**	**3803**

Features were calculated for the disfluent sequence itself and also for the two contiguous words, before and after the disfluent sequence. The following set of features was used for each word in those regions: f_0 and energy raw and normalized mean, median, maxima, minima, and standard deviation, as well as part-of-speech tags, number of phones, and durations. Energy and f_0 slopes were calculated based on linear regression. Pitch and energy were extracted using the *Snack Sound Toolkit* (http://www.speech.kth.se/snack/). Durations of phones, words, and interword-pauses were extracted from the recognizer output.

Prosodic analysis

Our analysis shows that pitch and energy increase from the disfluency region ("disf" or *reparandum*) to the repair of fluency ("disf+1"). Results from a Kruskall-Wallis test show significant differences (*p-value* < 0.001) in "disf-1", "disf" and "disf+1", concerning pitch ($X^2(12)=53.82$; $X^2(12)=161.54$ and $X^2(12)=34.62$; respectively) and energy slopes ($X^2(12)=56.57$; $X^2(12)=78.09$ and $X^2(12)=152.47$; respectively) within a word as well as in the differences of pitch and energy amongst those regions (pitch and energy difference between "disf-1" and "disf" $X^2(12)=139.32$ and $X^2(12)=92.61$; between "disf" and "disf+1" $X^2(12)=378.34$ and $X^2(12)=104.95$; respectively). Thus, pitch

and energy slopes are significantly different within the words immediately before and after the disfluency (but not before and after that). Pitch and energy increase from the disfluency to the repair, independently of the disfluency type and stand for the majority of our speakers. There are, however, degrees in the resets of the next unit (the highest pitch and energy resets occur after a filled pause and a repetition, respectively).

Results show that the prosodic contrast strategy (pitch and energy increase) does not apply exclusively to error correction categories (substitutions, deletions, fragments and complex sequences). Substitutions, *e.g.*, show similar significant pitch/energy increase differences on the onset of the repair, or even on the slope within the repair. Thus, results do not support the use of a contrast strategy exclusively on the error corrections, as described in Levelt & Cutler (1983). There is a more general tendency towards a contrast marking strategy, regardless of the specific disfluency type.

As for tempo analysis, the averages of the different regions are represented in Figure 1. The disfluency is the longest event, the silent pause between the disfluency and the following word is longer in average than the previous one, and the "disf+1" word is the shortest.

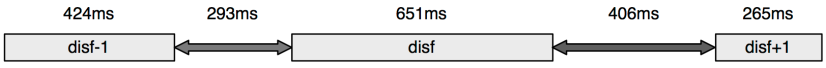

Figure 1. Durations of all the events.

Contrarily to the previous prosodic parameters, tempo patterns exhibit significant differences (*p-value* < 0.001) per speaker and disfluency type in the units "disf-1"; "silent pause before", "disf", "silent pause after", and "dif+1" ($X^2(6)=336.34$; $X^2(6)=128.16$; $X^2(6)=178.82$; $X^2(6)=401.10$; $X^2(6)=250.21$; $X^2(12)=377.93$; $X^2(12)=534.72$; $X^2(12)=1485.84$; $X^2(12)=176.86$; $X^2(12)=449.0$; respectively). For instance, when uttering a filled pause, the previous silent pause is longer (486ms) than the one after (269ms). The articulation and speech rate as well as the phonation ratio (Cucchiarini *et al.* 2002) per speaker are quite distinct as well (ranging from 12.8 to 20.3; 12.5 to 16.7; and 69.3 to 89.3 respectively).

There are degrees in mastering all the features described. Thus, the acoustic correlates of the most proficient speaker (S6) are expressed by means of: (i) the highest energy slope within the repair; (ii) a considerable pitch increase also in the repair; and (iii) the highest articulation and speech rates - correlates which are frequently associated in fluent sequences with higher level strategies of language use. Note that the combination of all those strategies is not found in the production of the remaining speakers.

Conclusions

Three main conclusions arise from the data. Firstly, different regions of a disfluent sequence are uttered with distinct prosodic properties and speakers contrast those areas with the minimum context possible. Secondly, there are different contrastive degrees in using the prosodic parameters (filled pauses are the most distinct type in what regards pitch increase and durational aspects, and repetitions in what concerns energy rising patterns). Finally, when repairing fluency, speakers overall produce both pitch and energy increases, but they monitor tempo aspects in an idiosyncratic way.

Our results, supported on a considerable amount of data, point out to domain specificities. The systematic way in which the prosodic properties are used to repair fluency points out to higher level strategies of language use. Our work, thus, contributes to a definition of fluency markers which incorporates surgical and contrastive strategies in the production of the so called disfluencies and of fluency repairs. Future work will tackle comparable studies for other domains, and also for other languages in the classroom domain.

Acknowledgements

This work was supported by national funds through FCT – Fundação para a Ciência e a Tecnologia – under Ph.D grant SFRH/BD/44671/2008 and projects CMU-PT/HuMach/0039/2008, PTDC/CLE-LIN/120017/2010, and by DCTI - ISCTE-IUL – Lisbon University Institute.

References

Cole, J., Hasegawa-Johnson, J., Shih, C., Kim, H., Lee, E., Lu, H., Mo, H. & Yoon, T. 2005. Prosodic parallelism as a cue to repetition and error correction disfluency. In Proc. of DISS'2005, 53-58, Aix-en-Provence, France.

Cucchiarini, C., Strik, H. & Boves, L. 2002. Quantative assessment of second language learner's fluency: comparisons between read and spontaneous speech. Journal of the Acoustic Society of America 111(6), 2862-2873.

Eklund, R. 2004. Disfluency in Swedish human-human and human-machine travel booking dialogues. Ph.D Dissertation, University of Linköping.

Levelt, W. & Cutler, A. 1983. Prosodic marking in speech repair. Journal of Semantics 2, 205-217.

Savova, G. & Bachenko, J. 2003. Designing for errors: similarities and differences of disfluency rates and prosodic characteristics across domains. In Proc. Of Interspeech'2003, 229-232, Geneva, Switzerland..

Trancoso, I., Martins, R., Moniz, H., Mata, A. I. & Viana, M. C. 2008. The Lectra corpus – classroom lecture transcriptions in European Portuguese. In Proc. Of LREC'2008, Marrakesh, Marocco.