Proceedings of Disfluency in Spontaneous Speech (DiSS) 2015
University of Edinburgh, Scotland, UK. 8-9 August 2015

UM AND UH AS DIFFERENTIAL DELAY MARKERS: THE ROLE OF CONTEXTUAL FACTORS

Ralph ROSE

Waseda University
rose@waseda.jp

ABSTRACT

The English filled pauses *uh* and *um* have been argued to correspond respectively to shorter and longer anticipated delays in speech production. This study looks at some contextual factors that might cause this difference by investigating filled pause instances in monologue and conversation speech corpora. Results are consistent with previously observed delay differences and further show that discourse-level processing may influence differential delay marking though monologue results are more conclusive than conversation results. However, no evidence was found that lexical factors (word type, frequency) correlate with filled pause choice. The findings suggest a limited view of how speakers use filled pauses as delay markers: Not all contextual factors may trigger differential delay marking.

Keywords: filled pause, delay, contextual factors

1. INTRODUCTION

According to the perceptual loop theory [1,2], speakers are constantly monitoring their speech production and, upon identifying a problem in their production, may initiate repair of the problem. In many cases, the repair may be covert and entirely unnoticeable to the listener. But in other cases, the repair takes some overt form. In this way, filled pauses (FPs)—like other hesitation phenomena (e.g., repeats, self-repairs, lengthenings, silent pauses)—can be said to mark such a repair sequence and thus constitute a delay in the communication of the speaker's message [cf., 3,4].

In English, the FP inventory is extremely limited, consisting almost exclusively of just an open syllable *uh* or a closed syllable *um*. Furthermore, the English FPs *uh* and *um* show some difference with respect to the length of the associated delay [3-7; but see 8-9 for counter-evidence]. *Uh* corresponds to a lower likelihood of an immediately following silent pause and an overall shorter delay (FP duration plus following silent pause duration) than *um*. Thus, it has been proposed [3-5] that when speakers detect a minor problem in their speech production and thus predict a minor delay to repair the problem, they are more likely to mark this delay with *uh* rather than *um*. Conversely, when speakers detect a major problem in their speech production, they are more likely to mark the associated delay with *um*.

Given the empirical observations of a differential delay between *uh* and *um*, then the next problem is to pinpoint the cause or causes of the respective minor and major delays that trigger the FPs. Several contextual factors have been observed to correspond with the occurrence of FPs. For example, much evidence exists to show that filled pauses are more likely to be used at major rather than minor discourse boundaries [7,10]. For instance, consider the hypothetical spoken discourse shown in (1). The beginning of the discourse {A} is a major discourse boundary requiring much planning effort as the storyteller plans the entire discourse to follow. The sentence boundary {B} is a less major boundary in the discourse as a whole, while the beginning of the subordinate clause {C} is a minor boundary and the clause-internal point {D} is a more minor boundary.

(1) {A}Yesterday I was walking down the street when I saw a surprising thing. There was this guy selling toys {E} in a small {F} stall and everyone was watching him because he was so unique. {B} He would balance several toys at once in one hand {C} while demonstrating a new toy {D} with the other hand. All the kids couldn't help but watch and so many parents had no choice but to buy something!

Major discourse boundaries like {A} and {B} incur greater speech production difficulty than minor boundaries like {C} or {D} and therefore increase the likelihood of a delay.

FPs have also been observed to occur more often before content words than before function words [11] and before words with low rather than high contextual frequency [12]. Hence, in (1), there is a greater probability of a FP at point {F} before the low-frequency word *stall*, than there is at point {E} before the high-frequency word *in*. More effort is required to retrieve low-frequency words from memory and thus the likelihood of a delay increases [13]. Since content words as a whole are lower frequency than function words, the same explanation applies, meaning a higher probability of FPs before content than function words.

Because such contextual factors as discourse boundary level and word frequency have been observed to incur speech delay as measured by the occurrence of FPs, a further hypothesis is that gradient differences in these contextual factors correspond to a speaker's anticipation of greater or lesser delay and thus greater or lesser choice of *um* or *uh*, respectively. This paper reports on a test of this hypothesis using speech corpora.

2. EXPERIMENT

2.1. Methods

In order to test the hypothesis in a broad range of speech contexts, this study used two English speech corpora: one of monologues and the other of conversations. The monologue corpus is the Corpus of Presentations in English (COPE: [14]) in which native English participants spoke in response to a prompt for about ten minutes, following minimal preparation time. Participants spoke in front of a small audience of peers. The conversation corpus is the Santa Barbara Corpus (SBC: [15]) which consists of unstructured, non-task-oriented conversations. Recordings were taken as speakers engaged in normal, everyday activities and conversations in non-laboratory settings (cf., BNC Spoken Corpus [16]).

A sample of FPs from each corpus were taken and the following delay measurements were made: FP duration, presence of immediately following silent pause, duration of following silent pause, and total delay (calculated as FP duration plus duration of any immediately following silent pause). Subsequently, the following contextual measurements were also made: syntactic location (clause boundary or clause-internal: a simplified view of discourse boundary level, corresponding to points A, B, and C versus D in (1) above), following word type (content or function), and following word frequency (using frequency counts from Brown Corpus [17]).

2.2. Results

The analysis is based on 163 FPs contained in 20 minutes of the COPE (monologue) corpus and 149 FPs contained in 165 minutes of the SBC (conversation) corpus. Turn-final FPs in the conversation corpus—where a following silent pause could arguably be attributed not to the speaker but rather to an interlocutor's latent uptake—were not included in this analysis. The results for the delay measures are shown in Tables 1-4 (with test statistics for main effects).

Table 1: Mean duration of FPs.

Speech type	*uh*	*um*	
monologue	328 ms	423 ms	F(1,308)=5.8 p<0.05
conversation	340 ms	363 ms	
	F(1,308)=17.0, p<0.001		

Table 2: Proportion of FPs followed by silent pauses

Speech type	*uh*	*um*	
monologue	16.4%	38.2%	n.s.
conversation	16.9%	36.9%	
	p<0.001 (log. regr.)		

Table 3: Mean duration of silent pauses following FPs

Speech type	*uh*	*um*	
monologue	398 ms	825 ms	n.s.
conversation	744 ms	910 ms	
	F(1,133)=3.9, p<0.05		

Table 4: Mean total delay (duration of FP plus following silent pause).

Speech type	*uh*	*um*	
monologue	465 ms	876 ms	n.s.
conversation	592 ms	818 ms	
	F(1,308)=16.0, p<0.001		

The duration of FPs (Table 1) differs with respect to the type of FP (*uh*, *um*) and the speech type (monologue, conversation), and even the interaction is significant [F(1,308) = 6.1, p<0.05]. While *um* is longer on the whole, this difference is only significant in monologue speech [t(143)=4.6, p<0.001]. Although *um* consists of two phonemes compared to the one in *uh*, this does not necessarily mean that *um* consistently takes longer to articulate. The articulation of FPs seems to be partially modulated in the presence of interlocutors.

Results further show that there is more likely to be a silent pause (Table 2) after *um* than *uh*, but there is no difference between speech types in this regard. However, when looking at the length of the following silent pauses (Table 3), results show that despite the apparently very short pauses after *uh* in monologues, there is only a marginal difference between *um* and *uh* barely reaching significance at α =0.05 and no difference between speech types. This result therefore waters down differences between the FPs themselves (Table 1) when looking at the effect of the total delay (Table 4): There is a difference between *um* and *uh*, but no difference between the speech types and no interaction.

The results for the total delay show, therefore, that *um* marks a longer delay than *uh*. This replicates previous findings [3-7], and shows the results to be consistent across both monologue and conversational speech. However, the composition of that delay (i.e., the relative length of *uh/um* and its accompanying pause) differs between speech types.

As for the hypothesis that gradient differences in contextual factors corresponds to differential delays, the results for the contextual measures are shown in Tables 5-8.

Table 5: Proportion of clause boundary FPs that are closed FPs (*um*)

Speech type	boundary	internal	
monologue	74.7%	45.6%	n.s.
conversation	54.4%	58.6%	
	$p<0.05$ (log. regr.)		

Table 6: Mean total delay (duration of FP plus following silent pause) at clause locations

Speech type	*uh*	*um*	
monologue			Interaction: F(1,304)=7.3 $p<0.01$
boundary	433ms	1004ms	
internal	486ms	581ms	
conversation			
boundary	552ms	680ms	
internal	641ms	963ms	
	F(1,304)=16.4, $p<0.001$		

Table 7: Proportion of following words that are content words

Speech type	*uh*	*um*	
monologue	27.9%	27.4%	$p=0.08$ (log. regr.)
conversation	39.3%	35.5%	
	n.s.		

Table 8: Mean (log) frequency of following word

Speech type	*uh*	*um*	
monologue	7.52	7.81	F(1,296)=9.4 $p<0.005$
conversation	6.54	6.94	
	n.s.		

Overall, when a FP is used clause-internally, the probability of it being an *um* rather than *uh* seems to be near chance (Table 5). However, when a FP is used at a clause boundary, it is more likely that it will be an *um* than an *uh*. This result appears to be driven mostly by the monologue speech results, but in fact the difference between speech types is not significant. Hence, the overall result suggests that major discourse boundaries prompt greater use of *um*. However, when looking at the actual total delay results (Table 6), the situation is somewhat more complicated. While there is a main effect of FP with *um* longer than *uh* (i.e., same as shown in Table 4 above), there is an interaction between speech type and clause location. The monologue speech data shows that boundary *ums* are longer than others. Together with the results shown in Table 5, this follows the prediction that major discourse boundaries prompt longer delays which are marked by *um*. The conversation speech data seems to show something of an opposite trend where internal FPs have a longer duration than boundary FPs. However, this difference is not significant. Thus, only the monologue speech data is conclusive with regard to the contextual effect of discourse boundary on differential delay.

In contrast, as for the following word, choice of *um* and *uh* shows no connection with the type (Table 7) nor with the frequency (Table 8) of the following word. But the results do show differences between speech types with FPs in conversation (more so than in monologue) followed by more content words and by lower-frequency words. Results relating word type and frequency to mean total delay showed no relevant significant effects and are not shown here.

3. DISCUSSION

The aim of this study was to evaluate the hypothesis that differences in delays associated with FPs may be attributed to certain contextual factors. Results first show that *uh* and *um* do correspond with shorter and longer total delay, respectively (consistent with previous work). Results further show that *um* is more likely to be used at a clause boundary than clause-internally suggesting that processing associated with a major discourse boundary may be a factor that impels speakers to choose to use an *um* over an *uh*. Positive support for this conclusion comes only from the monologue speech data. The conversation speech data was inconclusive on this point. More discussion of this difference between the monologue and conversation data will be given below.

Despite the influence of discourse boundary level, comparable findings could not be obtained for the type or frequency of words following FPs, suggesting that lexical access effects are not sufficient to influence differential delay marking by FP choice.

These results suggest a limited view of FPs as delay-marking devices: While there may be many sorts of linguistic processing problems that speakers experience and which may cause an anticipation of delay such as the various contextual factors examined here, using a FP as a signal to mark the degree of that expectation (i.e., short vs. long) is not a generic technique. Rather, it seems to be limited to cases where speakers recognize difficulty in processing major discourse constituents. Other delays might lead to different techniques.

Finally, the results here bear vaguely on a larger question regarding the use of FPs in spontaneous speech: the speaker's intent. As noted in the background, when speakers detect a minor delay in speech production, they mark it with *uh*, and mark a detected major delay with *um*. This can be unpacked into two hypotheses. First, there is the hypothesis that different FPs in English correspond to different delay lengths. This may be called the *differential delay hypothesis*. Previous studies as well as the present study confirm this hypothesis.

A second hypothesis, though, is that speakers intend to convey their anticipation of a delay differentially to their interlocutors. This can be called the *differential conveyance hypothesis*. None of the studies cited in the background provide clear evidence on this hypothesis: That is, there seem to be no tests of the speaker's intent to convey something different between *uh* and *um*.

Evaluating intent is surely a difficult task, but perhaps differences between the corpora used in this study are suggestive. In the monologue corpus, speakers were obliged to speak for a target amount of time, in order to accomplish the investigator's task. Furthermore, they were doing so in front of an audience. Here, speakers may feel more compulsion to communicate about their anticipated delays: Hence, their intent may be taken as a given.

On the other hand, in the conversation corpus, speakers were under no investigative time or task constraints: They could, so to speak, take their time freely within their conversation. In this context, they were under less compulsion to communicate about their anticipated delays to their interlocutors.

If this distinction is valid, then the prediction of the differential conveyance hypothesis would be that FPs would be used differently between the monologue and conversation corpora. The results here do in fact show this (cf., Table 6) where the differential use of *um* and *uh* at different discourse boundary levels occurs in the monologue but not the conversation speech data. Therefore the results provide support for the hypothesis. Of course, this is highly speculative and warrants much further examination.

4. FURTHER WORK

Although this work has looked at a broad sample of data with both monologue and conversational speech, the number of samples used was relatively limited and could be expanded to confirm the findings. Also, other factors that might lead to expectation of delay could be examined such as articulation, (co)reference processing, or syntactic and semantic effects. These could be investigated in corpus studies as performed here or in controlled production or perception experiments to see whether and how these various factors are related to differential delay as marked by *uh* and *um*.

5. REFERENCES

[1] Levelt, W.J.M. 1983. Monitoring and self-repair in speech. *Cognition* 14, 41-104.

[2] Levelt, W.J.M. 1989. *Speaking: From intention to articulation. Cambridge*. MA: MIT Press.

[3] Clark, H., Fox Tree, J.E. 2002. Using uh and um in spontaneous speaking. *Cognition* 84, 73-111.

[4] Clark, H., Fox Tree, J.E. 2014. On thee-yuh fillers uh and um. Language Log, November 11, http://languagelog.ldc.upenn.edu/nll/?p=15718

[5] Smith, V., Clark, H. 1993. On the course of answering questions, *Journal of Memory and Language* 32, 25-38.

[6] Kendall, T. 2013. *Speech Rate, Pause and Sociolinguistic Variation: Studies in Corpus Sociophonetics*. London: Palgrave Macmillan.

[7] Rose, R.L. 1998. *The Communicative Value of Filled Pauses in Spontaneous Speech*, Unpublished Master's Thesis, University of Birmingham.

[8] O'Connell, D., Kowal, S. 2005. Uh and Um Revisited: Are They Interjections for Signaling Delay? *Journal of Psycholinguistic Research* 34. 555-576.

[9] Corley, M., Stewart, O.W. 2008. Hesitation Disfluencies in Spontaneous Speech: The Meaning of um. *Language and Linguistics Compass* 2, 589-602.

[10] Swerts, M. 1998. Filled pauses as markers of discourse structure, *Journal of Pragmatics* 30, 485-496.

[11] Maclay, H., Osgood, C. 1959. Hesitation Phenomena in Spontaneous English Speech. *Word* 15, 19-44.

[12] Beattie, G.W., Butterworth, B.L. 1979. Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech* 22, 201-211.

[13] Oldfield R.C., Wingfield A. 1965. Response latencies in naming objects. *The Quarterly Journal of Experimental Psychology* 17, 273–281.

[14] Watanabe, M. Corpus of Oral Presentations in English (COPE). Unpublished corpus data.

[15] Du Bois, J.W., Chafe, W.L., Meyer, C., Thompson, S.A. 2000. *Santa Barbara corpus of spoken American English*, Part 1. Philadelphia: Linguistic Data Consortium.

[16] Crowdy, S. 1995 The BNC spoken corpus. In Leech, G., Myers, G. Thomas, J., eds. *Spoken English on computer: transcription, mark-up and application* Harlow: Longman, 224-235.

[17] Kucera, H., Francis, W.N. 1967. *Computational Analysis of Present-day American English*. Providence: Brown University press.