MISSING DISFLUENCIES

Robin J. Lickley
Dept. of Linguistics, University of Edinburgh

ABSTRACT

Everyday experience suggests that many disfluencies pass unnoticed by listeners attending to speech. This paper presents the results of a perception experiment on a corpus of spontaneous Dutch speech, where the subjects are asked to detect disfluencies as they compare a transcript with the recording they are hearing. The results show that many disfluencies are missed by listeners even when they are trying to spot them.

INTRODUCTION

Spontaneous speech contains frequent occurrences of filled pauses (*uh, um*), repetitions (*the the door is on the left*) and false starts (*put the move the plant to the left*), all of which may be referred to as disfluencies or repairs. These phenomena occur with great frequency in normal, spontaneous speech (e.g. [1, 2]), and yet we as listeners rarely seem to notice them. Researchers who have to transcribe normal speech often report finding it hard to detect disfluencies, to transcribe them correctly and to place them correctly even when they are specifically listening for them or doing verbatim transcriptions. Computational models of speech understanding, on the other hand, will attempt to assign lexical descriptions to everything in the speech signal and then try to resolve anomalies via mainly syntactic information (e.g. [3, 4]). It seems that the human listener may have a very useful ability to avoid some of the processing problems suggested by computational models: many disfluencies may be "filtered" out before lexical or syntactic processing commences.

Previous work on the perception of speech with disfluency has shown that under certain conditions disfluency can be detected very early in the speech signal, even before the recognition of the first word after the interruption, and that prosodic information might play an important part [1, 5, 6, 7]. But under more normal listening conditions, it may be that many disfluencies are missed altogether. Some previous work has discussed the phenomenon we examine here to a limited extent. Martin and Strange (1968) [8] suggested that listeners noticed very few disfluencies and tended to displace the few that they heard towards clause boundaries. Duez (1993) [9] found that "prepausal" lengthening was a valuable cue to the *detection* of self-interruptions in French speech. But in neither case was much distinction made between types of disfluency. The eventual aim of the present study is to discover, via perceptual experiments, which types of disfluency are most likely to be missed by listeners and to look for explanations for this perceptual illusion. The work described in this paper is a first step towards this aim. We also describe a larger project for which this work can be seen as a pilot study

CORPUS

The spontaneous speech materials used for the experiments were a set of 6 instructional monologues in Dutch collected and transcribed orthographically by Blaauw [10] and based on an idea by Terken [11]. In each monologue, a different male native speaker of Dutch described the construction of a picture of a house from pieces of coloured card to a listener who was neither visible nor audible to him. The length of the six monologues varied from about 3 minutes to about 17 minutes, depending on the amount of detail each speaker considered necessary for successful completion of the task.

Disfluencies were marked in the transcription by the author and checked by two other phoneticians.

The corpus consisted of a total of 4885 words in which 762 disfluencies, including silent pauses, filled pauses, repetitions and false starts, were identified. Disfluencies thus occurred every 6.4 words overall, with rates for individual speakers ranging from every 5.3 words to every 9.6 words. The disfluency types which concern us for the present study are filled pauses ($N = 208$), repetitions ($N = 65$) and false starts ($N = 96$).

EXPERIMENT

Materials

The speech materials used for the experiment were the full recordings of the six monologues described above. These were presented over stereo headphones from DAT tapes. The original transcriptions of the monologues were edited so that all disfluencies were removed as well as all labels and then printed out with double spacing between lines and triple spacing at major instruction boundaries to make it easier for subjects to follow them. In addition to the transcriptions, several pieces of coloured card were provided, which could be used to form a picture of a house.

Subjects and Procedure

Twenty subjects were paid to take part in the experiment. All were native speakers of Dutch, students or staff of the University of Utrecht, between the ages of 18 and 30, who reported no hearing defects.

Subjects were seated in a soundproof booth and given an instruction sheet to read, describing the task. When it was clear that the subjects understood the instructions clearly, the experiment began. Subjects were asked to listen to the tape and follow the transcriptions, marking with a cross any point at which the speech and the script differed. At the same time, subjects carried out the house-building task described in the monologues: this ensured that they were actually attending to the meaning of the speech, rather than concentrating fully on spotting anomalies, and thus made the listening task more realistic.

At the end of each monologue, the tape was stopped and the house that the subject had built was examined. Each subject was tested individually. The running time for the experiment was about 50 minutes.

Results

For each filled pause, repetition and false start, the number of subjects detecting the disfluency was totaled and these totals were then averaged for each type of disfluency. In many cases the outcomes for an individual disfluency were confused by the adjacency of another or by the fact that one occurred within another (a filled pause or a repetition might be found within a false start, for example) so it was impossible to decide which to count as having been detected or missed. For this reason, from the original totals of disfluencies found in the corpus, we were left with 125 "clean" filled pauses, 16 "clean" single word repetitions, insufficient numbers of other repetitions, and 29 "clean" false starts.

Subjects were able to detect filled pauses 55.2% of the time. A clear difference was found between detection of filled pauses which were *within* sentences and those *between* sentences. Within sentences, filled pauses were correctly spotted 51.4% of the time (N = 88), while between sentences they were more easily detected (65.4%, N = 37).

A difference in the detectability of repetitions and false starts is suggested by the outcomes for those with single-word reparanda: single-word repetitions were detected only 27% of the time (N = 16), where false starts of the same length (in number of words) were detected at a rate of 39.3% (N = 7). In addition, longer false starts appeared to be easier to spot, with a 50% success rate (N = 22) for those with a reparandum of two words or more. Note that longer and more complex disfluencies were excluded from the analysis because of their complexity but that their rate of detection could be estimated at 90-100%.

DISCUSSION

A transcription-checking experiment tested the ability of listeners to detect disfluencies in spontaneous speech.

All types of disfluency included in the analysis showed lower than optimal rates of detection, ranging from 27% for single-word repetitions to 65.4% for filled pauses.

Differences were found between certain types. Filled pauses between sentences were easier to detect than those within sentences. There are a number of possible explanations for the difference, which will be investigated further in later work: filled pauses between sentences may simply be acoustically more prominent than those within either because of features of the filled pauses themselves (e.g. mid-clause filled pauses have pitch features which vary with their context [12]) or because of a more prominent pause context [9]; cognitive processing load on the listener may be lighter between sentences, allowing greater attention to the detection task, rather than to understanding the message. Single-word repetitions were harder to detect than false starts of the same length: one possible explanation for this is that "clean" repetitions are likely to be "prospective" and thus less likely than others to be accompanied by pause [13] which has been suggested as a possible cue to detection [9]; another explanation may be that since such repetitions are most likely to be short function words [1], they will be less prominent acoustically and perceptually than the words occurring in false starts of the same or greater length. Finally, the expected result that disfluencies with longer reparanda would be easier to detect was confirmed.

The number of "clean" tokens available in this corpus makes it difficult to come to firm conclusions on these results alone. However, the indications suggested here provide useful seeds for further study, which will involve a considerably larger amount of data.

WORK IN PROGRESS

The work described in this paper provides useful input to a larger project currently underway at the University of Edinburgh. Using the HCRC map task corpus [14] and performing a series of transcription experiments and acoustic and prosodic analyses we investigate the relationship between the tendency of disfluencies *not* to disrupt the processing of speech and the acoustic and prosodic features of such speech. It is hoped that this research will be of great value to our general understanding of how listeners process spontaneous speech.

ACKNOWLEDGEMENT

This work was carried out while the author was working at the Research Centre for Language and Speech (OTS), Utrecht University, The Netherlands. The author wishes to thank Noortje Blaauw, Sieb Nooteboom and Hugo Quené for their help and advice on the project. The current project at the University of Edinburgh is supported by ESRC Award No. R0002352.

REFERENCES

[1] Lickley, R.J. (1994), *Detecting disfluencies in spontaneous speech*, Unpublished PhD thesis, University of Edinburgh.
[2] Shriberg, E.E. (1994), *Preliminaries to a Theory of Speech Disfluencies*, UC Berkeley.
[3] Hindle, D. (1983), *Deterministic parsing of syntactic non-fluencies*, in Proceedings of the 21st annual meeting of the Association for Computational Linguistics, pp 123-128.
[4] Shriberg, E.E., Bear, J. and Dowding, J. (1992), *Automatic Detection and Correction of Repairs in Human-Computer Dialog*, in Proceedings of the DARPA Speech and Natural Language Workshop. Marcus, M. (ed.).
[5] Lickley, R.J., Bard, E.G. and Shillcock, R.C. (1991), *Understanding Disfluent Speech: is there an Editing Signal?*, Proceedings of the ICPhS, Aix-en-Provence, France, vol 4 pp 98-101.
[6] Lickley, R.J. Shillcock, R.C. and Bard, E.G. (1991), *Processing Disfluent Speech: How and When are Disfluencies Found?*, Proceedings of Eurospeech 91, Genova, Italy, pp 1499-1502.
[7] Lickley, R.J. and Bard, E.G. (1992), *Processing Disfluent Speech: Recognising Disfluency Before Lexical Access*, Proceedings of The ICSLP, Banff, Alberta, Canada, pp 935-938.
[8] Martin, J.G. and Strange, W. (1968), *The Perception of Hesitation in Spontaneous Speech*, in Perception and Psychophysics 3(6), pp 427-438.
[9] Duez, D. (1993), *Acoustic correlates of subjective pauses*, in The Journal of Psycholinguistic Research 22(1), pp 21-39.
[10] Blaauw, E. (1994), *The contribution of boundary markers to the perceptual difference between read and spontaneous speech*, in Speech Communication 14, pp 359-375.
[11] Terken, J.M.B. (1984), *The distribution of pitch accents in instructions as a function of discourse structure*, Language and Speech 27(3), pp 269-290.
[12] Shriberg, E.E. and Lickley, R.J. (1993), *Intonation of clause-internal filled pauses*, Phonetica 50, pp 172-179.
[13] Hieke, A. (1981), *A content-processing view of hesitation phenomena*, Language and Speech 24(2), pp 147-160.
[14] Anderson, A.H. *et al.* (1991), *The HCRC map task corpus*, Language and Speech 4, pp 351-366.