

FILLED PAUSES AS MARKERS OF DISCOURSE STRUCTURE

Marc Swerts[†], Anne Wichmann[‡] and Robbert-Jan Beun[†]

[†]Institute for Perception Research (IPO), Eindhoven, The Netherlands

[‡]University of Central Lancashire, Preston PR1, UK

ABSTRACT

This study aims to test quantitatively whether filled pauses (FPs) may highlight discourse structure. More specifically, it is first investigated whether FPs are more typical in the vicinity of major discourse boundaries. Secondly, the FPs are analyzed acoustically, to check whether those occurring at major discourse boundaries are segmentally and prosodically different from those at shallower breaks. Analyses of twelve spontaneous monologues (Dutch) show that phrases following major discourse boundaries more often contain FPs. Additionally, FPs after stronger breaks tend to occur phrase-initially, whereas the majority of the FPs after weak boundaries are in phrase-internal position. Also, acoustic observations reveal that FPs at major discourse boundaries are both segmentally and prosodically distinct. They also differ with respect to the distribution of neighbouring silent pauses.

1. INTRODUCTION

1.1. General

Spontaneous speech is sometimes believed to represent a deficient form of language usage. In such a view, speakers would produce ‘incorrect’ utterances while talking spontaneously, due to, for instance, limitations in memory and attention, or to physiological constraints. As a consequence, this type of speech distinguishes itself considerably from what an ‘ideal speaker’ in an ‘ideal communicative setting’ would produce: it typically exhibits flaws in syntactic constructions, articulation errors, hesitations, self-repairs, etc.

However, for a number of reasons, it is no longer fashionable to consider spontaneous speech to be imperfect. It has, for instance, been shown that media professionals in interviews tend to speak spontaneously instead of reading their questions in order to secure effective communication, showing that ‘spontaneous’ is not necessarily synonymous with ‘functionally inadequate’ [8]. Also, numerous studies have indicated that spontaneous speech is very regular, even to the extent that one may distinguish consistencies in the production faults. This is for instance apparent from the fact that the tongue slips into patterns [11], and that mispronun-

ciations are repaired in a rule-governed way [9].

In fact, such errors in linguistic performance are legitimate topics for psycholinguistic research, since they constitute a window on the speech generation process. They give an idea about the mental processes underlying the generation of speech which is otherwise a black box. Slips of the tongue and spoonerisms are for instance often thought to be informative about how the brain controls the speech production mechanism [5].

The above implies that most investigations of speech errors are speaker-oriented. A few perceptual studies have been performed to find out how these faults are dealt with by a listener. It appears that many disfluencies tend to go unnoticed [13]. This implies of course that they may have no communicative relevance. However, as will be shown in the next section, a specific category of ‘errors’, i.e. filled pauses, is argued to be highly relevant perceptually, and even to have some communicative import.

1.2. Specific

Much work has been devoted to the treatment of hesitations, in particular filled (or voiced) pauses (henceforth: FPs). The way in which people hesitate may to some extent be language-specific: for instance, in Japanese there are very typical hesitation ‘morphemes’ such as *eto*, that are unknown in indo-european languages. Even closely related languages may differ in this respect: speakers of RP English typically have different FPs from those speaking Scottish English [3]. This study will concentrate on Dutch, in which language FPs appear as *uh* [əh], *um* [əm] or *mm* [mm].

Past literature has shown that it can be useful to distinguish between a micro- and macro-structure of hesitations. At the micro-level, it appears that FPs may indicate word-searching problems [6]. They would reflect moments of ‘time out’ in a speaker’s planning process. Words following hesitations have a low transitional probability and thus a high information value. Consequently, FPs may be useful for listeners, since they may presignal upcoming important linguistic materials. The latter was explicitly tested in the work by Fox Tree

[4] who found that the filler *um* seems to help listeners to recognize an upcoming target word faster.

At the macro-level, there is evidence that filled pauses may have a discourse function. In conversations, speakers may use them to signal that they want to ‘hold the floor’ [15], or to indicate uncertainty about their answer to a previously asked question [14]. Filled pauses are also claimed to occur at the onset of major discourse units ([12], [1], [17]).

This last idea will be further explored in the present study, more specifically addressing the question whether filled pauses (in Dutch) may carry information about discourse structure. In the literature, the latter term generally refers to a number of phenomena such as the distribution of given/new information or the specification of speech act types, but here it will be restricted to the delimitation of larger-scale discourse units (‘paragraphs’, ‘topical units’). It is first investigated quantitatively whether FPs are more typical in the vicinity of major discourse boundaries. Secondly, the FPs are analyzed acoustically, to check whether those occurring at major discourse boundaries are segmentally and prosodically different from those at shallower breaks and whether they are distinct with respect to the distribution of neighbouring silent pauses.

2. DATA

The investigations were based on an analysis of twelve spontaneous monologues: six painting descriptions produced by two female speakers of Dutch (amounting to 46.5 minutes of speech). These materials contained 226 FPs in total. Discourse structure was established by instructing nineteen subjects, while having access to speech, to mark perceived paragraph boundaries in orthographic transcriptions of the monologues (see also [16]). Additionally, prosodic phrase boundaries were determined by an independent labeller other than the authors. Strong “paragraph boundaries” were then defined as those phrase boundaries identified by at least 75% of the labellers as a paragraph transition; other phrase boundaries were considered to represent weak boundaries.

The monologues were digitized at a sampling frequency of 16 kHz. For the prosodic analyses discussed below, fundamental frequency (F_0) was determined by the method of subharmonic summation ([7]). Duration was measured directly in the digitized waveform.

3. RESULTS

3.1. Quantitative

To get a first idea of their potential relevance for discourse structure, the distribution of FPs was determined in the following way: first, it was checked whether a phrase following either a weak or a strong discourse boundary contained any FPs; if it did, it was determined whether or not the phrase had a FP in phrase-initial position. If a FP occurred between

	DBS <.75		DBS >.75	
with FP	176		50	
	initial	non-initial	initial	non-initial
	74	102	43	7
without FP	262		14	
Total	438		64	

Table 1: Phrases with one or more filled pauses and phrases without filled pauses as a function of discourse boundary strength (DBS)

two phrases, it was analysed as being at the beginning of the second phrase, rather than at the end of the first one. Since the two speakers behave very similarly, their results are pooled (see table 1). (The same procedure is followed in the next tables.)

Table 1 reveals first that FPs are indeed more typical in the vicinity of stronger boundaries, since 50 out of 64 phrases following a strong boundary have one or more FPs, compared with only 176 out of 438 following a weak one. Moreover, the majority of the phrases with FP following a stronger boundary have a FP in initial position. The converse is true for the weaker boundaries. In other words, these first countings suggest that discourse structure indeed has an influence on the relative occurrence of FPs. However, in accordance with findings by Chafe ([1]), it appears that, although one cannot predict a boundary from the presence of a hesitation, the presence of a boundary makes a hesitation highly likely.

The next sections further elaborate on the segmental and prosodic characteristics of the filled pauses, and investigates the distribution of neighbouring silent pauses.

3.2. Segmental

A natural distinction in types of filled pauses in the monologues analysed appears to be one between those that have a nasal component (*um* and *mm*) and those that do not (*uh*). Clark ([2]) has presented evidence that in the London-Lund corpus the fillers *uh* and *um* differ with respect to their communicative function: the former are often used to signal short interruptions and the latter to signal more serious ones. Additionally, Shriberg [13] found that the two forms show systematic differences in sentence position, with *um* being more typical for initial position.

This leads to the hypothesis that ‘nasalized’ FPs may be more typical at the onset of major discourse units than other FPs. Following [13], the former “may be used relatively more often during planning of larger units, and “uh” may be relatively more likely to reflect local lexical decision-making” (p. 154).

Table 2 shows that there is indeed a preference for *uh* in non-initial position, and vice versa a preference for *um* in initial

DBS <.75				DBS >.75			
234				76			
initial		non-initial		initial		non-initial	
72		162		45		31	
uh	um	uh	um	uh	um	uh	um
25	47	143	19	4	41	27	4
<i>35</i>	<i>65</i>	<i>88</i>	<i>12</i>	<i>9</i>	<i>91</i>	<i>87</i>	<i>13</i>

Table 2: Number of “uh” and “um” as a function of discourse boundary strength (DBS) and position within phrase (percentages are given in italics)

DBS <.75		DBS >.75	
183		192	
initial	non-initial	initial	non-initial
193	178	201	180

Table 3: Mean pitch (in Hz) of filled pauses as a function of discourse boundary strength (DBS) and position within phrase

position, which is true for phrases following both weak and strong boundaries. However, whereas the relative distribution of *uh* and *um* is basically the same for non-initial position, there is a stronger tendency for *um* to occur in initial position if the phrase follows a stronger discourse break.

3.3. Prosodic

There is not much work devoted to prosodic features of filled pauses. Shriberg ([13]) concentrated on durational and melodic properties of FPs in clause-internal position only. From her study, it appears, first, that the filled pauses are typically longer than similar vowels in other words or in the homophonous determiner “a”. Also, FPs show a gradual, roughly linear F_0 fall and tend to be lower than surrounding words.

Here, the prosodic properties of all FPs were analysed from the perspective of discourse structure. First, the average F_0 of a filled pause was measured (expressed in Hz), and its duration (in centiseconds). Results are given in tables 3 and 4.

The tables clearly show that FPs after stronger breaks are higher and longer in duration than those after weaker bound-

DBS <.75		DBS >.75	
42		55	
initial	non-initial	initial	non-initial
54	36	66	38

Table 4: Mean duration (in centisecc.) of filled pauses as a function of discourse boundary strength (DBS) and position within phrase

aries; but this effect is due to the FPs in phrase-initial position. In other words, FPs at the onset of major discourse units are prosodically different from those in other positions. An ANOVA revealed a significant overall difference in terms of F_0 between the filled pauses in the four different contexts ($F_{(3,306)}=49.03$, $p<0.001$). Post-hoc Scheffé comparisons brought to light that all these versions differ significantly from each other ($p<0.05$), except for the difference between the non-initial FPs after stronger boundaries and those after weaker ones. There was also a significant overall effect of duration ($F_{(3,306)}=9.01$, $p<0.001$). Post-hoc analyses revealed that initial FPs differed significantly from non-initial ones, which was true both after stronger and weaker boundaries; paired comparisons of the two types of initial FPs showed no significant effect, which was also true for the non-initial FPs.

3.4. Pausal context

Finally, the FPs were also analysed with respect to the occurrence of surrounding silent pauses. Clark [2] already noted that nasalized FPs differ from other ones with respect to the length of the following pauses, with longer pauses after *um*. Also, speakers are more likely to use *um* than *uh* when they anticipated further pauses. This points towards a distinct communicative function of the two types of FPs.

Given the observations mentioned above that *um* was much more typical in phrase-initial position, especially when this coincided with the onset of a next discourse unit, it is further tested here whether there are any regularities in surrounding pauses from the perspective of discourse structure. Results are given in table 5.

The table shows that there is a tendency - in phrases following both strong and weak boundaries - to have a pause only *after* the FP when the latter does not occur in phrase-initial position. When it is in initial position, there is a preference for the FP to have silent pauses both *before* and *after*, but this preference is much stronger after a strong discourse boundary.

4. DISCUSSION

Summarizing the results, it appears that FPs may carry information about discourse structure. Stronger breaks in the flow of information are more likely to co-occur with FPs than weaker ones. The FPs at stronger breaks also tend to be segmentally and prosodically different from other ones, and more often have preceding and following silent pauses. FPs differing in segmental and prosodic features may reflect different planning processes, but they could also be different rhetorical devices that are explicitly controlled by the speaker to signal something to a communication partner.

Therefore, it would be useful to explore in future research whether FPs are useful for listeners, too. Potentially, listeners may profit from FPs that point towards major changes in topics. The mere occurrence of a FP alone is in itself not

DBS <.75 234								DBS >.75 76							
initial 72				non-initial 161				initial 45				non-initial 31			
silent pause				silent pause				silent pause				silent pause			
both	before	after	no	both	before	after	no	both	before	after	no	both	before	after	no
29	25	11	7	16	15	69	62	37	8	0	0	2	0	25	4
<i>40</i>	<i>35</i>	<i>15</i>	<i>10</i>	<i>10</i>	<i>9</i>	<i>43</i>	<i>39</i>	<i>82</i>	<i>18</i>	<i>0</i>	<i>0</i>	<i>6</i>	<i>0</i>	<i>81</i>	<i>13</i>

Table 5: Number of filled pauses with respect to surrounding silent pauses as a function of discourse boundary strength (DBS) and position within phrase (percentages are given in italics)

sufficient to mark major units, but if a listener also takes into account segmental and prosodic properties, the fillers may have communicative relevance.

Next, the results are potentially relevant for automatic speech recognition. At the micro-level, the usefulness of FPs appears from a study by Nakatani and Hirschberg (1994) demonstrating that an automatic speech recogniser may profit from the occurrence of filled pauses to detect self-repairs.

Similarly, Swerts and Ostendorf ([18]) found that in human-machine interactions, turns introducing a new topic tended to have more disfluencies than other turns, showing that a speech recognizer may exploit these disfluencies to detect discourse structure. Also, a ‘paragraph detector’ or a ‘speech browser’ that needs to find major discourse breaks may use the occurrence and prosodic structure of FPs, in combination with other prosodic and lexical information.

ACKNOWLEDGEMENTS

Marc Swerts is also affiliated with the University of Antwerp (UIA). Thanks are due to René Collier for commenting upon an earlier version of this paper.

REFERENCES

1. W.L. Chafe (1980), The Pear Stories: cognitive, cultural and linguistic aspects of narrative production. Norwood, N.J.: Ablex.
2. H.H. Clarck (1994), “Managing problems in speaking,” *Speech Communication* 15, 243-250.
3. A. Cruttenden (1986), *Intonation*. Cambridge, Cambridge University Press.
4. J.E. Fox Tree (1993), *Comprehension after speech disfluencies*, Unpublished doctoral dissertation, Stanford University, Stanford, CA.
5. V.A. Fromkin (Ed.) (1973), *Speech errors as linguistic evidence*. The Hague, Paris: Mouton.
6. F. Goldman-Eisler (1968), *Psycholinguistics: Experiments in Spontaneous Speech*. New York: Academic Press.
7. D.J. Hermes (1986), “Measurement of pitch by subharmonic summation,” *Journal of the Acoustical Society of America* 83, pp. 257-264.
8. S. Kowal, M.R. Bassett and D.C. O’Connell (1985), “The spontaneity of media interviews,” *Journal of psycholinguistic research* 14, pp. 1-18.
9. W.J.M. Levelt and A. Cutler (1983), “Prosodic marking in speech repair,” *Journal of semantics* 2 (2), pp. 205-217.
10. C.H. Nakatani and J. Hirschberg (1994), “A corpus-based study of repair cues in spontaneous speech,” *Journal of the Acoustical Society of America* 95 (3), pp. 1603-1616.
11. S.G. Nootboom (1969), “The tongue slips into patterns,” in: A.G. Sciarone, A.J. van Essen, and S.G. Nootboom (eds.), *Nomen Society, Leyden Studies in linguistics and phonetics*. The Hague: Mouton, pp. 114-132.
12. E.A. Shergloff (1979), “The relevance of repair for syntax-for-conversation”, in T. Givon (Ed.): *Syntax and semantics. Vol 12: Discourse and syntax*. New York, SanFrancisco, London: Academic Press, pp. 261-286.
13. E.E. Shriberg (1994), *Preliminaries to a theory of speech disfluencies*. Unpublished doctoral dissertation, University of California at Berkeley.
14. V.L. Smith and H.H. Clark (1993), “On the course of answering questions,” *Journal of Memory and Language* 32, 25-38.
15. A.- B. Stenström (1990), “Pauses in monologue and dialogue,” in J. Svartvik (Ed.): *London-Lund Corpus of Spoken English: Description and Research*. Lund: Lund University Press.
16. M. Swerts (subm.), “Prosodic features at discourse boundaries of different strength,” submitted.
17. M. Swerts and R. Geluykens (1994), “Prosody as a marker of information flow in spontaneous speech,” *Language and Speech* 37 (1), pp. 21-43.
18. M. Swerts and M. Ostendorf (subm.), “Prosodic indications of discourse structure in human-machine interactions,” submitted.