Effects of open and directed prompts on filled pauses and utterance production

Robert Eklund[1,2,3] *& Mats Wirén*[4]
[1] *Karolinska Institute / Stockholm Brain Institute, Stockholm, Sweden*
[2] *Voice Provider Sweden, Stockholm, Sweden*
[3] *Linköping University, Linköping, Sweden*
[4] *Department of Linguistics, Stockholm University, Stockholm, Sweden*

Abstract

This paper describes an experiment where open and directed prompts were alternated when collecting speech data for the deployment of a call-routing application. The experiment tested whether open and directed prompts resulted in any differences with respect to the filled pauses exhibited by the callers, which is interesting in the light of the "many-options" hypothesis of filled pause production. The experiment also investigated the effects of the prompts on utterance form and meaning of the callers.

Introduction

Spontaneous speech differs from (most) printed text in that it includes *disfluency* (to use the most common term), i.e. pauses (unfilled/silent and filled), repetitions, segment prolongations, repetitions, truncated words and so on, with a reported average frequency of around 6% of all "words" uttered (Fox Tree, 1995; Oviatt, 1995; Brennan & Schober, 2001; Bortfeld et al., 2001; Eklund, 2004). From an automatic speech recognition perspective this poses a problem in the design of automated services, since disfluency is not always easy to detect and recognize, and consequently difficult to either "recognize-and-disregard" or to interpret and exploit. In this paper we analyse filled pause incidence in a Wizard-of-Oz (WOZ) data collection, using real customer care agents and real customers with authentic problems, the latter being unaware of their calls being recorded and analysed. More specifically, the phenomenon analysed in this paper is the incidence of filled pauses in customer utterances following either *directed* or *open* system prompts, asking them to describe their reason for calling. Furthermore, we investigated the effects of the prompts on utterance form and meaning of the callers.

Filled Pause hypotheses

Except unfilled pauses (UPs), filled pauses (FPs) constitute the most common form of hesitation in spontaneous speech, and Eklund (2004) reported that approximately 25% of all disfluencies were filled pauses.

Already in the 1950s it was shown that FPs exhibit different distribution and behavior as compared to all other types of disfluency (Mahl, 1958; Christenfeld & Creager, 1996).

Over the years, FPs have been explained according to a number of different hypotheses as to their function(s) in speech, and some (not all) of these will be summarised in the following paragraphs. Note that we will use our own names for the presented hypotheses.

Floor-holding hypothesis. Maclay & Osgood, (1959) were probably the first to suggest that FPs can be used to maintain the floor in conversation, i.e. as a means to prevent interlocutors from breaking in. This view was also forwarded by Livant (1963).

Help-me-out hypothesis. That FPs can be used as a signal asking for interlocutor help was suggested by Clark & Wilkes-Gibbs (1986), or that FPs simply signal to the listener that the speaker is encountering slight timing problems in the production of speech was proposed by Clark (2002). When a speaker is looking for a word or term which is not available to them, uttering "uh" signals to the listener that some help is desired.

Self-monitoring/error detection hypothesis. Levelt (1989) suggested that FPs are a sign of internal error detection, a thread that was extended by Christenfeld & Creager (1996) who were of the opinion that anything that halted speech production could result in emitted FPs,

making FPs adhere to Baumeister's (1984) notion of "choking under pressure".

Many-options hypothesis. Lounsbury (1954) proposed that FPs "correspond to the points of highest statistical uncertainty in the sequencing of units in any given order" (ibid., p. 99), i.e. at the beginning of clauses, before the speaker has "committed" to anything, and where speech planning consequently is most difficult. This has been repeatedly confirmed by e.g. Beattie & Barnard (1979) who observed that 55.3% of all FPs produced by customers in telephone conversations (directory enquiries) occurred at the beginning of utterances. Along the same lines, Cook (1971) observed that FPs tended to occur before the first, second or third word of a clause. Shriberg (1994) and Eklund & Shriberg (1998) reported that speakers used FPs at the beginning of utterances more often than in any other position of an utterance. Eklund (2004) reported that 45.3% of FPs were utterance-initial in a large set of corpora, while Boomer (1965) observed that the most frequent position for hesitations were after the first word of phonemic clauses. Perhaps the most striking confirmation of the many-options hypothesis is found in a study by Schachter et al. (1991) who, in order to test the many-options-hypothesis studied hesitations in lectures within three disciplines with varying degrees of inherent optionality: (1) natural science, with very few options (there are very few options to describe the orbit of a planet or the outcome of a chemical reaction); (2) social science (with an intermediate degree of available options); and (3) humanities (with an high number of ways to describe, for example, what Shakespeare really meant with a certain passage). They found that lecturers within the humanities used more FPs than lecturers within social sciences, who, in turn, used more FPs than did lecturers within natural sciences. To rule out individual differences, the same set of lecturers also gave talks on a common subject, in which case they all produced an equal number of FPs.

Attention-getting signal. Lalljee & Cook (1974) reported a number of experiments aimed at testing the floor-holding hypothesis, all of which failed to provide support for the floor-holding function of FPs. Instead, they suggested that FPs might simply fill an attention-getting function, which could also explain the oft-reported high incidence of FPs in utterance-initial positions. However, they also suggested that filled pauses might serve several different functions in conversation, and that any experiment designed to test only one particular hypothesis may not produce significant results because it fails to account for other functions.

Summing up, filled pauses have been assigned several different functions, and several of the hypotheses have been supported by experimental data. One thing to stress, as we have already mentioned was pointed out by Lalljee & Cook (1974) is, of course, that the hypotheses described above are not mutually exclusive, and that FPs might serve several functions, possibly even more than just one function at the same time.

However, there is strong support for the many-options hypothesis, or as Christenfeld (1994) summarizes his study: "more options did produce more filled pauses" (ibid., p. 192).

Semantic categories

Caller utterances were analyzed both with respect to linguistic form and meaning. To represent the meaning of utterances, we used the same tripartite semantic categories (*family*, *intent*, *object*) as in the system that was later deployed (Boye & Wirén, 2007). The first of these elements, *family*, corresponds to the general product family which the call concerns (e.g. fixed telephony, mobile telephony, broadband, etc.), whereas *intent* represents the action associated with the request (e.g. order, want-info, change-info, activate, want-support, report-error, etc.), and *object* represents the specific product or entity (e.g. particular names of products, or concepts like "telephone number", "SIM card", or "password"). For the purposes of our analysis, there were 10 families, around 30 intents, and about 170 objects.

Each slot in a semantic triplet can take the value "unknown", representing the absence of information. For instance, the most accurate semantic category for the common fragmental utterance "broadband" is (*broadband*, *unknown*, *unknown*), since this request conveys nothing about either the *intent* or *object*.

The present study

The aim of this paper is two-fold: First, to study filled pause production in the speech of customers in a customer care entrance, following either a directed system prompt or an open prompt. The hypothesis is that if the many-options hypothesis is true, then FP frequency should be higher in the open-prompt settings. Second, to study whether the prompts have any effect on the semantic triplets.

Data collection

The data analysed this paper were collected during a pilot project carried out at TeliaSonera between December 2004 and February 2005 at the TeliaSonera Customer Service Call Center in Sundbyberg (Sweden).

The aim of the project was to prepare the ground for the launching of speech-based call routing in the Telia residential customer care, a service reached at the number 90200, handling 14 million calls annually.

Call routing is the task of directing callers to a service agent or a self-service that can provide the required assistance.

The data were collected using a novel variant of the Wizard-of-Oz (WOZ) technique (for historical descriptions of WOZ, see Fraser & Gilbert,1991; Dahlbäck, Jönsson & Ahrenberg, 1993; Eklund, 2004), using authentic agents as wizards and authentic customers who were not aware of the fact that the calls were being recorded. Consequently, the quality of the data collected can be assumed to be even better than that of traditional WOZ collections, where neither agents/wizards nor customers are authentic, but are acting out roles given to them by researchers, a critique often raised against WOZ (Allwood & Haglund, 1992; von Hahn, 1986). A detailed description of the present data collection is given in Wirén et al. (2006).

The general structure of the dialogue is as follows: First the (simulated) system plays an initial open prompt, containing a welcome message and an invitation to the caller to describe their reason for call.

If the utterance contains sufficient information to route the call, no more dialogue is needed. If, on the other hand, the utterance contains some but not all information necessary to route the call, the system asks a *disambiguation question* to try to obtain the information required to route the call.

Directed vs open prompts

The experiment described here examined how customers reacted linguistically when asked to express their business, comparing two disambiguation prompts: one **directed prompt**, giving some hints as to possible ways to describe themselves, and one **open prompt**, giving no hints on how to formulate their business.

The two prompts were:

(1) Directed prompt:

Jag behöver veta lite mer om ditt ärende. Gäller det till exempel beställning, prisinformation eller support?

('I need some additional information about the reason for your call. Is it for example about an order, price information or support?')

(2) Open prompt:

Kan du säga lite mer om vad du vill ha hjälp med?

('Could you please tell me some more about the reason for your call?')

The dialogs between the wizards (authentic agents) and the (authentic) customers were transcribed by an independent consulting company, STTS (**www.stts.se**), following the Nuance Guidelines. Although transcription did not focus on disfluency labelling, one type of disfluency was labeled, i.e. the filled pause, which was indicated by the item **@hes@** in the transcriptions. All instances of these hesitation labels were located and listened to (by the first author) to confirm that they were in fact cases of (Swedish) filled pauses, most often transcribed as "eh".

Comments on data collection

As explained above, the data collection and the experiment described above allow us to take a look at some of the hypotheses proposed concerning the role of the filled pause. It could be argued that the data set is fairly limited, but it has the advantage of being entirely naturalistic (to the point that "experimental" is almost a misnomer) and that it effectively pits directed prompts against open prompts in an otherwise natural setting, with no "roles" assigned, and where all speakers were completely unaware of their speech being recorded for analysis.

Results

The collected data are summarised in *Table 1*.

Utterance form

As can be seen in *Table 1*, the use of an open prompt has dramatic effects on the syntactic-categorical behavior of the customers' utterances. Following the directed prompt, 72% of the utterances are telegraphic noun-only utterances, and sentences (that contain a finite verb) constitute less than 10% of the utterances. Following the open prompt, more than 40% of the utterances are clauses (including a finite verb) and (one) noun-only utterance are down to less than 20%.

*Table 1. Summary Statistics for the directed prompt and the open prompt, and a syntactic-categorical analysis of the customers' utterances and ratios for all categories divided by number of utterances and words. **Legend:** **S** = clause containing (at least one) finite verb; **Noun** = single noun; **NP** = noun phrase; **VP** = verb phrase; **AdvP** = adverbial phrase; **AP** =*

*Table 1. Summary Statistics for the directed prompt and the open prompt, and a syntactic-categorical analysis of the customers' utterances and ratios for all categories divided by number of utterances and words. **Legend:** **S** = clause containing (at least one) finite verb; **Noun** = single noun; **NP** = noun phrase; **VP** = verb phrase; **AdvP** = adverbial phrase; **AP** =*

*adjective phrase; **Y/N** = "yes" or "no"; **Interj** = interjection; – = no response given.*

Prompt	Utts	Words	Syntax	% Utts	% Words
Directed	118	216	N = 85	72.0	39.4
			S = 11	9.3	5.1
			Y/N = 8	6.8	3.7
			NP = 6	5.1	2.8
			– = 3	2.5	1.4
			Y/N,Noun = 2	1.7	0.9
			VP = 1	0.8	0.5
			AdvP = 1	0.8	0.5
			AP = 1	0.8	0.5
Open	121	791	S = 49	40.5	6.2
			NP = 26	21.5	3.2
			Noun = 24	19.9	3.0
			VP = 11	9.1	1.4
			AP = 5	4.1	0.6
			AdvP = 2	1.6	0.2
			– = 2	1.6	0.2
			Y/N = 1	0.8	0.1
			Interj = 1	0.8	0.1

Also, utterances following the open prompt are on average three times longer than utterances following the directed prompt. All this clearly shows that the use of an open prompt has clear effects on the linguistic behavior of the callers.

Utterance meaning

Following Boye & Wirén (2007), we can regard every element in the semantic triple as one "concept". We can then obtain a measure of how information increases in the dialogue by computing the difference between the triples representing each user utterance, where "difference" means that the values of two corresponding elements are not equal. The results for semantic concepts are shown in *Table 2*.

As can be seen in *Table 2*, although there is a gain in the number of semantic concepts retrieved from the customers' utterances, the gain is marginal and not statistically significant, either using a *t* test (two-sampled, two-tailed: p=0.16 with equal variances assumed; and p=0.158 with equal variances not assumed) or Mann-Whitney *U* test (two-tailed, p=0.288).

*Table 2. Summary Statistics for semantic concept triplets following the directed and the open prompt. Ratios are given for number of concepts compared to number of utterances and words, as well as totals and ratios for the differences (**DIFFs**) between concepts in and concepts out, i.e., how many concepts you "win" by asking the disambiguation prompt.*

Prompt	Concepts In	Concepts Out	DIFFs Total	DIFFs Change	DIFFs /Utts	DIFFs /Words
Directed	136	244	108	0	0.9	0.5
Open	144	248	122	18	1.01	0.15

As was pointed out in Wirén et al. (2006), however, two other observed differences were that there were no instances following the directed prompt where an already instantiated concept (e.g. `fixedTelephony`) was changed to something else (e.g. `broadband`), while this happens 18 times following the open prompt. Furthermore, following the directed prompt, one never "gains" more than one new concept, while there are 26 instances following the open prompt where the gain is two concepts, and even two cases where the gain is three concepts (which also means that one concept is changed).

Filled pause frequency

FP frequency is shown in *Table 3*.

Table 3. Summary Statistics for utterances, words and filled pauses, and ratios for FPs/Utts and FPs/Words.

Prompt	Utterances & Words			Filled Pauses		
	Utts	Words	Words /Utts	FPs	FPs /Utts	FPs /Words
Directed	118	216	1.8	16	0.14	0.074
Open	121	791	6.5	60	0.50	0.076
Σ	239	1007	4.2	76	0.32	0.075

Needless to say, there is a striking stability across the two settings from a FPs/number-of-words point of view. While number of words per utterance increased by a factor three following the open prompt (with an ensuing difference in number of FPs per utterance), FP occurrence divided by number of words is almost exactly the same following the two prompts. However, the figures, 7.4% and 7.6%, respectively, are considerably higher than the approximately 3.5% reported in Eklund (2004, p. 235) for Swedish in a similar setting, which could

possibly indicate that real problems and genuine planning (as was the case here) leads to a higher FP rate than what is observed in a more traditional WOZ data collection with the planning of "pretend" tasks (which was used in Eklund, 2004), which in and by itself is of interest, but needs independent corroboration.

Filled pause distribution

As was previously mentioned it has been shown repeatedly that FPs tend to occupy initial positions in utterances. FP distribution in the present study is shown in *Table 4*.

Table 4. Summary Statistics for FP position, either utterance-initial or in other position.

Prompt	FP position		Σ
	Initial	Other	
Directed	14	2	16
Open	36	24	60
Σ	50	26	76
Proportion	65.8%	34,2%	100%

As is shown in *Table 4*, not only do the majority of FPs occur in utterance-initial position, they do so markedly more so than was reported in Eklund (2004, p. 239), where 1178 out of 2601 (45.3%) of FPs were utterance-initial. The difference is statistically significant given a *Z*-test for two proportions (two-tailed, $p<0.01$).

However, once again there is no statistically significant difference between the two prompt settings. This seemingly repeats the results for FP frequency that subjects with real-world problems in a real-world setting produce more FPs than subjects in a WOZ collection, even when great care is taken to make the WOZ collection as authentic as is possible.

Discussion

Although it could, quite reasonably, be argued that the data set studied in this paper is too small to allow any far-reaching conclusions to be drawn, a counterargument would be that the data are as ecologically valid as is possible, which makes the results interesting, especially when compared to more traditional WOZ data collections, e.g. Eklund (2004).

The first, obvious, result is that it that there is no difference in FP production with respect to the two prompts—the greater number of FPs following the open prompt is most likely a result of the longer utterance following the open prompt. It would seem that to the extent that the many-options hypothesis is valid, the "real-world-authenticity" has already defined and delimited the planning problems the customers might have, and that at least these two prompts in no serious way address that particular problem. From this follows that if FP production is indeed a "planning metric" – and there is much support for that hypothesis, as we have seen – then it would seem that the use of directed prompts in a call center does not help customers in their general speech planning, or at least that this assumption receives no support in the present study.

The different prompts do, however, have other effects on customer utterances, as we have seen above, in that open prompts lead to longer and less telegraphic utterances. It could be argued that the two types of prompts mainly address the *form* of the customers' utterances, rather than the *content*, since there is no significant difference with respect to the "gain" in meaning following open prompts as compared to utterances following directed prompts.

However, different utterances following the open prompt still exhibit a greater *variation* with respect to the increase in meaning: As mentioned in earlier, there were no instances following the directed prompt where an already instantiated concept was changed to something else, while this happens 18 times following the open prompt. Furthermore, following the directed prompt, one never "gains" more than one new concept, while there are several instances following the open prompt where the gain is two or even three concepts.

Finally, the fact that more FPs were produced in this corpus than in the similar WOZ collection reported in Eklund (2004), both as such and in utterance-initial position, could indicate that authentic data possibly can reveal processes that remain hidden in WOZ collections, even if these are well-designed, and also be taken as support for the many-options hypothesis, even if no observable difference was found between the two prompt settings.

Conclusions

In conclusion, we found that FP occurrence is almost exacly the same following the two prompts. However, FP incidence in general was considerably higher than that reported in Eklund (2004), and the majority of FPs occurred in utterance-intial position – markedly more so than was reported in Eklund (2004, p. 239).

Taken together, this might indicate that an authentic setting differs from a WOZ collection, however well-designed.

Concerning utterance form, the two prompt settings gave dramatic differences: Caller utterances following the open prompt were much longer, and also much more conversational in the sense that the utterances more often constituted full clauses, including a finite verb. On the other hand, it was somewhat surprising that there was no significant difference of the gains in meaning with respect to utterances following the two prompts, although the variation was much larger after the open prompt.

Acknowledgements

The data analysed in this paper are covered by a right of use agreement ("nyttjanderättsavtal") between TeliaSonera and Karolinska Institute (637/10), signed on 8 October 2009. Thanks to Jan Lindgren for help with terminology.

References

Allwood, Jens & Björn Haglund. 1992. *Communicative Activity Analysis of a Wizard of Oz Experiment*. Internal Report, PLUS ESPRIT project P5254.

Baumeister, Roy F. 1984. Choking Under Pressure: Self-Consciousness and Paradoxical Effects of Incentives on Skillful Performance. *Journal of Personality and Social Psychology*, vol. 46, no. 3, pp. 610–620.

Bortfeld, Heather, Silvia D. Leon, Jonathan E. Bloom, Michael F. Schober & Susan E. Brennan. 2001. Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, vol. 44, no. 2, pp. 123–147.

Boomer, Donald S. 1965. Hesitation and grammatical encoding. *Language and Speech*, vol. 8, no. 3, pp. 148–158.

Boye, Johan & Mats Wirén. 2007. Multi-slot semantics for natural-language call routing systems. *Proc. Bridging the Gap: Academic and Industrial Research in Dialog Technologies.* NAACL Workshop, 26 April 2007, Rochester, New York, USA.

Brennan, Susan E. & Michael F. Schober. 2001. How Listeners Compensate for Disfluencies in Spontaneous Speech. *Journal of Memory and Language*, vol. 44, pp. 274–296.

Christenfeld, Nicholas. 1994. Options and Ums. *Journal of Language and Social Psychology*, vol. 13, no. 2, pp. 192–199.

Christenfeld, Nicholas & Beth Creager. 1996. Anxiety, Alcohol, Aphasia, and *Ums*. *Journal of Personality and Social Psychology*, vol. 70, no. 3, pp. 451–460.

Clark, Herbert H. 2002. Speaking in time. *Speech Communication*, vol. 36, pp. 5–13.

Clark, Herbert H. & Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, vol. 22, pp. 1–39.

Cook, Mark. 1971. The incidence of filled pauses in relation to part of speech. *Language and Speech*, vol. 14, part 2, pp. 135–139.

Dahlbäck, Nils, Arne Jönsson & Lars Ahrenberg, Wizard of Oz Studies — Why and How. 1993. *Knowledge-Based Systems*, vol. 6, no. 4, pp. 258–266. Also in: Mark Maybury & Wolfgang Wahlster (eds.). 1998. *Readings in Intelligent User Interfaces*, San Francisco, CA: Morgan Kaufmann.

Eklund, Robert. 2004. *Disfluency in Swedish human–human and human–machine travel booking dialogues*. PhD thesis, Dept. of Computer and Information Science, Linköping University.

Eklund, Robert & Elizabeth Shriberg. 1998. Cross-linguistic Disfluency Modelling: A Comparative Analysis of Swedish and American English Human–Human and Human–Machine Dialogues. *Proceedings of ICSLP 98*, Sydney, 30 Nov–5 Dec 1998, vol. 6, pp. 2631–2634.

Fox Tree, Jean E. 1995. The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *Journal of Memory and Language*, vol. 34, pp. 709–728.

Fraser, Norman M. & G. Nigel Gilbert. Simulating speech systems. 1991. *Computer Speech and Language*, vol. 5, pp. 81–99.

von Hahn, Walther. 1986. Pragmatic considerations in man–machine discourse. *Proc. COLING*, 25–29 August 1986, Bonn, Germany, pp. 520–526.

Lalljee, Mansur & Mark Cook. 1974. Filled Pauses and Floor-Holding: The Final Test? *Semiotica*, vol. 12, no. 3, pp. 219–225.

Levelt, Willem J. M. 1989. *Speaking. From Intention to Articulation*. Cambridge, MA: MIT Press.

Livant, William Paul. 1963. Antagonistic functions of verbal pauses: filled and unfilled pauses in the solution of additions. *Language and Speech*, vol. 6, part 1, pp. 1–4.

Maclay, Howard & Charles E. Osgood. 1959. Hesitation Phenomena in Spontaneous English Speech. *Word*, vol. 5, pp. 19–44.

Mahl, George F. 1958. On the use of "ah" in spontaneous speech: Quantitative, developmental, characterological, situational, and linguistic aspects. *American Psychologist*, vol. 13, p. 349.

Oviatt, Sharon. 1995. Predicting spoken disfluencies during human–computer interaction. *Computer Speech and Language*, vol. 9, pp. 19–35.

Schachter, Stanley, Nicholas Christenfeld, Bernard Ravina & Frances Bilous. 1991. Speech Disfluency and the Structure of Knowledge. *Journal of Personality and Social Psychology*, vol. 60, no. 3, pp. 362–367.

Shriberg, Elizabeth Ellen. 1994. *Preliminaries to a Theory of Speech Disfluencies*. PhD thesis, University of California, Berkeley.

Wirén, Mats, Robert Eklund, Fredrik Engberg & Johan Westermark. Experiences of an In-Service Wizard-of-Oz Data Collection for the Deployment of a Call-Routing Application. *Proc. Bridging the Gap: Academic and Industrial Research in Dialog Technologies.* NAACL Workshop, 26 April 2007, Rochester, New York, USA.