Memory & Cognition
2001, 29 (2), 320-326

Listeners' uses of *um* and *uh* in speech comprehension

JEAN E. FOX TREE
University of California, Santa Cruz, California

Despite their frequency in conversational talk, little is known about how *um*s and *uh*s affect listeners' on-line processing of spontaneous speech. Two studies of *um*s and *uh*s in English and Dutch reveal that hearing an *uh* has a beneficial effect on listeners' ability to recognize words in upcoming speech, but that hearing an *um* has neither a beneficial nor a detrimental effect. The results suggest that *um* and *uh* are different from one another and support the hypothesis that *uh* is a signal of short upcoming delay and *um* is a signal of a long upcoming delay.

Hardly a conversation goes by without an *um* or an *uh*; but despite their frequency, they are often considered undesirable and unnecessary. People take courses to learn how to avoid saying them, reporters strike them from verbatim accounts of what someone has said, and professional broadcasters digitally splice them out of interviews. Perhaps because of their disagreeable status, *um*s and *uh*s have been overlooked by or excluded from models of speech comprehension. The possible effects that *um*s and *uh*s might have on on-line speech comprehension were tested across two languages—English and Dutch.

Ums and Uhs as Signals of Upcoming Delay

One proposal for the function of *um*s and *uh*s is that they serve as signals of upcoming delay (Clark, 1994; Clark & Wasow, 1998; Smith & Clark, 1993). When people are thinking of answers to factual questions, they say *um* when the delay before answering is going to be long and *uh* when the delay is going to be short (Clark, 1994; Clark & Wasow, 1998; Smith & Clark, 1993). This proposed function might be what underlies a number of disparate proposals about the functions of *um*s and *uh*s, such as that they mark syntactic structure or discourse structure (Maclay & Osgood, 1959; Martin, 1967; Swerts, 1998), that they indicate a desire to maintain control of the floor (Maclay & Osgood, 1959; Rochester, 1973; Schegloff, 1981; Siegman, 1979), or that they indicate a speaker's speech production trouble, including the need for more time to plan upcoming speech (Christenfeld, Schacter, & Bilous, 1991; Jefferson, 1974; Kasl & Mahl, 1987; Levelt, 1989; Martin, 1967; Reynolds & Paivio, 1968; Schacter, Christenfeld, Ravina, & Bilous, 1991; Siegman, 1979; Tannenbaum, Williams, & Hillier, 1965). In all these cases, *um*s and *uh*s might serve as forewarnings of delays associated with these processes, instead of forewarning the actual processes.

As indicators of varying lengths of delay, *um* and *uh* may have different effects on on-line speech comprehension. For example, as an indicator of a brief delay, *uh* could focus the listener's attention on immediately upcoming information; but an indicator of a longer delay might not be as useful to listeners. When expecting a minor delay after hearing *uh*, it is to a listener's advantage to focus on upcoming speech in anticipation of a continuation. But when expecting a major delay after hearing *um*, it might not be useful to focus on upcoming speech. One reason for this is that it might be a long time before the speaker begins to talk again, and it might not be possible for listeners to maintain heightened attention for that long. A second reason is that a major delay after *um* might arise from problems that are not benefitted by a listener's heightened attention. For example, listeners may gain no benefit from heightening their attention if speakers are having trouble conceptualizing what they want to say. Yet another twist is that anticipated major delays might even shift the focus from listeners' anticipation of upcoming speech to listeners' working to help produce that speech. Listeners may play active roles in helping speakers complete their ideas (from Jefferson, 1974, p. 186); for example,

(1) Ken: I like driving. I really do. I enjoy it very much.

Louise: I used to like it until I became the complete sl-uhm, (1.0)

Ken: 'Slave'? Yeah.

I thank Herbert Clark and Nanda Poulisse for making their tapes and transcripts available and the Max Planck Institute for Psycholinguistics for technical and administrative assistance with the Dutch experiments. I also thank Susan Brennan, Maureen Callanan, and Herbert Clark for comments on earlier drafts of this manuscript. Some of the results were presented at the 38th Annual Meeting of the Psychonomic Society, Philadelphia, in 1997. The research was supported by an NSF-NATO Postdoctoral Fellowship and by faculty research funds granted by the University of California, Santa Cruz. Transcripts of the materials can be obtained from the author. Address correspondence to J. E. Fox Tree, Psychology Department, Social Sciences II, University of California, Santa Cruz, CA 95064 (e-mail: foxtree@cats.ucsc.edu).

They may help speakers advance the topic (adapted from Svartvik & Quirk, 1980; asterisks indicate overlapped speech); for example,

(2) C: I remember it over aunt Matty. One was always having to find out how many steps there were in places, before one knew whether one could take her there. This house for instance, there are thirteen up to this room

D: yes, but I really meant not so much that, which is bad enough as you know, but places where there are one or two or three steps

C: yes, right, I'm sure

D: um up and down to *places*

C: *m* (pause)

D: um

C: ramps

D: yes, the people are doing a lot with ramps.

So, even though they both signal delay, *um* and *uh* might have different effects on on-line comprehension because of the different lengths of delays they signal.

How Ums and Uhs Might Affect On-Line Speech Comprehension

The present studies tested one level of comprehension, on-line sentence processing—in particular, listeners' incorporation of spoken words into a representation of what was said. Hearing *um* or *uh* could affect on-line comprehension in three ways: There could be no effect, a detrimental effect, or a beneficial effect.

There are two different reasons that *um*s and *uh*s can be predicted to have no effect on on-line comprehension. One is that they might be filtered out of the speech stream before word integration begins. *Um*s and *uh*s have predictable F0 frequencies within clauses (Shriberg & Lickley, 1992, 1993), which might set them apart from the rest of a person's speech, allowing a processor to detect and remove them before they are even sent to the lexicon for identification, even before word incorporation can take place. If *um*s and *uh*s are automatically filtered out, people should have a difficult time noticing them in speech, and they do (Christenfeld, 1995; Lindsay & O'Connell, 1995; Martin, 1971; Martin & Strange, 1968). However, people can notice them when they want to (Christenfeld, 1995), showing that an automatic filter must at times be nonautomatic and under conscious control. The automatic filter proposal is especially attractive for theories of *um* and *uh* function that describe them as being by-products of the speech production process without their having any implications for the listener (Martin, 1971; Martin & Strange, 1968).

Another reason *um*s and *uh*s can be predicted to have no effect on on-line comprehension is that they might function at a different level of comprehension. Instead of being filtered out, *um*s and *uh*s might be processed and noted without their influencing on-line word integration. This proposal is especially attractive for theories of *um* and *uh* function that describe them as being involuntary emotional reactions to nervousness or stress (Lalljee & Cook, 1969, 1973; see also Christenfeld & Creager, 1996, and Rochester, 1973, for reviews of relevant research). Listeners might notice them and recognize the speaker's state, which in turn might influence listeners' interpretations of speaker intention and the shared conversational goals, yet they might not have any effect on on-line word recognition or integration.

In contrast to the no-effect predictions, there are at least two reasons to predict that *um*s and *uh*s could be disruptive to on-line comprehension. One reason concerns the filter hypothesis. It is possible that *um*s and *uh*s are filtered out, but that the filtering process is disruptive. The identification of *um*s and *uh*s in the speech stream, the flagging of them as nonpropositional information, and the inhibition of their incorporation into discourse can take time. This would mean that every time an *um* or *uh* is encountered, the on-line speech processor is slowed.

The second reason *um*s and *uh*s might be disruptive has nothing to do with filtering, but rather with failed attempts at incorporation. Several models of speech comprehension propose that, in understanding speech, people identify words, assign them grammatical roles, and then connect them to a syntactic representation of the utterance that is built word by word, as speech is being heard (for reviews, see Carroll, 1994; Mitchell, 1994). *Um*s and *uh*s would pose a serious stumbling block to these systems because they cannot be combined with surrounding words to form syntactic constructions (Clark & Fox Tree, 2000). Speech comprehension systems might try to incorporate them, and when can't, they resort to repair procedures to recover from the misparses. Both the disruptive filter proposal and the stumbling block proposal are especially attractive for theories of *um* and *uh* function that describe them as unwanted speech disruptions (see Postma, Kolk, & Povel, 1990, for discussion).

Finally, there is reason to believe that *um*s and *uh*s have beneficial effects on comprehension. As signals of upcoming delay, *um*s and *uh*s could benefit on-line speech comprehension by prompting listeners to pay more attention to upcoming speech. This benefit would be lost if *um*s and *uh*s were absent. This proposal fits well with the findings that *um*s and *uh*s affect listeners' off-line interpretations of talk (Brennan & Williams, 1995; Christenfeld 1995; Fox Tree, 1999).

To test how *um*s and *uh*s affect on-line speech processing, I compared how long it took listeners to detect a word in a speech stream after a spontaneously produced *um* or *uh* with how long it took them to detect the same word when the *um* or *uh* had been digitally excised. In this task, the people listen for a word in an utterance and pressed a button if they hear that word. The speed at which they press the button is related to their ability to integrate information up to that point (Fox Tree, 1995; Fox Tree & Schrock, 1999; Marslen-Wilson & Tyler, 1980). If *um*s and *uh*s are disruptive to on-line comprehension, word monitoring should be slower after an *um* or *uh* than when the *um* or *uh* has been excised. But, if *um*s and *uh*s are beneficial for speech processing, word monitoring should be faster after an *um* or an *uh* than when the *um* or *uh* has been excised. If they are neither detrimental nor beneficial, there should be no effect.

EXPERIMENT 1
English *Um*s and *Uh*s

Method

Participants. Thirty-four native English speakers from the University of California participated in this experiment in exchange for course credit.

Materials. The materials were taken from the spontaneous speech of students telling face-to-face stories to each other. The corpus was collected by Herbert Clark at Stanford University. The materials consisted of a stretch of speech from one speaker that contained at least one spontaneously occurring *um* or *uh*, which was followed immediately by a word that had not occurred earlier in the stimulus (i.e., the *post-um-or-uh-word*). Each stimulus began at the beginning of an idea and finished with a completed thought.

Eighty-eight stimuli were selected: 40 critical stimuli, 40 filler stimuli, and 8 practice stimuli. All stimuli were similar in length and content. In the critical stimuli, the post-um-or-uh-word was the target word. In the filler stimuli, a target word was chosen that was similar semantically and, if possible, phonologically to the post-um-or-uh-word but that did not occur in the stimulus (e.g., if *question* were the post-um-or-uh-word, *query* could be the target word). Target words were chosen from a variety of form classes and were one to four syllables long. The filler stimuli helped ensure that the participants did not adopt a strategy of immediately responding after hearing an *um* or *uh*. This strategy would not work in this case because the stimuli contained 24 *um*s and 17 *uh*s in addition to the 20 *um*s and *uh*s of interest that each participant heard (see the Design section).

Eight of the critical *um*s and eight of the critical *uh*s (40% of the critical stimuli) were matched on the local syntactic constituent structure immediately surrounding the fillers and targets. For example,

(3) Um: And he said why sure well what kind of *um* price range are you looking for?

Uh: Then he also tol- sold her on uh a couple of *uh* furniture items for the ant.

Matching was done to test the role of syntactic location in causing any effects found. For example, if *uh*s usually occur within clauses and *um*s between clauses, as has been found in one Dutch corpus (Swerts, 1998), *uh*s might be more noticeable than *um*s and therefore have a stronger effect. *Um*s and *uh*s tested in the present experiment occurred both within and between clauses.

For each of the critical stimuli and for half the filler stimuli, second versions were created in which the critical *um* or *uh* was digitally excised. The *um* or *uh* was not replaced by a pause, although any pauses preceding or following the *um* or *uh* were retained. The technique of creating edited versions of materials kept constant a number of variables that might otherwise have influenced comparisons between speech with and without *um*s and *uh*s, such as target word frequency or pronunciation, prosodic stress, or syntactic construction. The only difference across conditions was whether or not there was an *um* or an *uh*, as well as any processing time associated with the lengths of each (350 msec on average).

Detectability of editing. A follow-up study tested whether the listeners were able to detect editing in the stimuli. The 80 critical stimuli were divided into two lists, each containing 10 unedited *um* items, 10 edited *um* items, 10 unedited *uh* items, and 10 edited *uh* items. Twenty people who did not participate in the main experiment listened to each list, 10 per list. These participants indicated on an answer sheet whether they thought each trial had been edited. They were asked to "spot the splice" where the materials may have been digitally edited. Two practice trials helped them understand the instructions. The participants performed no better than chance at identifying the edited speech. The inability of the listeners to detect editing is consistent with listeners' inability to detect editing in other similar studies (Fox Tree, 1995; Fox Tree & Schrock, 1999). Although it is possible that splicing can be detected beneath conscious awareness, such detection would be too unsystematic to drive the effects found here; in other similar studies, the edited conditions were sometimes responded to more quickly, sometimes more slowly, and sometimes in the same amount of time as the unedited conditions (Fox Tree, 1995; Fox Tree & Schrock, 1999).

Design. Two lists were created. List 1 contained the practice stimuli, 10 unedited critical *um* stimuli, 10 unedited critical *uh* stimuli, 10 edited critical *um* stimuli, 10 edited critical *uh* stimuli, and 40 filler stimuli. List 2 contained the same practice and filler stimuli, but had the edited versions of the List 1 unedited critical stimuli and the unedited versions of the List 1 edited critical stimuli. The randomized order of presentation was the same across lists. Each participant listened to only one list and therefore heard only one version of each stimulus (a within-subjects and within-items counterbalanced design).

Procedure. Each trial had the following structure. First, a 500-msec tone was heard indicating that the participants should focus their attention on the computer screen. The tone was followed by a 500-msec silence. A word then appeared on the computer screen for 1,000 msec, followed by a 1,000-msec silence, after which a sound file was played. The participants held the word in memory while they listened to the sound file and immediately pressed a button in front of them if they heard the word. Critical stimuli elicited buttonpresses; filler stimuli did not. Reaction times (RTs) were measured from the onset of the target words in milliseconds. Response times timed out after 1,500 msec. All stimuli played to the end, regardless of whether a button was pressed. There was a 1,500-msec silence between trials. The experiment lasted about half an hour.

Results

Three items were removed from the analyses, one because of an experimental error and two because they were not responded to by more than 20% of the participants. The low response items were the first and last items in the stimulus list and, in addition, contained target words that might have been hard to hear. Response times more than two standard deviations from the mean were treated as outliers and removed from the analyses (4% of the data). This eliminated both false alarms and misses. There was no difference in error rates across conditions [unedited *um*: 3.8%, edited *um*: 3.2%, $F_1(1,33) = 0.48$, n.s., $F_2(1,18) = 0.03$, n.s.; unedited *uh*: 5.5%, edited *uh*: 4.7%, $F_1(1,33) = 0.04$, n.s., $F_2(1,17) = 0.29$, n.s.].

The main finding was that there was a significant difference in the participants' speed at recognizing target words following *uh*s compared with their speed at recognizing the same words when the *uh*s had been excised [$F_1(1,33) = 13.77, p = .001; F_2(1,17) = 5.93, p < .05$], but there was no significant difference for the *um*s [$F_1(1,33) = 0.52$, n.s.; $F_2(1,18) = 0.41$, n.s.]. Means and standard deviations are presented in Table 1.

A number of other possible differences between *um*s and *uh*s can be ruled out as factors causing the different

Table 1
Summary of Means (*SD*) for English *Ums* and *Uhs* Averaged Across Participants and Items in Milliseconds

	Unedited		Edited		
	M	*SD*	*M*	*SD*	Difference
Um	561	132	548	126	13
Uh	554	121	601	121	−47

effects of *um* and *uh*. One is the length of the *um*s or *uh*s: If the *uh*s had been longer, the listeners might have paid more attention to them than to the *um*s. In fact, the *um*s tended to be longer than the *uh*s. The *um*s were on average 384 msec long (range, 205–557 msec) and *uh*s were on average 327 msec long [range, 214–451 msec; $t(35) = 1.93$, $p = .06$].

A second difference is the amount of pausing before and after *um*s and *uh*s: If the *uh*s had longer pauses surrounding them, they might have been more prominent in the speech stream. But the pauses before and after *um*s and *uh*s did not vary systematically in this set of sentences [average pause before *um* was 362 msec, average pause before *uh* was 349 msec, $t(35) = .09$, n.s.; average pause after *um* was 334 msec, average pause after *uh* was 355 msec; $t(35) = -.13$, n.s.].

A third possible difference is where the *um*s and *uh*s were located: If the *uh*'s effect was due to its position, the effect should be stronger for the more wide ranging, unmatched *uh* stimuli than for the constrained, matched *uh* stimuli. Similarly, if *um*'s lack of an effect was due to its position, an effect might appear with those *um*s that match the syntactic location of the *uh*s. But there was no interaction between the presence or absence of *um* or *uh* and whether the stimuli were matched or not [editing × matching; *um*: $F_1(1,33) = 0.0$, n.s., $F_2(1,17) = 0.10$, n.s., one item was removed from analysis because of low response; *uh*: $F_1(1,33) = 0.07$, n.s., $F_2(1,16) = 0.45$, n.s., two items were removed from analysis because of low response and experiment error]. Although it may be true that *um*s and *uh*s vary in usual location of occurrence and that the present experiment might have had a different proportion of typical to atypical *um*s versus *uh*s, this variance cannot account for the results found here.

A fourth possible difference is in the target words that followed the *um*s and *uh*s: If *uh* targets were less frequent words, more softly spoken, or longer, maybe they would benefit more from a preceding filler. But there was no difference between the overall RTs to *um* targets and *uh* targets [mean RT *um*, 551, *SD*, 92; mean RT *uh*, 567, *SD*, 89; $t(35) = .55$, n.s.]. So, functionally, targets occurring after *um* and *uh* can be treated as being similar enough to not be considered as the cause of the effects found here.

A fifth possible difference is in the pauses in the edited conditions. After the *um*s and *uh*s were edited out, the original pauses that preceded or followed them were retained in the edited items. Instead of *uh*s' benefitting recognition, the pauses might have slowed recognition. Of course, this does not explain why no similar detrimental effect of pauses was found for the *um* items, given that the amount of pausing before and after *um*s and *uh*s was similar. In fact, there was no correlation between the differences in RTs across conditions and the amount of pausing that remained in the edited conditions (pause before *um* or *uh* plus pause after; $r = .19$, $p =$ n.s.).

A related alternative explanation is that the differences between *um*s and *uh*s were due to coarticulation effects. If targets after *uh* are more likely to be coarticulated with the *uh* than are targets after *um*, removing the *uh* could also remove some important word recognition information. Although the listeners could not detect the editing, they might nonetheless have been deprived of crucial acoustic cues, but only in the *uh* conditions. This explanation can be ruled out by considering that only four *um*s and four *uh*s did not have a pause between them and the targets. Even without intervening pauses, coarticulation effects would be surprising, given that *um*s and *uh*s are never cliticized onto the next word (Clark & Fox Tree, 2000).

Discussion

The present study provides evidence that *uh*s are beneficial to listener's abilities to recognize words in upcoming speech. Neither beneficial nor detrimental effects were found for *um*s. Several alternative explanations for these results have been ruled out; that is, the results were not due to (1) differences in the editing quality across *um* and *uh* stimuli, (2) the lengths of the *um*s and *uh*s, (3) the amount of pausing before or after the *um*s and *uh*s, (4) the syntactic position of the *um*s and *uh*s, or (5) characteristics of the targets following *um*s and *uh*s. The effect also cannot be due to a listener strategy of responding positively immediately after hearing an *uh*. There were many other *uh*s in the materials, both in the critical stimuli and the filler stimuli, so this strategy would have failed more often than it would have succeeded. It is also unlikely that the listeners adopted the strategy of responding faster only after *uh* but not after *um*.

Experiment 2 attempted to replicate these findings cross-linguistically. A further discussion of why *um*s and *uh*s differ is in the General Discussion section.

EXPERIMENT 2
Dutch *Um*s and *Uh*s

Method

Participants. The participants were 32 native Dutch speakers from the Max Planck Institute participant pool. They were each paid FL8.50 for their participation.

Materials. The materials were taken from the spontaneous speech of students describing abstract figures to an experimenter who acted as a silent listener. The corpus was collected by Nanda Poulisse at the Max Planck Institute for Psycholinguistics (Poulisse, 1989). The selection and preparation of the materials were similar to those in Experiment 1, except that there were 10 additional filler stimuli, targets ranged from one to six syllables, and 50% of the critical stimuli were syntactically matched.

Detectability of editing. In the Dutch detection study, the listeners followed along with a transcript and were allowed to hear each stimulus up to six times. Twelve participants (6 per list) listened to each stimulus and were instructed to mark on the transcript where

Table 2
Summary of Means (*SD*) for Dutch *Ums* and *Uhs* Averaged Across Participants and Items in Milliseconds

	Unedited		Edited		
	M	*SD*	*M*	*SD*	Difference
Um	585	119	591	96	−6
Uh	506	84	534	76	−28

they thought it might have been digitally manipulated. A practice trial and discussion of that trial helped them understand the instructions.

The listeners greatly overestimated the amount of editing in this task. Overall, 90% of the stimuli were judged to have been edited at some point, when in fact only 50% had been edited. Among the edited stimuli, 27.5% of the edits were detected. However, on no item was editing accurately detected by all 6 participants who heard it, and on nine items, none of the participants detected the editing. Importantly, there was no difference in the detectability of editing between *um* stimuli and *uh* stimuli [on average, 1.9 participants detected an edit in an *um* stimulus, and 1.4 detected an edit in an *uh* stimulus; $t(38) = 1.25$, n.s., i.e., 32% of the *um* stimuli and 23% of the *uh* stimuli were judged to have been edited].

Design. The design was the same as that in Experiment 1, except that there were 10 additional filler stimuli.

Procedure. The procedure was the same as that in Experiment 1, except that the tone was followed by a 1,500-msec silence, the word appeared for 715 msec, and the word was followed by a 285-msec silence.

Results

Response times more than two standard deviations from the mean were treated as outliers and removed from the analyses (7% of the data). This eliminated both false alarms and misses. There was no difference in error rates across conditions [unedited *um*, 5.9%, edited *um*, 7.4%, $F_1(1,31) = 0.66$, n.s., $F_2(1,19) = 0.63$, n.s.; unedited *uh*, 6.6%, edited *uh*, 8.2%, $F_1(1,31) = 0.62$, n.s., $F_2(1,19) = 1.11$, n.s.].

The results replicated those of Experiment 1: There was a significant difference in the listeners' speed at recognizing the target words following *uh*s compared with their speed at recognizing the same words when the *uh*s were excised [$F_1(1,31) = 7.5$, $p = .01$; $F_2(1,19) = 4.28$, $p = .05$]; there was no significant difference for the *um*s [$F_1(1,31) = 0.04$, n.s.; $F_2(1,19) = 0.35$, n.s.]. Means and standard deviations are presented in Table 2.

Once again, many alternative explanations for the differences between *um*s and *uh*s can be ruled out. One is that, if the *uh*s were longer, listeners might have paid more attention to them than to the *um*s. In fact, the *um*s were longer [average length of *um* was 615 msec, average length of *uh* was 476 msec; $t(38) = 2.20$, $p < .05$]. Another alternative is that the amount of pausing before and after the *um*s and *uh*s could have caused the effect. In contrast to the English materials, in the Dutch materials, pausing occurred more often before and after *um*s than before and after *uh*s [average pause before *um* was 592 msec, before *uh*, 181 msec; $t(38) = 2.77$, $p < .01$; after *um*, 412 msec, after *uh*, 199 msec; $t(38) = 1.92$, $p = .06$]. This extra pausing should have highlighted the presence of an *um*, yet the *um*s had no effect on the listeners' word recognition. There was also no correlation between the differences in RTs across conditions and the amount of pausing remaining in the edited conditions ($r = .12$, $p =$ n.s.). Another alternative is that the location of the *um*s and *uh*s could have driven the effects. But there was no interaction between the presence or absence of *um* or *uh* and whether the stimuli were matched or not [i.e., editing X matching; *um*: $F_1(1,31) = 0.20$, n.s., $F_2(1,18) = 0.75$, n.s.; *uh*: $F_1(1,31) = 0.04$, n.s., $F_2(1,18) = 0.01$, n.s.].

One alternative explanation that cannot be ruled out, as it has been for the English materials, is that there was something about the target words that varied systematically across the *um* and *uh* items. Overall RTs to the *um* targets were slower than to the *uh* targets [mean RT *um*, 589 msec, $SD = 128$; mean RT *uh*, 522 msec, $SD = 69$; $t(29.38) = 2.03$, $p = .05$]. Of course, given the number of post hoc tests (on length of *um* or *uh*, pausing, syntactic location), a p level of .05 might be considered suspect. The slowness of *um* target recognition was not caused by a loss of articulatory information: 16 targets after *um* had an intervening pause (7 targets after *uh* also had an intervening pause). Although it is possible that an effect of *um* on upcoming targets was drowned out by slower processing, this would not explain why no effect of *um* was found in the English experiment.

Discussion

Experiment 2 replicated cross-linguistically the results of Experiment 1 that showed that *uh*s were beneficial to the listeners' recognition of words in upcoming speech, with *um*s being neither beneficial nor detrimental. As before, many alternative explanations were ruled out; that is, the results were not driven by (1) differences in the editing quality across the *um* and *uh* stimuli, (2) the lengths of the *um*s and *uh*s, (3) the amount of pausing before or after the *um*s and *uh*s, or (4) the syntactic position of the *um*s and *uh*s. Unlike the English targets, the Dutch targets after *um* were recognized more slowly than were the targets after *uh* (at $p = .05$, without a Bonferroni correction). Future investigation will clarify what the implications this has for the differences between *um* and *uh*.

GENERAL DISCUSSION

In two word-monitoring studies, in two different languages, *uh*s were found to increase the speed at which listeners were able to recognize words in upcoming speech, whereas *um*s were found to have no effect on the listeners' speed of recognition. The fact that *um*s and *uh*s are frequently avoided in speech or edited from reported speech suggests that there may be something wrong with them. Against this backdrop, it is surprising that *um*s and *uh*s do not inhibit on-line processing and, even more surprising, that *uh*s aid processing.

There are two alternative explanations for the present findings. According to the editing account, it is not that *uh*s speed up RTs, but that editing slows them. Even if people cannot overtly detect editing, their RTs may still be hindered by an editing by-product that occurs beneath overt awareness. According to the *time-to-target account*, splicing out *um*s and *uh*s shortens the processing time available to listeners, slowing responses in the edited conditions. Both accounts are unlikely for several reasons.

The editing account is unlikely because (1) it cannot explain why editing has not always yielded a negative effect, either in prior studies (see Fox Tree, 1995; Fox Tree & Schrock, 1999) or in the present studies, in which only the *uh* stimuli would have been so affected; (2) it cannot explain why similar effects were found in both English and Dutch, despite stimulus variation that could have presumably affected such an editing by-product, such as the fact that the materials were from two languages that sound different phonologically, and that the stimuli were created by using different digitizing and editing systems and were played on different kinds of audio equipment; and (3) it cannot explain why RTs are sometimes faster in edited conditions (see Fox Tree, 1995). The difficulty in explaining why a hypothetical subdetection editing signal would hinder responses in only certain situations leads to a rejection of the editing account.

The time-to-target account is unlikely because (1) there was no correlation between pausing and RT differences in either experiment; (2) in other similar experiments, whether the edited item was replaced by a pause or not had no affect on the results (Fox Tree, 1995; Fox Tree & Schrock, 1999); and (3) it cannot explain why slowing would occur only for the *uh*s and not for the *um*s; this is especially problematic given that the *um*s tended to be longer than the *uh*s, which means that more processing time would have been lost in the edited versions of the *um* stimuli, incorrectly predicting a stronger effect for *um*s. Given the lack of concordance between the processing time predictions and the three outcomes listed above, this account has been rejected.

The present results pose problems for parsing models that describe parsing as a process that involves identifying words in the speech stream, assigning them grammatical roles, and fitting them into syntactic representations. According to these models, the occurrence of an *um* or an *uh* should lead to parsing failure since they do not combine with surrounding words to form syntactic constructions. The negative reputation of *um*s and *uh*s in the public eye fits nicely with this predicted negative influence on word recognition. The problems for these models posed by the nondetrimental effects found here may seem to be easily overcome by an automatic filtering system account, in which *um*s and *uh*s are detected and filtered out before the processes of word identification and syntactic construction begin. This filtering process should be effortless and have no effect on processing. Unfortunately, an automatic filtering system would also remove the demonstrated benefits of *uh*.

The present results are consistent with the proposal that *um*s and *uh*s signal different lengths of upcoming delay. The brief delays signaled by *uh* heighten listeners' attention for upcoming speech. The longer delays after *um* do not appear to alter listeners' attention in the same way, either because heightened attention is not as useful when the length of delay is indeterminant, or because listeners focus their attention elsewhere in an effort to help speakers get their thoughts out, or for some other reason.

The present research strongly suggests that utterances with *um*s and *uh*s excised cannot automatically be considered better versions of the originals. In fact, cleaned-up versions might be lacking important information that listeners use to process spontaneous speech. Instead of their being undesirable, speakers may choose to use one or the other, and addressees seem to make use of the distinction.

REFERENCES

Brennan, S. E., & Williams, W. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory & Language*, **34**, 383-398.

Carroll, D. W. (1994). *Psychology of language* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Christenfeld, N. (1995). Does it hurt to say um? *Journal of Nonverbal Behavior*, **19**, 171-186.

Christenfeld, N., & Creager, B. (1996). Anxiety, alcohol, aphasia, and ums. *Journal of Personality & Social Psychology*, **70**, 451-460.

Christenfeld, N., Schacter, S., & Bilous, F. (1991). Filled pauses and gestures: It's not coincidence. *Journal of Psycholinguistic Research*, **20**, 1-10.

Clark, H. H. (1994). Managing problems in speaking. *Speech Communication*, **15**, 243-250.

Clark, H. H., & Fox Tree, J. E. (2000). Uh *and* um *as English words*. Manuscript in preparation.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, **37**, 201-242.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory & Language*, **34**, 709-738.

Fox Tree, J. E. (1999). *Between-turn pauses and ums*. Paper presented at the International Conference of Phonetic Sciences Satellite Meeting on Disfluencies, Berkeley, CA.

Fox Tree, J. E., & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory & Language*, **40**, 280-295.

Jefferson, G. (1974). Error correction as an interactional resource. *Language in Society*, **3**, 181-199.

Kasl, S. V., & Mahl, G. F. (1987). Speech disturbances and experimentally induced anxiety. In G. F. Mahl (Ed.), *Explorations in nonverbal and vocal behavior* (pp. 203-213). Hillsdale, NJ: Erlbaum.

Lalljee, M. G., & Cook, M. (1969). An experimental investigation of the function of filled pauses in speech. *Language & Speech*, **12**, 24-28.

Lalljee, M. [G.], & Cook, M. (1973). Uncertainty in first encounters. *Journal of Personality & Social Psychology*, **26**, 137-141.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lindsay, J., & O'Connell, D. C. (1995). How do transcribers deal with audio recordings of spoken discourse? *Journal of Psycholinguistic Research*, **24**, 101-115.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, **75**, 19-44.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, **8**, 1-71.

Martin, J. G. (1967). Hesitations in the speaker's production and listener's reproduction of utterances. *Journal of Verbal Learning & Verbal Behavior*, **6**, 903-909.

Martin, J. G. (1971). Some acoustic and grammatical features of spontaneous speech. In D. Horton & J. Jenkins (Eds.), *The perception of language* (pp. 47-68). Columbus, OH: Merrill.

Martin, J. G., & Strange, W. (1968). The perception of hesitation in spontaneous speech. *Perception & Psychophysics*, **3**, 427-438.

Mitchell, D. C. (1994). Sentence parsing. In M. A. Gernsbacher (Ed.),

Handbook of psycholinguistics (pp. 375-409). San Diego: Academic Press.
Postma, A., Kolk, H., & Povel, D.-J. (1990). On the relation among speech errors, disfluencies, and self-repairs. *Language & Speech*, **33**, 19-29.
Poulisse, W. (1989). *The use of compensatory strategies by Dutch learners of English.* Doctoral dissertation, de Katholieke Universiteit te Nijmegen.
Reynolds, A., & Paivio, A. (1968). Cognitive and emotional determinants of speech. *Canadian Journal of Psychology*, **22**, 164-175.
Rochester, S. R. (1973). The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research*, **2**, 51-81.
Schacter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality & Social Psychology*, **60**, 362-367.
Schegloff, E. A. (1981). Discourse as an interreactional achievement: Some uses of 'uh huh' and other things that come between sentences. In D. Tannen (Ed.), *Georgetown University round table on languages and linguistics, 1981.* Washington, DC: Georgetown University Press.
Shriberg, E. E., & Lickley, R. J. (1992). *The relationship of filled-pause F0 to prosodic context*. Paper presented at the IRCS Workshop on Prosody in Natural Speech, University of Pennsylvania.
Shriberg, E. E., & Lickley, R. J. (1993). Intonation of clause-internal filled pauses. *Phonetica*, **50**, 172-179.
Siegman, A. W. (1979). Cognition and hesitation in speech. In A. W. Siegman & S. Feldstein (Eds.), *Of speech and time* (pp. 151-178). Hillsdale, NJ: Erlbaum.
Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory & Language*, **32**, 25-38.
Svartvik, J., & Quirk, R. (Eds.). (1980). *A corpus of English conversation.* Lund: Gleerups.
Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, **30**, 485-496.
Tannenbaum, P. H., Williams, F., & Hillier, C. S. (1965). Word predictability in the environments of hesitations. *Journal of Verbal Learning & Verbal Behavior*, **4**, 134-140.

(Manuscript received December 9, 1999;
revision accepted for publication June 26, 2000.)