Journal of Memory and Language **44**, 274–296 (2001)
doi:10.1006/jmla.2000.2753, available online at http://www.academicpress.com on IDEAL®

How Listeners Compensate for Disfluencies in Spontaneous Speech

Susan E. Brennan

State University of New York at Stony Brook

and

Michael F. Schober

New School for Social Research

Listeners often encounter disfluencies (like *uhs* and repairs) in spontaneous speech. How is comprehension affected? In four experiments, listeners followed fluent and disfluent instructions to select an object on a graphical display. Disfluent instructions included mid-word interruptions (*Move to the yel- purple square*), mid-word interruptions with fillers (*Move to the yel- uh, purple square*), and between-word interruptions (*Move to the yellow- purple square*). Relative to the target color word, listeners selected the target object more quickly, and no less accurately, after hearing mid-word interruptions with fillers than after hearing comparable fluent utterances as well as utterances that replaced disfluencies with pauses of equal length. Hearing less misleading information before the interruption site led listeners to make fewer errors, and fillers allowed for more time after the interruption for listeners to cancel misleading information. The information available in disfluencies can help listeners compensate for disruptions and delays in spontaneous utterances.

Key Words: repairs; disfluencies; pauses; fillers; spontaneous speech; comprehension; parsing; paralinguistic cues.

Spontaneous human speech is notoriously disfluent. Speakers hesitate, interrupt themselves mid-phrase or mid-word, repeat or replace words, abandon phrases to start afresh, and season their talk with expressions like *um*, *uh*, *or*, *I mean*, and *oh*. A conservative estimate (excluding silent hesitations) for the rate of disfluencies in spontaneous speech is 6 words per 100 (Fox Tree, 1995). A similar rate of 5.97 per 100 words was found in a corpus of speech by young, middle-aged, and older people, married and strangers, about familiar and unfamiliar topics (Bortfeld, Leon, Bloom, Schober, & Brennan, in press). Disfluency rates may be even higher in certain content domains (Schachter, Christenfeld, Ravina, & Bilous, 1991), although they appear to be lower in speech directed at machines (Oviatt, 1995).

Disfluent speech poses a *continuation problem* for listeners (Levelt, 1989), who must edit out disfluencies in order to make sense of speakers' utterances. Consider a hypothetical listener who hears *Call Don Harw- I mean, Harrison*. Recovering a fluent version requires determining that there is a problem with the utterance, what the problem is, and how to repair it (including how far to back up to replace corrected information); to do this, the parser must identify three adjoining intervals, the *reparandum*, the *edit interval*, and the *repair interval* (see Levelt, 1983; Nakatani & Hirschberg, 1994). The reparandum contains fluent speech up until the *interruption site*, where the speaker leaves off speaking fluently. In the case of an overt error

This material is based on work supported by the National Science Foundation under Grants IRI9402167, IRI9711974, and SBR9730140. We are grateful to Elizabeth Shriberg, Arthur Samuel, Fred Conrad, Richard Gerrig, Charles Metzing, and three anonymous reviewers for their comments on an earlier draft as well as to Jonathan Bloom, Ricardo Carrion, Julia Kung, Angela Lawrence, Maria Malzone, Kimiko Ryokai, Leora Schefres, Darron Vanaria, and Maurice Williams for their assistance in preparing the stimuli and running the experiments.

Address correspondence and reprint requests to Susan E. Brennan, Department of Psychology, State University of New York, Stony Brook, NY, 11794-2500 or Michael F. Schober, Department of Psychology, AL-303, New School for Social Research, 65 Fifth Avenue, New York, NY, 10003. E-mail: susan.brennan@sunysb.edu or schober @newschool.edu.

0749-596X/01 $35.00

repair, this interval contains material that the speaker finds problematic (in our example, *Harw-*). The edit interval begins at the interruption site and ends with the onset of the repair (here, *I mean* serves as an editing expression or explicit comment that the speaker has made an error; Hockett, 1967; Levelt, 1989). The repair interval may or may not retrace material from the reparandum (in our simple example, the repair *Harrison* does not include any retracing). Determining these intervals is not a trivial task, as developers of machine speech recognition systems have discovered (Core & Schubert, 1999; Nakatani & Hirschberg, 1994; Shriberg, Bear, & Dowding, 1992).

Evidence about how people process spontaneous speech (with all its natural disfluencies) is scant, as most studies of human speech comprehension have focused on fluent, idealized utterances. The tacit assumptions of these studies seem to be that (1) disfluencies uniformly present obstacles to comprehension and (2) disfluencies need to be excluded in order to study comprehension in its "purest" form. We question both of these assumptions. Regarding the first assumption, some disfluencies may actually present information that listeners can use to help compensate for what might otherwise hinder processing, as we discuss shortly. As H. H. Clark (1994, 1996) has argued, people have a number of resources for managing the process as well as the content of conversation. That is, not only do they produce and interpret utterances that address the main purposes at hand, but they produce informative secondary (or paralinguistic) signals about the utterances themselves.

As for the second assumption, studying comprehension under normal "noisy" conditions should be just as important as studying it in "pure" conditions. As Fox Tree (1995) has noted, many current approaches to parsing the structure and interpreting the meaning of incoming utterances are built on data about sanitized utterances and cannot handle disfluencies at all. Successful comprehension of spontaneous, disfluent speech is particularly intriguing given the fact that much of the time, listeners don't experience disfluencies as disruptive, and when they do detect disfluencies, they have trouble categorizing or locating them precisely (see Bond & Small, 1984; Cooper, Tye-Murray, & Nelson, 1987; Ferber, 1991; Fromkin, 1973; Laver, 1973; Martin & Strange, 1968; Tent & Clark, 1980). Instead, listeners make the appropriate parsing decisions, solve the continuation problem, and interpret speakers' intentions without much apparent difficulty. In the studies that follow, we examine how disfluencies, ubiquitous as they are in spontaneous speech, affect and inform comprehension processes.

HOW MIGHT SPEECH DISFLUENCIES INFORM COMPREHENSION?

As many speech production studies have shown (e.g., Dell, 1986; Fromkin, 1973; Garrett, 1975; Levelt, 1989; Smith & Clark, 1993), spontaneous speech is systematically shaped by the problems speakers encounter while planning messages, retrieving lexical items, and articulating a speech plan. For instance, analyses of corpora show that interruptions are usually located very close to the word that is the source of the trouble, since speakers monitor their speech closely and tend to interrupt themselves just as soon as they detect trouble (Levelt, 1989; Nooteboom, 1980; although see Blackmer & Mitton, 1991, who argue that when interruptions are followed immediately by repairs, the trouble must have been detected before the interruption). If the problem is detected *after* a troublesome word has already been uttered, the speaker is somewhat more likely to finish the current word (but not the current phrase) before stopping to repair it (Levelt, 1983, 1989). Levelt found that when an interruption occurs midword, the problem is with the interrupted word itself (Levelt, 1983, 1989); so an interrupted word may signal what a speaker does *not* mean.

Another systematicity within speech disfluencies is in the occasional use of editing expressions like *I mean*, *sorry*, or *that is*. Sixty-two percent of the repairs in Levelt's (1983) corpus of spontaneous task-oriented utterances included some type of editing expression. The most common of these was *er* (corresponding to what many American English speakers pronounce as *uh*), occurring in 30% of repairs. *Er* seemed to mark those repairs made very early; the most

frequent use of *er* in Levelt's corpus was in covert repairs (that is, *er* filled a hesitation in which no problematic word was actually uttered). Often, *er* was used immediately after an interruption, as in *Left- er- right in front of me* (Levelt, 1989, p. 484), and its use declined as the interruption occurred further away from the source of the problem. In other words, *er* was more likely to be used the earlier a problem was detected.

While linguists and psycholinguists have considered disfluencies largely from a production standpoint, computational linguists have considered them from a recognition standpoint, often with the goal of improving machine recognition of spontaneous speech. This work has focused on characterizing disfluencies in speech corpora, identifying potential cues that disfluencies have occurred, identifying repairs and reparanda, and comparing the performance of algorithms based on various combinations of cues. Hindle (1983), for instance, originally suggested that an "edit signal" serves as a cue that fluent speech has been interrupted. Although no evidence for a single such cue has been found (Bear, Dowding, & Shriberg, 1992; Lickley & Bard, 1998; Nakatani & Hirschberg, 1994), several corpus studies have found *combinations* of cues that could be used by algorithms to identify disfluencies and repairs with reasonable success (Bear, Dowding, & Shriberg, 1992; Nakatani & Hirschberg, 1994; Shriberg et al., 1992; Shriberg, 1999). These potential cues include word or syllable lengthening, interrupted words, glottalization, laryngealization, silent pauses in the edit interval, fillers and other editing expressions, and increased stress on a repair word versus a reparandum word.

Our question is how *human* listeners cope with disfluent input. A handful of linguists and psycholinguists have looked at disfluencies from the listener's point of view, proposing that cues such as interruptions, hesitations, and prosody in spontaneous speech may help listeners solve the continuation problem and that perhaps listeners can infer speakers' intentions by exploiting regularities in the distributions of types of speech errors. Lickley and Bard (1998) used a word gating paradigm (increasing the length of utterances one word at a time while playing them over and over to listeners) to discover how much information is necessary for detecting disfluency. They found that nearly 80% of the disfluencies in their corpus were detectable (rated "disfluent" or "maybe disfluent") at the first word gate in a repair, sometimes even before lexical access of that word had occurred. Fox Tree's (1995) studies of listeners in Dutch and English demonstrated that while some fresh starts slowed word-monitoring performance, repetitions did not. Brennan and Williams' (1993) studies, using as stimuli spontaneous answers to general knowledge questions, showed that listeners were able to use the information available in pauses of various lengths and in fillers (*um* and *uh*) vs silent pauses to correctly judge how confident the speakers were in their answers to the questions. And in a study of synthesized speech played to listeners, pauses before repairs and stress on repair words led to judgments of higher comprehensibility and helped listeners initiate repetitions of the disfluent utterances more quickly (Howell & Young, 1991).

While these studies represent first steps toward discovering how listeners use the information in spontaneous disfluent speech, the tasks (gating and judging if disfluencies were present, monitoring for words, rating answers, choosing which of two disfluent utterances with and without a particular feature was more comprehensible, and repeating a fluent version of a just-heard disfluent utterance) differ from what listeners do when they hear spontaneous speech. One goal of the present studies was to examine the processing of disfluencies in a comprehension task that is closer to what people do in everyday language use.

If the forms of some disfluencies bear useful information, there should be situations in which a target word in a disfluent utterance is more quickly comprehended (without loss of accuracy) than in a comparable utterance in which the disfluency, or one of its features, is absent. Consider a situation in which there are two objects mutually known to a speaker and a listener (say, *the purple mug* and *the yellow mug*). If the speaker begins to name one and then stops and names the other (*hand me the pur- uh, yellow*

mug), the interruption (*pur- uh*), which displays what the speaker did not intend (*purple*), might tip the listener off early as to what the speaker *does* intend (*yellow*). In this situation the listener may be faster to recognize the speaker's intentions to refer to a target object in a disfluent utterance than in a comparable, more fluent utterance (for instance, *hand me the <pause> yellow mug*).

For the studies that follow, we set up just such a situation: Listeners tried to pick out a unique referent on a display in response to fluent and disfluent versions of the same utterances. They heard disfluent utterances spontaneously produced by one speaker, like *Move to the orange-purple circle* or *Move to the pur- uh, yellow square*, and we measured their comprehension by seeing how quickly and accurately they could press a key corresponding to the target square. We used spontaneously produced disfluencies because of the possibility that artificially created or performed ones would sound unnatural (Fox Tree, 1995). Listeners also heard three kinds of relatively fluent comparison utterances. The most obvious comparison is with (1) spontaneous fluent utterances by the same speaker, as in *Move to the yellow square*. Because such a fluent utterance may differ from a disfluent utterance in prosody or other uncontrolled characteristics, we also created (2) disfluency-excised controls in which the disfluency (in this case, *pur- uh*) was electronically excised from a copy of the disfluent utterance and the remaining parts "zipped up" to create *Move to the yellow square*. Because both the fluent and disfluency-excised utterances are shorter in duration than the disfluent utterances, we also created (3) controls in which the disfluency was removed from a copy of the utterance and replaced with a pause of equal length, as in *Move to the <pause> yellow square*. Any comprehension difference for a disfluency may be due to the form of the disfluency or it could be due simply to timing. The disfluency-replaced-by-pause controls thus provide the most stringent comparison because they hold time constant. These and the disfluency-excised controls have the additional advantage that they vary disfluency within the same token of an utterance.

In Experiments 1 and 2 we examine two hypotheses (not mutually exclusive) about how certain features of disfluencies may support the repair process in comprehension such that the net effects of disfluencies are not harmful. Hypothesis 1 is that mid-word interruptions (like *yell- orange*) are better signals than between-word interruptions (like *yellow- orange*) that a word was produced in error and that the speaker intends to replace it. This follows Levelt's (1989, p. 481) proposal that "by interrupting a word, a speaker signals to the addressee that that word is an error. If a word is completed, the speaker intends the listener to interpret it as correctly delivered." Hypothesis 2 is that interruptions marked by the filler *uh* (like *yell- uh orange*) are better error signals than interruptions without *uh* (like *yell- orange*). This follows Levelt's (1989, p. 481) proposal that an editing expression like *er* or *uh* may "warn the addressee that the current message is to be replaced." These two hypotheses are not in competition; they simply specify two kinds of features that may help in monitoring and repair.

The logic of our studies, then, is that certain cues in disfluent speech may be informative to listeners (as opposed to being simply noise to be filtered out). If so, then there should be faster and no less accurate comprehension of target words following disfluencies than target words in utterances in which these cues are absent. On the other hand, if disfluencies lack compensatory features, then listeners should choose targets more slowly and less accurately after disfluent utterances than after the controls. If Hypothesis 1 is correct, mid-word interruptions should facilitate comprehension more (or harm comprehension less) than between-word interruptions. If Hypothesis 2 is correct, interruptions marked by *uh* should facilitate comprehension more (or harm comprehension less) than those that are not. Experiment 1 compares response times and error rates for three types of disfluent utterances to three types of edited and natural controls in a logically constrained context, in which listeners select objects from two-object referent arrays. Experiment 2 decreases the informativeness of the disfluency by increasing the number of potential referents. Experiment 3 examines whether

the disfluency advantage found in Experiments 1 and 2 is due to the phonological or temporal characteristics of fillers, and Experiment 4 examines the effect when an interruption does not always predict a replacement.

COLLECTION OF SPONTANEOUS FLUENT AND DISFLUENT UTTERANCES FOR EXPERIMENTS 1–4

We began by collecting a database of spontaneous utterances to use as stimuli. A volunteer who was a male native speaker of English and naive to our purposes produced the utterances. He viewed a series of simple computer displays of geometric objects (orange, yellow, purple, red, green, and blue squares and circles) shown three at a time and arranged horizontally on the screen against a white background. After a display of three objects appeared, there was an orienting tone and one object was highlighted. The speaker's task was to say out loud: *Move to the <highlighted object>*. Between trials, the screen was blank. After 10 practice trials, the speaker described 150 displays in each of three sessions.

Meanwhile, the experimenter who had instructed the speaker sat at a similar display in the same room and carried out his instructions by moving her cursor to the target object (neither could see the other's display). She interacted freely with the speaker (e.g., by saying "okay" when she had moved). This minimal feedback proved necessary in order to elicit spontaneous and natural-sounding utterances from the speaker. (In our pilot attempts to collect instructions with no addressee present, speakers appeared to fall into a routine and produced utterances with the mechanical-sounding intonation characteristic of reading a list.)

We elicited disfluent utterances using a procedure modified from van Wijk and Kempen (1987). Our speaker was aware that the highlight might jump to another object and was instructed to update his instructions in those cases as quickly as he could. In about two-fifths of the trials, the highlight suddenly jumped to another object, and the rest of the time it remained on the same object. When the highlight remained the same, about one-third of the time it flickered slightly (as if it might jump). Meanwhile, a second experimenter monitored the speaker's utterances in another room and controlled the timing of the highlights using a remote keyboard attached to the speaker's computer in order to elicit disfluencies and self-corrections from the speaker. The speaker was unaware of the presence of the second experimenter until he was debriefed. The second experimenter hit a key during each trial, usually while the speaker was speaking, which sometimes made the highlight jump to another object. The speaker produced disfluencies or self-corrections in about 34% of the trials. These were produced not only in response to the changing highlight, but also apparently because of interference; the displays appeared in rapid succession and were quite similar. About three-quarters of the displays involved objects with two-syllable colors (*yellow*, *purple*, or *orange*) to increase the likelihood of the speaker interrupting himself mid-word.

Since our goal in this project was to study the effects of disfluencies upon listeners, we used only one speaker so that listeners would not need to calibrate to different voices (see Nygaard, Sommers, & Pisoni, 1994) and so that their task would be less unnatural than if they had several (virtual) partners.

Stimuli

The audiotapes of fluent and disfluent spontaneous utterances were transcribed and checked. The utterances were then digitized using Signalyze for the Macintosh. Those that mentioned objects with two-syllable colors (yellow, purple, or orange) were categorized as (1) 27 mid-word interruptions followed immediately by replacement of a color word, such as *Move to the yel- purple square*; (2) 12 mid-word interruptions followed by *uh* and then the replacement of a color word, such as *Move to the yel- uh, purple circle*; (3) 30 between-word interruptions followed immediately by replacement of color words, such as *Move to the yellow- purple square*; (4) 195 fluent utterances; and (5) *other* utterances, which included all directions that did not fall into the first four categories. These consisted of restarts that involved repeating parts of phrases, such as *Move to the yel- to the purple square*, as well as utterances with more than one

restart or with a restart of the shape word, as in *Move to the yel- orange cir- square*. Fluent and disfluent utterances involving one-syllable color words (red, green, and blue) were not categorized or used. Both authors and a research assistant categorized the utterances. Disfluent utterances from the three disfluency categories that had been difficult to categorize or that were associated with any background noise were eliminated from further consideration. This left 14 between-word interruptions, 14 mid-word interruptions, and 6 mid-word interruptions with fillers, for a total of 34 *Naturally Disfluent* utterances. These served both as the critical stimuli and as the basis for the digitally edited controls. Two involved the speaker's replacing *yellow* with *orange*, 12 *orange* with *yellow*, 8 *yellow* with *purple*, 5 *purple* with *yellow*, 5 *orange* with *purple*, and 3 *purple* with *orange*. We chose a total of 68 utterances from the 195 fluent utterances to serve as fillers and as natural controls. These *Fluent* utterances were chosen randomly, but such that their color words were in roughly the same relative proportions as the second (target) color words in the Naturally Disfluent utterances (there were 12 orange, 24 purple, and 32 yellow objects mentioned in the Fluent controls).

In order to vary fluency within utterances, we made digital copies of each of the 34 spontaneous disfluencies and used SoundEdit Pro for the Macintosh to edit them. For each disfluency, we created a *Disfluency-replaced-by-pause* version by replacing the first color word (and any filler associated with it) with a silent pause of equal length. The silent pause contained the same ambient room sound as the material it was replacing. Recall that the purpose of these items was to enable us to examine whether an interrupted word is more informative about a speaker's intention than a silent pause. In addition, we created a *Disfluency-excised* version by removing the interrupted color word and any associated filler, shortening the utterance by an equal amount (exactly equal to the length of the pause inserted into the Disfluency-replaced-by-pause version and with the same edit points).

Finally, for each utterance, a reference point for the target color word was determined using SoundEdit Pro so that response latencies relative to the onsets of the target word could be calculated. These points were located as precisely as possible at the earliest point where information about the onset of the target color word (*yellow*, *purple*, or *orange*) was available. For each trio based on the same utterance (Naturally Disfluent, Disfluency-excised, and Disfluency-replaced-by-pause), the reference points were exactly the same with respect to the target color word, since the members of a trio were based on the same spontaneously disfluent utterance. Reference points for the onsets of each of the 68 *Fluent* controls were measured independently.

In all, the main set of stimulus utterances consisted of 34 spontaneously disfluent utterances, 34 edited Disfluency-replaced-by-pause versions, 34 edited Disfluency-excised versions, and 68 spontaneous Fluent utterances, for a total of 170 utterances. So one-fifth of the items contained an overt lexical disfluency and four-fifths did not. Since each utterance (e.g., *move to the orange square*) consisted of 5–6 words (counting the word fragments in the disfluencies), this amounts to a disfluency rate of less than 4 per 100 words (less than Fox Tree's, 1995, and Bortfeld et al.'s, in press, estimated rate of approximately 6 disfluencies per 100 words). Three-fifths of the utterances in the set were spontaneously produced and two-fifths were digitally edited. In Experiments 1 and 2, only 6 utterances (< 4% of all utterances) contained fillers; in Experiments 3 and 4, only 12 and 13, respectively, contained fillers (< 7%).

All the disfluent utterances had their interruption sites within or immediately following the word to be repaired. These sorts of disfluencies are relatively frequent in spontaneous speech; for instance, 69% of the interruption sites in Levelt's (1983) corpus occurred within (18%) or immediately after (51%) problem words. Interrupted words appear to be even more common in repairs of speech directed at machines (Bear et al., 1992, found 60% of repairs in their corpus contained word fragments, and Nakatani & Hirschberg, 1994, found 73%).

Prosodic Characteristics of the Stimuli

Pitch accents increase the prominence of words in speech; they can convey that something is new

TABLE 1

Stimuli: *F*0 of Color Words (in Hertz)

Stimuli	Reparandum color word	Target color word
Fluent instructions		104.0
Naturally disfluent instructions		
Between-word	102.8	125.1
Mid-word	101.3	117.6
Mid-word w/filler	102.8	124.7
Mean (naturally disfluent)	102.8	122.0

rather than given or that it contrasts with other information mentioned in an utterance (Pierrehumbert & Hirschberg, 1990). Repair words that supply new semantic content are stressed (with a higher *F*0) relative to the words they are correcting (Bear et al., 1992; Howell & Young, 1991; Levelt & Cutler, 1983; O'Shaughnessy, 1992; Shriberg et al., 1992), while repeated words are not (Bear et al., 1992; O'Shaughnessy, 1992). To characterize our stimuli, we measured *F*0 peaks on each color word (see Table 1). Target color words in Naturally Disfluent utterances had higher *F*0s than those in Fluent utterances, $t(100) = 8.04$, $p < .001$. And in the Naturally Disfluent utterances, the *F*0 of (target) color words in repairs averaged 19.2 Hz higher than the *F*0 of color words in reparanda, $t(34) = 9.47$, $p < .001$. This difference was consistent for all three types of disfluencies. There was no difference in *F*0 between the first (reparandum) color words in Naturally Disfluent utterances and those in Fluent utterances.

Ratings of Edited and Nonedited Stimuli

Before presenting the stimuli to listeners as instructions to follow in a comprehension task, we needed to determine whether listeners could hear the electronic edit points in the edited controls. We also wanted to determine whether the Disfluency-excised and Fluent utterances were perceptually distinguishable (since lexically and syntactically they were exactly the same). Sixteen undergraduate Stony Brook students (11 women and 7 men) listened to the set of 170 utterances and judged the degree to which they sounded natural or electronically edited. The students, native speakers of English, volunteered in exchange for research credit in a psychology course.

Each utterance was played once in a different random order for each listener, using SuperLab on a Macintosh computer. After each utterance, listeners took as much time as necessary to make a rating on a 7-point scale where the left endpoint was labeled *(1) Edited*, the midpoint was labeled *(4) ????*, and the right endpoint was labeled *(7) Natural*. Listeners participated alone and made their ratings using the top row of number keys (1–7) on the keyboard. Particular care was taken to instruct listeners about was meant by *Natural* and *Edited*, with the instructions: "Some of the utterances will be played naturally, that is, exactly as the speaker spoke them, and others have been electronically edited, that is, changed using the computer afterward." Since we wanted to know whether the electronic edit points were detectable, we needed listeners to judge whether utterances sounded as if they had been edited *electronically* as opposed to whether the speaker himself corrected (or "edited") them as he was speaking. Mean naturalness ratings were calculated on a scale of 1 (edited) to 7 (natural).

Ratings are shown in Table 2. There was no evidence that listeners were able to detect the electronic edits. Disfluency-replaced-by-pause utterances were rated as no more edited than Naturally Disfluent (nonedited) utterances, $t1(15) = .67$, *ns*; $t2(33) = 1.21$, *ns*. This suggests that Disfluency-replaced-by-pause utterances are appropriate controls for testing the effects of the presence and absence of a disfluency within the same utterance. Although they were worded identically, Fluent and Disfluency-excised utterances were *not* rated alike by listeners; Fluent utterances were rated as more natural than Disfluency-excised ones, $t1(15) = 5.04$, $p < .001$; $t2(100) = 14.90$, $p < .001$. Disfluency-excised utterances may have sounded less natural because of interrupted intonation contours (as opposed to any audible edits).

Even though none of the Naturally Disfluent utterances contained electronic edits, mid-word interruptions with fillers were rated as less

TABLE 2

Stimuli: Ratings of Edited and Nonedited Instructions (1 = *Edited*, 7 = *Natural*)

Disfluency type	Naturally disfluent	Disfluency-replaced-by-pause	Disfluency-excised	Fluent
Between-word (n = 14)	3.88	3.80	3.75	
Mid-word (n = 14)	3.67	3.62	3.50	
Mid-word w/filler (n = 6)	4.86	4.22	4.23	
Mean	3.97	3.80	3.75	5.63

edited/more natural than the other two types considered together, planned comparison, $F1(1, 15) = 9.53$, $p < .007$; $t2(31) = 4.69$, $p < .001$. Fluent utterances were rated as only slightly (.77) more natural than mid-word interruptions with fillers, significant by items but not by subjects, $t1(15) = 1.28$, $p = .22$; $t2(72) = 4.28$, $p < .001$.

Disfluency-replaced-by-pause utterances appear to be the most appropriate edited controls of Naturally Disfluent utterances for three reasons: they vary disfluency within items, there are identical *F*0s and delays before the target words as with Naturally Disfluent utterances, and both these types of utterances were judged as equally electronically nonedited. However, we included the Disfluency-excised and Fluent utterances as additional controls in case the pauses within Disfluency-replaced-by-pause controls had any detrimental effects on comprehension. Neither Disfluency-excised nor Fluent controls had any lexical disfluency or hesitation; target words in Disfluency-excised utterances had higher *F*0s than those in Fluent utterances, while Fluent utterances were judged as less edited.

EXPERIMENT 1

Methods

Participants. The participants were 50 undergraduate students (37 women and 13 men) from the State University of New York at Stony Brook who volunteered in exchange for research credit in a psychology class. None had participated in generating or rating the stimuli, and all identified themselves as native speakers of English (those who were bilingual had learned English before the age of 8). Data from 5 students who produced a large total number of errors (placing them distinctly outside the distribution of the other participants) were replaced. Those replaced made errors approximately 2 *SD*s (or more) above the mean of the others. One additional student was replaced for not following instructions.

Design and stimuli. Each listener heard the same set of Naturally Disfluent, Disfluency-replaced-by-pause, Disfluency-excised, and Fluent utterances (two of the 68 Fluent controls were omitted due to a programming error). So approximately 80% of the utterances contained no lexical disfluency. Of the 20% that did, 14 contained between-word interruptions before the repair, 14 contained a mid-word interruption, and 6 contained a mid-word interruption and a filler. There were two versions of the experiment (A and B) for counterbalancing purposes; if the target object for an utterance appeared on the left-hand side in List A, it appeared on the right in List B and vice versa. Half the listeners received each list.

Procedure. Stimuli were presented on a Macintosh Quadra computer by the SuperLab program and a Sony speaker. Two keys on the keyboard were used for input. Participants viewed a series of displays of pairs of geometric objects (either two squares or two circles) arranged horizontally on the screen; with each display, they heard an utterance instructing them about a target object. They were instructed to press the key on the same side as the object intended as the target by the speaker. When a key was pressed, the object corresponding to that key was highlighted on the screen. Listeners were told that they should press the correct key as quickly as possible, and if they pressed the

TABLE 3

Experiment 1 (Two-Object Displays): Response Times in Milliseconds from Onset of Target Color Word (Error Rates in Parentheses)

Disfluency type	Naturally disfluent	Disfluency-replaced-by-pause	Disfluency-excised	Fluent
Between-word	682 (.19)	667 (.02)	702 (.02)	
Mid-word	673 (.11)	675 (.03)	707 (.01)	
Mid-word w/filler	617 (.05)	695 (.03)	726 (.02)	
Mean	666 (.13)	675 (.02)	708 (.02)	719 (.02)

wrong one, they should then press the correct one as quickly as possible (and that the next trial would not begin until they pressed the correct one). One second after the correct key was pressed, the next trial began. Half of the time, the target object appeared on the right and half, on the left. After 15 practice trials, the 170 experimental and control trials began, appearing in a different random order for each listener. Listeners were told to respond as quickly and as accurately as possible.

Responses and response times were collected for each trial. Time was measured from the start of each utterance until the correct key was pressed. Response times were then calculated relative to the correct target color words by subtracting the time of onset of the target color word (established earlier in order to provide a reference point) from the response time for that trial. Response times from trials in which responses were incorrect or times were greater than approximately 3 *SD*s from the mean (1400 ms) were discarded.

Results and Discussion

For this experiment and the ones that follow, we first computed a difference score for each item within which disfluency was varied (via an edited and intact version of the same token). We did this by subtracting the item's Naturally Disfluent response time from its Disfluency-replaced-by-pause response time whenever both responses were correct. These difference scores represent a *disfluency advantage* whenever they were positive-that is, when the presence of a disfluency led to faster response than its absence (in the form of a pause of equal length) in the same utterance. We examined these difference scores using a repeated-measures ANOVA whose factors were disfluency type (between-word interruption, mid-word interruption, and mid-word interruption with filler) and side of the target object.[1] Planned comparisons were used to examine the impact of the interruption point (between vs mid-word) as well as the impact of the filler (mid-word interruptions with vs without fillers). The same planned comparisions were used to examine error rates in an additional repeated-measures ANOVA comparing the three kinds of disfluency types. An additional ANOVA was used to compare responses for Naturally Disfluent utterances to those for nonedited Fluent controls. Table 3 shows response times and error rates for each type of utterance.

Hypothesis 1 predicted a comprehension advantage for target words after mid-word versus between-word interruptions. This hypothesis was supported by a difference in error rates: mid-word interruptions without fillers led to lower error rates than between-word interruptions, $F1(1, 49) = 17.67$, $p < .001$; $t2(31) = 3.35$, $p = .002$. Interrupting the reparandum mid-word prevented listeners from committing themselves to the wrong interpretation. However, Hypothesis 1 was not supported by the response-time difference scores: those for target

[1] Target side did not affect response times for Naturally Disfluent, Disfluency-replaced-by-pause, and Disfluency-excised utterances. For Fluent utterances, right-side targets were 30 ms slower than left-side targets, $F1(1, 49) = 20.00$, $p < .001$, $F2(1, 65) = 15.69$, $p < .001$. This may be due to a tendency for listeners to scan the display from left to right. Since side was counterbalanced across the two versions of the experiment, we do not report it further.

words after between-word interruptions were no different (15 ms slower than Disfluency-replaced-by-pause controls) than those for target words after mid-word interruptions (2 ms faster than Disfluency-replaced-by-pause controls), planned comparison, $F1(1, 48) = .03$, *ns*, $t2(31) = .20$, *ns*.

As for Hypothesis 2, which predicted that marking interrupted words with fillers would be helpful, there was a clear response time disfluency advantage. Difference scores for mid-word interruptions with fillers (78 ms) were higher than those without (2 ms), supporting the idea that fillers may signal the replacement of an interrupted word, planned comparison, $F1(1, 48) = 15.76$, $p < .001$; $t2(31) = 3.75$, $p = .001$. This response-time advantage is particularly interesting because it occurred with relatively little cost in accuracy: Listeners were inaccurate only .05 of the time after mid-word interruptions marked with fillers, much less than with mid-word interruptions not so marked, $F1(1, 49) = 13.53$, $p = .001$; $t2(31) = 2.05$, $p < .05$, and at a rate only marginally higher than after Fluent utterances, $F1(1, 49) = 3.47$, $p < .07$; $t2(96) = 1.62$, *ns*. The implication is that while disfluent utterances can be misleading, those with fillers are less so.

Listeners chose correct targets after Naturally Disfluent utterances 53 ms faster than after Fluent ones, $F1(1, 49) = 20.83$, $p < .001$; $F2(1, 98) = 13.63$, $p < .001$. Responses to all three types of disfluent utterances were significantly faster than responses to Fluent utterances (between-word interruptions vs Fluent utterances, $F1(1, 48) = 12.80$, $p = .001$; $t2(96) = 1.95$, $p < .07$; mid-word vs Fluent, $F1(1, 48) = 12.67$, $p = .001$; $t2(96) = 2.43$, $p < .02$; mid-word with filler vs Fluent, $F1(1, 48) = 38.12$, $p < .001$; $t2(96) = 3.58$, $p = .001$. Although listeners must wait longer to hear the target color word in a disfluent utterance than in a fluent utterance, their speeded response to the target word in a disfluent utterance partially compensates for this delay. As for accuracy, listeners responded no less accurately to mid-word interruptions with fillers than to Fluent utterances. However, they did respond less accurately to both of the other types of disfluencies than to Fluent utterances [between-word interruptions vs Fluent utterances, $F1(1, 49) = 36.27$, $p < .001$; $t2(96) = 15.05$, $p < .001$, and mid-word interruptions vs Fluent utterances, $F1(1, 49) = 24.66$, $p < .001$; $t2(96) = 8.28$, $p < .001$].

That responding to *orange* after a between-word interruption like *yellow- orange* could be faster than responding to *orange* in a Fluent utterance seems surprising at first. This advantage could be due to the target color word in a Naturally Disfluent utterance receiving contrastive stress relative to the first color word (as in *Move to the yellow- ORANGE square*) (see Cutler, 1983; Levelt & Cutler, 1983; Shriberg et al., 1992). Converging support for this possibility comes from the finding of marginally speeded response times for Disfluency-excised over Fluent utterances (different by subjects but not by items), $F1(1, 49) = 4.07$, $p < .05$; $F2(1, 98) = .80$, *ns*.

Overall, responses to Fluent, Disfluency-excised, and Disfluency-replaced-by-pause utterances were equally (and highly) accurate; none of these utterances contained unintended color word information that could have caused interference or false alarms in the task. The response time means in Table 3 show the same pattern of differences among the three types of disfluencies for both Disfluency-excised and Disfluency-replaced-by-pause controls. That the times for Disfluency-replaced-by-pause controls were no longer than for Disfluency-excised controls suggests that silent pauses may not be harmful to comprehension (we consider pauses further in Experiments 3 and 4).

These results show that that some features of disfluencies enable listeners to partially compensate for any potential disruption or delay in comprehension. When a speaker interrupts an unintended word rather than completing it before repairing it, this disadvantages the listener less. And when the speaker marks the interruption with a filler, the listener is not only more accurate but also faster to recognize the correct target word than when the speaker does not do so. These results support speculations by Clark (1994, 1996) and Levelt (1989) that a word fragment lets listeners know that the speaker is having difficulty with that word.

Experiment 1 used a logically constrained context in which there were only two objects to choose from. Interruptions may have signaled to

listeners that the speaker was having trouble with a word or, more specifically, that he or she meant to cancel that word. With only two possibilities, if one is cancelled, then the other is certain. Next, we examined the disfluency effect in a context where knowing that a word is troublesome is somewhat less informative.

EXPERIMENT 2

In Experiment 2, we expanded the set of possible referents to reduce the information value of the disfluencies: Listeners chose from three objects rather than two. With three objects, when the speaker interrupts himself with *yell- uh, orange*, the information in the reparandum (*yell- uh*) may signal an intent to cancel *yellow*, but if it does, this no longer uniquely determines which of the other two objects the speaker must mean. So if the disfluency advantage in Experiment 1 depended entirely on an inference enabled by the unusual logical constraints of the context, then the response time advantage should be eliminated with three objects. On the other hand, if an advantage still occurs (perhaps in attenuated form, from having to choose among three objects rather than two), this supports the idea that a disfluency can aid comprehension by signaling what a speaker is having trouble with.

Methods

Participants. Fifty native speakers of English (31 women and 19 men) volunteered to participate in exchange for research credit in an introductory undergraduate psychology course. None had participated previously. One additional subject was replaced because of technical difficulties and another was replaced for making a large number of errors (> 2 *SD*s above the mean).

Design, stimuli, and procedure. Experiment 2 used the same software, hardware, and stimuli as described earlier (68 Fluent utterances and 34 Naturally Disfluent utterances matched to 34 Disfluency-replaced-by-pause and 34 Disfluency-excised versions). The procedure, utterances, and practice trials were the same as those in Experiment 1, with the following minor differences: Listeners were told that if they pressed the wrong key, they would hear a sound (rather than getting a visual highlight as in Experiment 1) and that the next trial would begin automatically, shortly after they made their selection. The important difference from Experiment 1 was that each visual display showed three rather than two objects (either an orange, a purple, and a yellow square or an orange, a purple, and a yellow circle, arranged horizontally in varying order). To create the three-object displays, we modified the displays from Experiment 1 by adding a third (irrelevant) object to the same position (left, middle, or right) on both the List A and List B displays for a particular utterance so that for all Naturally Disfluent items, when the object mentioned in the reparandum appeared to the left of the object mentioned in the repair in one list, it appeared to the right in the other. Irrelevant objects appeared equally often in the left, middle, and right positions. Each listener experienced the trials of either List A or List B in a different random order. For input, listeners used both index fingers and the middle finger of their dominant hand, poised over three adjacent keys. Incorrect or extremely slow response times (>1600 ms, ~3 *SD*s from the mean) were discarded.

TABLE 4

Experiment 2 (Three-Object Displays): Response Times in Milliseconds from Onset of Target Color Word (Error Rates in Parentheses)

Disfluency type	Naturally disfluent	Disfluency-replaced-by-pause	Disfluency-excised	Fluent
Between-word	781 (.08)	743 (.01)	762 (.02)	
Mid-word	761 (.04)	749 (.01)	751 (.01)	
Mid-word w/filler	741 (.01)	769 (.00)	789 (.02)	
Mean	765 (.05)	750 (.01)	762 (.02)	786 (.01)

Results and Discussion

Response times along with error rates are shown in Table 4. Listeners were noticeably slower to choose a referent out of a set of three than out of a set of two; response times for Fluent utterances were 67 ms slower than in Experiment 1. Although listeners received the same instructions to try to be both fast and accurate in both experiments, three-object displays presented a harder task than did two-object displays, and this seems to have made listeners trade off speed for accuracy. Experiment 2's listeners took longer to respond but got more trials correct, making on average 3.5 errors to Experiment 1's 7.3.

As before, we compared within-item response-time difference scores (Disfluency-replaced-by-pause minus Naturally Disfluent) to test for a relative advantage for mid-word as opposed to between-word interruptions (Hypothesis 1). Just as before, there was no difference, planned contrast, $F1(1, 49) = 2.27$, *ns*, $t2(31) = .91$, *ns*, but as before, listeners were more accurate after mid-word than between-word interruptions, $F1(1, 49) = 12.43$, $p < .001$, $t2(31) = 2.33$, $p < .05$. Once again, difference scores for mid-word interruptions with fillers were greater than those without (Hypothesis 2), planned contrast, $F1(1, 49) = 15.32$, $p < .001$, $t2(31) = 2.39$, $p < .05$. As before, this response-time disfluency advantage incurred little cost in accuracy; errors were less common with mid-word interruptions with fillers than without (by subjects but not by items), $F1(1, 49) = 5.62$, $p = .02$; $t2(31) = .97$, *ns*. In fact, errors occurred no more often to utterances with fillers than to Fluent utterances (1% of which incurred errors), $F1(1, 49) = 0.05$, *ns*; $t2(98) = .14$, *ns*. So listeners responded to target words faster and more accurately after mid-word interruptions marked with fillers than those not so marked, and this disfluency advantage for response times did not occur for between-word interruptions or mid-word interruptions not marked with *uh*. This pattern of results across the three types of disfluencies is the same as that for two-object displays.

The important difference between responses to two- and three-object displays is that the whole pattern of difference scores shifted downward with three objects (see Fig. 1). The difference score for utterances with fillers was only 28 ms (down from 78 ms in Experiment 1), reliably different from zero by subjects but not by items, $F1(1, 49) = 8.86$, $p = .005$; $F2(1, 5) =$

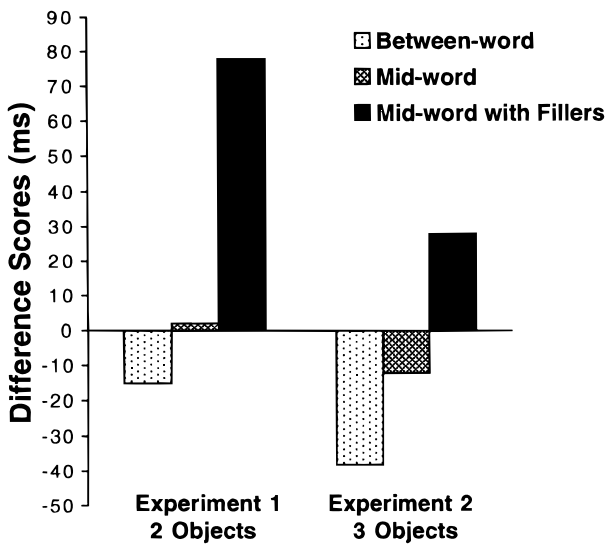

FIG. 1. Disfluency advantage (Disfluency-replaced-by-pause minus Naturally Disfluent response times) for the three kinds of disfluencies in Experiments 1 and 2.

2.71, $p = .161$, *ns*. That the filler's value as a cue is attenuated but not eliminated when listeners choose between three objects as opposed to two suggests that knowing what word the speaker is having trouble with helps listeners cancel the effect of hearing the unintended word.

Although listeners balanced speed and accuracy somewhat differently with three- than with two-object displays, making fewer errors with three-object displays, the pattern of errors remained the same. Mid-word interruptions with fillers were the only kind of disfluency with an error rate as low as that of Fluent utterances. Listeners were less accurate after between-word interruptions than after Fluent utterances, $F1(1, 49) = 29.62$, $p < .001$; $t2(98) = 7.74$, $p < .001$. They were also less accurate after mid-word interruptions than after Fluent utterances, $F1(1, 49) = 10.74$, $p < .005$; $t2(98) = 2.84$, $p < .01$.

As before, stress (in the form of elevated *F*0) appeared to contribute to response times. Listeners responded to target words in Disfluency-excised utterances 24 ms faster than in Fluent ones, planned contrast, $F1(1, 49) = 14.11$, $p < .001$; $F2(1, 98) = 2.82$, $p < .10$. Target side had no effect.

Apart from the finding that increasing the set of possible referents attenuated the disfluency advantage for mid-word interruptions, the patterns of response time and error data parallel those in Experiment 1. When listeners must choose from a set of three rather than two objects, the advantage of an interruption marked with *uh* is attenuated, but not eradicated. The response time advantage for the disfluency may diminish because the information value of the cue is lower, because it takes longer to choose from three alternatives than two, or for both reasons. Whether the disfluency is marked with a filler still contributes strongly to the error rates; responses to utterances containing mid-word interruptions marked with *uh* were just as accurate as responses to Fluent utterances.

EXPERIMENT 3

So far, listeners' responses to target color words after mid-word interruptions marked by fillers have been faster than when these disfluencies were absent (in both Disfluency-replaced-by-pause and naturally Fluent utterances). What is still unclear is why. What features of the utterance are responsible for the speeded responses and lower error rates, and what is the underlying process that takes advantage of these features? The disfluency advantage could be due to the phonological form of the filler alone; certain fillers could heighten listeners' vigilance to an upcoming target word (Fox Tree, 1993). If so, then a version minus the interrupted word and containing only the filler should be processed just as quickly as the nonedited Naturally Disfluent versions and faster than the Disfluency-replaced-by-pause versions. Alternatively, the response time advantage could be due to the phonological form of the filler *in combination with* the interrupted word. If the filler acts as an editing expression signaling that an interrupted word is being cancelled, then the Naturally Disfluent utterances should be processed faster than versions in which either the interrupted word *or* the filler are edited out. Finally, the advantage could be due to the extra time that elapses during the filler, together with the information about the cancelled target word in the reparandum. If this last hypothesis is true, then a signal that a word is to be replaced may help only when there is sufficient time to process such information; whether the interrupted word is followed by a filler or a silent pause of equivalent length should make no difference. Experiment 3 teased apart these three possibilities.

Although listeners in Experiments 1 and 2 were instructed to respond quickly, they sometimes waited until well after the end of the utterance before responding. Since we were interested in the incremental effects of disfluencies, Experiment 3 encouraged listeners to respond more quickly by placing a moderate deadline on their responses.

Methods

Participants. Forty-eight undergraduate psychology students (25 women and 23 men) participated in exchange for research credit in a psychology course. None had participated previously, and all identified themselves as native speakers of English. Data from 3 students who produced a large total number of errors and time-

TABLE 5

Intervals of Interest in the Stimuli for Experiments 3 and 4

Stimuli		Reparandum	Edit interval	Repair (target)
1. Disfluent	Move to the	ye-	uh,	orange square
2. Filler-removed	Move to the	ye-	|---|	orange square
3. Word-removed	Move to the	|---|	uh,	orange square
4. Disfluency-replaced-by-pause	Move to the	|---	---|	orange square
		Target		
5. Disfluency-excised	Move to the	| orange square		
6. Fluent	Move to the	orange square		

Note. Utterances are aligned to symbolize relative time elapsed. Vertical bars show location of electronic edits and horizontal lines show relative extent of silent pauses.

outs (placing them distinctly outside the distribution of the other participants) were replaced. Those replaced made errors and timeouts approximately 2 *SD*s (or more) above the mean.

Design and stimuli. Experiment 3 used the same stimuli as the previous experiments plus 12 additional items (see Table 5) that were edited from duplicates of the 6 Naturally Disfluent mid-word interruptions with fillers for a total of 182 stimuli. Six of the new items had the filler replaced with a pause of equal length (Filler-removed items), and 6 had the interrupted word replaced with a pause of equal length (Word-removed items). All displays consisted of two objects as in Experiment 1, and these were counterbalanced for target side using two versions of the displays as before.

Procedure. The same software, hardware, and input method were used as in the previous experiments. When listeners did not respond within 1000 ms after the onset of the target word, the trial timed out and a "too slow" message appeared before the next trial proceeded automatically.

Results and Discussion

The critical comparisons for Experiment 3 were those that contrasted features of the disfluencies within utterances, using four versions of each mid-word interruption with filler: Naturally Disfluent, Filler-removed, Word-removed, and Disfluency-replaced-by-pause. Correct response times to these four versions (as opposed to difference scores, which represent a comparison between only two versions) were compared in an ANOVA that included display side, with planned comparisons to test for relative contribution of fillers and mid-word interruptions. The results were clear: With respect to both response times (Fig. 2) and error rates, Filler-removed items were similar to Naturally Disfluent (nonedited) items and Word-removed items were similar to Disfluency-replaced-by-pause items. Listeners responded about 80 ms more slowly to utterances in which mid-word interruptions were replaced by silent pauses (leaving only the filler) than to the nonedited versions, $F1(1, 47) = 25.11$, $p < .001$; $F2(1, 5) = 64.92$, $p < .001$. And listeners responded just as quickly to utterances that kept the interrupted word but replaced the filler with a silent pause as they did to the nonedited versions, $F1(1, 47) = .004$, *ns*, $F2(1, 5) = .036$, *ns*. While removing the filler from Naturally Disfluent utterances did not harm response times, the presence of the filler in the Word-removed utterances led to faster response times than the Disfluency-replaced-by-pause versions by subjects but not by items, $F1(1, 47) = 4.16$, $p < .05$; $F2(1, 5) = 2.86$, $p = .15$. So it appears that some remodeling is in order for Levelt's (1989) original idea that a filler may be an informative cue, at least in our task situation. The facilitating aspect of a filler after a mid-word interruption appears to be *not* the phonological form of the filler, but the extra

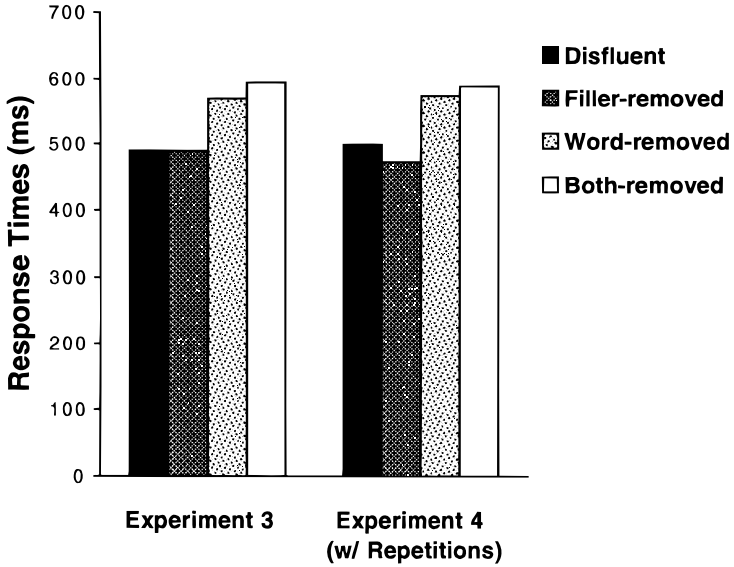

FIG. 2. Effects of interrupted words and fillers in Experiments 3 and 4.

time that elapses after the interruption, while the filler is being produced.

Error rates, which were extremely low for all four versions of the mid-word interruptions with fillers, were lowest when misleading information in the form of interrupted color words was removed, planned comparison of Word-removed and Naturally Disfluent versions, $F1(1, 47) = 5.96$, $p < .02$; $F2(1, 5) = 18.95$, $p < .01$. Disfluency-replaced-by-pause and Word-removed versions, the two versions with no misleading color information, had equally low error rates, $F1(1, 47) = .02$, *ns*; $F2(1, 5) = .00$, *ns*. And removing the filler from Naturally Disfluent versions made no difference, planned comparison, $F1(1, 47) = .20$, *ns*; $F2(1, 5) = .43$, *ns*. Error rates and response times are displayed in Table 6.

TABLE 6

Experiments 3 and 4 (Two-Object Displays): Response Times in Milliseconds from Onset of Target Color Words (Error Rates in Parentheses)

Stimuli	Experiment 3	Experiment 4
Fluent	628 (.01)	612 (.01)
Naturally Disfluent		
Replaced color words		
Between-word	525 (.14)	532 (.19)
Mid-word	541 (.09)	539 (.09)
Mid-word w/filler	490 (.03)	498 (.02)
Repeated Color Words		548 (.03)
Disfluent edited		
Filler-removed	489 (.03)	473 (.03)
Word-removed	571 (.00)	574 (.02)
Both word and filler removed (Disfluency-replaced-by-pause)	595 (.00)	589 (.02)

Difference scores (Disfluency-replaced-by-pause minus Naturally Disfluent response times) showed the same pattern as in Experiments 1 and 2. Difference scores for mid-word interruptions were no higher than for between-word interruptions, $F1(1, 47) = 1.74$, *ns*; $t2(31) = .64$, *ns*. As before, mid-word interruptions led to fewer errors than did between-word interruptions (by subjects but not by items), $F1(1, 47) = 8.21$, $p = .006$; $t2(31) = 1.62$, $p < .116$. Difference scores for mid-word interruptions with fillers were higher than for those without, $F1(1, 47) = 20.82$, $p < .001$; $t2(31) = 3.36$, $p = .002$. As before, there was no penalty to the disfluency advantage in response times: Mid-word interruptions with fillers led to fewer errors than did those without fillers (by subjects but not by

items), $F1(1, 47) = 10.14$, $p = .003$; $t2(31) = 1.48$, $p = .148$.

However, under moderate time pressure, listeners made more use of the information in disfluencies: Difference scores were positive not only for the mid-word interruptions with fillers, but for the other two types of disfluencies as well (unlike in Experiments 1 and 2). So this time, listeners responded more quickly to Naturally Disfluent between- and mid-word interruptions (without fillers) such as *Move to the purple- orange square* than they did to the Disfluency-replaced-by-pause versions of the same utterances such as *Move to the <pause> orange square*. Figure 3 shows that while the relative pattern of difference scores for the three types of disfluencies remained the same across Experiments 1, 2, and 3, the difference scores shifted upward with the earlier responses in Experiment 3. If we assume that recognition of the (highly predictable) color word in the repair took well under 200 ms (see Marslen-Wilson & Tyler, 1981) and the key press another 200 ms, then this places the listener's decision point after the color word in the repair (mean response times ranged from nearly 500 ms in Experiment 3 to the high 700s in Experiment 2). So by the time listeners made their decisions, they had had ample evidence that the reparandum word or word fragment was being cancelled from both the semantic and the prosodic contrasts between the repair and reparandum. Those under moderate time pressure (Experiment 3) benefited more from hearing cues in the disfluencies than those under less time pressure (Experiments 1 and 2). In the Disfluency-replaced-by-pause and Word-removed versions, without a good account for the hesitation before the target word, listeners could tell only that the speaker may have had some (covert) trouble.

So far, we have found that error rates are lowest for utterances with no misleading color word information, next lowest for those containing mid-word interruptions marked by a filler, next lowest for those not so marked, and finally, highest for between-word interruptions. However, we have not yet distinguished two competing explanations for this effect: Are the lower error rates for mid-word interruptions (compared to between-word interruptions) due to having a discrete cue in the form of the interrupted word, or are they simply due to hearing less information that is misleading-a continuously varying cue?

To address this question, we conducted a post hoc analysis of the stimuli. We measured

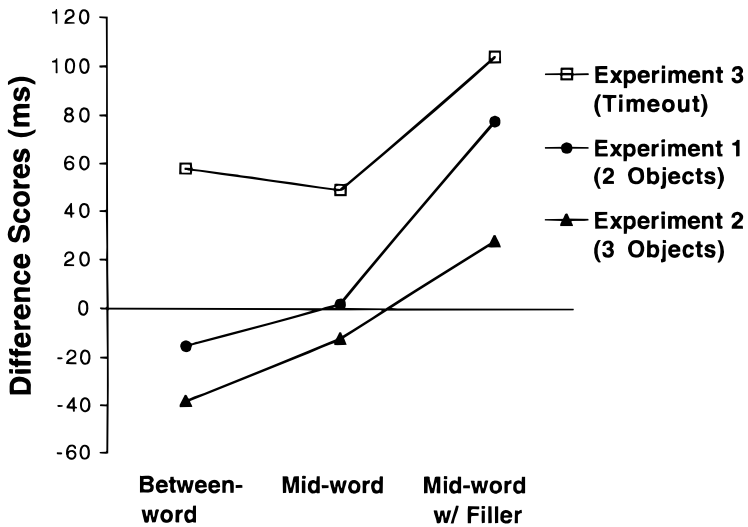

FIG. 3. Disfluency advantage for Experiments 1, 2, and 3.

the lengths of two intervals of interest: the reparandum and the editing interval (or delay before the repair). The reparandum in a disfluent utterance began with the onset of the *un*intended color word and ended with the interruption site. Mean reparandum lengths averaged 328, 276, and 175 ms for between-word, mid-word, and mid-word interruptions with fillers, respectively. The editing interval began with the interruption site and ended with the onset of the intended color word (corresponding to the beginning of the repair); this averaged 71, 94, and 390 ms for between-word, mid-word, and mid-word interruptions with fillers, respectively.[2] For each disfluent utterance, we computed correlations between each of these two intervals and the two dependent measures of interest: mean error rate and mean response time difference score (Disfluency-replaced-by-pause minus Naturally Disfluent) for each disfluent utterance in Experiment 3. To determine whether any reliable associations were carried solely by the items with fillers (which, after all, yielded the largest effects), we computed each correlation both with and without the six filler items.

Error rates were correlated with the lengths of the reparanda (or amount of misleading information), $r(33) = .462$, $p < .001$, but not with the lengths of the edit intervals (or delay before repair), $r(33) = -.305$, *ns*. This is consistent with the interpretation that listeners make fewer errors when they hear less misleading information rather than depend on an interrupted word as a discrete cue. On the other hand, the magnitude of disfluency advantages in response times was correlated with the lengths of the edit intervals, $r(33) = .560$, $p < .001$, but not with the lengths of the reparanda, $r(33) = -.228$, *ns*. So the longer the delay before the repair, the more hearing the information in the disfluency may have speeded response times. These results are not due simply to the mid-word interruptions with fillers (which have especially short reparanda and long editing intervals); when those utterances are removed from the correlations, the pattern is the same (although slightly attenuated). This evidence converges with our conclusion from the comparison of Filler-removed vs Naturally Disfluent items: It is not the phonological form of the filler that speeds recognition of the target color word in our task, but the processing time that elapses during the filler.

EXPERIMENT 4

The disfluencies in Experiments 1–3 were 100% informative that the color word before the interruption would be replaced by another color word. But interrupted utterances can also be continued with repeated material; an interruption does not always signal that a word will be cancelled. In a study of a spontaneous conversations during which pairs of speakers did a matching task (Bortfeld et al., in press), the speakers interrupted themselves and then repeated (without correcting) portions of their utterances fairly frequently (1.47 times every 100 words) although less often than interrupting themselves and replacing reparanda with new material (1.94 times every 100 words). When we elicited the disfluencies for the current experiments, we had expected to collect many tokens of repetitions such as *Move to the yell- yellow square* (in fact, that is why we set up about one-fifth of our elicitation trials so that the highlight of the object flickered but then remained on the same object). But the speaker in our corpus produced only seven instances of immediately repeated color words, not nearly enough to provide comparison tokens for our three types of interruptions (between-word, mid-word, and mid-word with filler).

[2] The lengths of these intervals fall within the range of those reported by Blackmer and Mitton (1991) for their corpus of natural examples from talk-show conversations. Their overt repairs had editing intervals that averaged 332 ms but that were sometimes as short as 0 ms. In Nakatani and Hirschberg's (1994) corpus, edit intervals averaged 289 ms after word fragments and 481 after nonfragments. In Lickley's (1996) sample of 30 disfluent utterances, silent pauses at the interruption site ranged from 34 to 1134 ms. That the edit intervals after nonfragments in our stimuli were so short suggests that the speaker in our limited domain did not need to take much time for replanning repairs. For the purpose of algorithmically identifying pauses associated with edit intervals, O'Shaughnessy (1992) proposed that these pauses should range from 80 to 400 ms; this is consistent with our edit intervals.

That the stimuli in Experiments 1–3 included no repetitions raises the possibility that the speeded response times may have resulted from some special strategy on the part of listeners. Listeners may have learned in our experiments that interruptions always signaled replacement. If this were the case, then reducing the informativeness of interruptions should make the disfluency advantage go away. Experiment 4 ruled this possibility out by including some interruptions followed by repetitions.

Methods

Participants. Fifty Stony Brook undergraduate native speakers of English (29 women and 21 men) participated in exchange for research participation credit in their psychology course. None had participated in any of the previous experiments. Data from 7 students who produced a large total number of errors and timeouts (2 *SD*s or more above the mean of the other participants) were replaced.

Design, stimuli, and procedure. For the experiments so far, the informativeness of interruptions in the task context has remained the same. In Experiments 1 and 2, of the 170 experimental trials and 15 practice trials, 37 (or 20%) included reparanda containing all or part of a color word, and 100% of the time these corresponded to a change of color word in the repair. In Experiment 3, the percentage of replaced color words increased slightly, to 22%, due to the addition of six items with interrupted color words for which the fillers were removed. It is possible that listeners adjusted to this predictability and responded via an unnatural strategy such as deciding what key to press before they heard the repair. Of the seven tokens of utterances with repeated color words in our database, one was a between-word interruption, four were mid-word interruptions, one was a mid-word interruption with a filler, and one was a mid-word interruption that recycled a short stretch of the previous utterance (*Move to the yel- to the yellow square*). These were too varied and too rare to use for controlled comparisons in Experiment 4; however, we included them in order to reduce the informativeness of interrupting a color word. So the informativeness of interruptions was reduced from 100 to 82%; that is, about one time in seven, the speaker did *not* cancel what was in the reparandum. Given our corpus, this was the most by which we could reduce the informativeness of interruptions without resorting to electronically manufacturing disfluent utterances. Note that, unlike the other disfluent utterances in which the color words were repaired, the utterances with repeated color words did not have the target color word contrastively stressed; *F*0s were the same for the first and second color words, $t(6) = 1.04$, *ns*.

One of the seven utterances with a repeated color word was presented in the practice trials so that listeners would be exposed at the outset to the possibility that color words could be repeated as well as replaced; the other six were included among the experimental trials, presented as before to listeners in different random orders. Apart from the seven new utterances, Experiment 4 used the same stimuli, procedure, and analyses as Experiment 3.

Results and Discussion

Table 6 shows response times and error rates. Even though interrupting an utterance after or within the color word no longer predicted that this word would be cancelled 100% of the time, results were consistent with those in Experiment 3: Response times for filler-removed utterances were most similar to those for Naturally Disfluent utterances, and response times for Word-removed utterances were most similar to those for Disfluency-replaced-by-pause utterances (see Fig. 2). As before, replacing the interrupted word with a silent pause slowed response times, planned comparison, $F1(1, 49) = 62.89$, $p < .001$; $F2(1, 5) = 107.99$, $p < .001$. Once again, it seems that it is not the phonological form of fillers that speeds processing; in fact, responses were slightly faster (but only marginally) after interrupted words when fillers had been replaced with pauses of equal length, planned comparison, $F1(1, 49) = 3.07$, $p = .09$; $F2(1, 5) = 6.87$, $p < .05$.

As before, error rates with utterances that included misleading interrupted color words (Naturally Disfluent mid-word interruptions with fillers and Filler-removed utterances) were no

different, planned comparison, $F1(1, 49) = .23$, *ns*; $F2(1, 5) = .31$, *ns*. And utterances without misleading information (Word-removed and Disfluency-replaced-by-pause versions) were no different in either response times, $F1(1, 49) = 2.74$, *ns*; $F2(1, 5)$, *ns*, or error rates, $F1(1, 49) = .53$, *ns*; $F2(1, 5) = .60$, *ns*. Unlike in Experiment 3, however, there was no decrease in the (already quite low) error rate when the interrupted color word was replaced with a pause, planned comparison, $F1(1, 49) = .04$, *ns*; $F2(1, 5) = .90$, *ns*. Finally, error rates for the nonedited versions (mid-word interruptions with fillers) were no different than for Fluent utterances, $F1(1, 49) = .20$, *ns*; $t(2) = .25$, *ns*.

For the different types of nonedited disfluent utterances (between-word, mid-word, and mid-word interruptions with fillers), response-time difference scores (Disfluency-replaced-by-pause minus Naturally Disfluent response times) and error rates followed the same pattern as in Experiments 1–3. Difference scores for between-word interruptions were the same as for mid-word interruptions without fillers, $F1(1, 49) = .51$, *ns*; $t2(31) = .59$, *ns*. Mid-word interruptions incurred fewer errors than did between-word interruptions, $F1(1, 49) = 35.25$, $p < .001$; $t2(31) = 2.97$, $p < .01$. Difference scores for mid-word interruptions with fillers were higher than those without, $F1(1, 49) = 17.56$, $p < .001$; $t2(31) = 3.26$, $p = .003$, and incurred fewer errors, $F1(1, 49) = 29.98$, $p < .001$; $t2(31) = 1.68$, $p = .10$. As in Experiment 3, difference scores for all three types of disfluencies were positive. That is, listeners responded to target words in all Naturally Disfluent utterances more quickly than to target words in Disfluency-replaced-by-pause utterances.

As for the six trials with repeated color words (or word fragments) that were included to break the perfect correlation between interruptions and replacing words, listeners made errors only 3% of the time. If listeners were treating an interrupted color word as a cue that the word would be replaced, then listeners should have made more errors with repeated color words. The fact that this error rate was so low is consistent with the proposal that the less misleading information in the reparandum, the less likely listeners are to choose the wrong object (utterances with repetitions contained disfluencies but no misleading color information). Correct responses to instructions with repeated color words were relatively fast, averaging 453 ms from the onset of the final color word (see Table 6). However, in 28 cases (9.6%) the correct response came *before* the onset of the final (repeated) color word, suggesting that in these cases listeners responded to the color word in the reparandum. In contrast, early correct responses were virtually nonexistent for disfluent instructions in which color words were *replaced* (because the first color word was always incorrect). If we adjust the means for repetitions by counting the 28 early responses as incorrect, then correct response times would average 548 ms and error rates 13% (similar to the data for between- and mid-word interruptions with replacements in Table 6). So even though instructions with repeated color words were relatively rare in our stimuli compared to those with replaced color words, they didn't appear to slow listeners down or lead them astray any more than did disfluencies in which color words were replaced. This result suggests that listeners did not adopt a strategy of simply treating an interrupted color word as a cue by itself that the word would be replaced.

GENERAL DISCUSSION

Across all four experiments, listeners responded to target words after disfluencies that had long edit intervals (e.g., *Move to the pur- uh yellow square*) faster than when disfluencies were absent (e.g., *Move to the yellow square*, with or without a silent pause before *yellow*). This was true whether the display included two objects or three (with three objects, the disfluency advantage for utterances with fillers was attenuated but still reliably positive). Together, the results show that there is information in disfluencies that partially compensates for any disruption in processing. This disfluency advantage was even clearer when listeners were under time pressure (Experiments 3 and 4); they were also able to use the information in disfluencies without fillers (with shorter edit intervals, e.g., *Move to the purple- yellow square*) to respond faster than when disfluencies were replaced by pauses of equal length.

In Experiments 1–4, regardless of how listeners traded off speed and accuracy, the relative pattern of error rates was the same (between-word interruptions led to more errors than mid-word interruptions, which led to more errors than mid-word interruptions with fillers). Of course, the reason measurable errors occurred at all was because our task required speakers to make a commitment; in more natural comprehension settings, reaching the wrong interpretation seldom leads to an irrevocable commitment. For future work we are eager to use an online technique, such as eye tracking, that is sensitive to the incremental decisions that are made (as well as unmade or postponed) during the interpretation of spontaneous speech.

From these data, we can construct an incipient account of how listeners are able to compensate for the disfluencies they hear in spontaneous speech. We began with predictions inspired by some ideas of Levelt's (1989); however, our data support an account that is not as simple as expected. Hypothesis 1, that mid-word interruptions should be easier for listeners to recover from than between-word interruptions, was supported by the consistent pattern of error data across all the experiments. However, the interrupted word does not appear to act as a discrete cue; rather, the less misleading information listeners hear, the less likely they are to make a commitment to the wrong interpretation. Listeners responded equally quickly to the target (repair) word whether the preceding reparandum was interrupted mid- or between-word.

As for the idea that fillers may help listeners process disfluent speech (Hypothesis 2), listeners responded consistently fastest to target words preceded by a mid-word interruption with a filler. Strikingly, these speeded responses incurred no error penalty; mid-word interruptions with fillers consistently led to fewer errors than did disfluencies without fillers, and in fact error rates were equal to or nearly as low as for Fluent items. But Experiments 3 and 4 showed quite clearly that it was not the phonological form of the filler that was driving the faster responses and lower error rates, but the extra time that elapsed during the filler before the repair. When the time between the interrupted word and repair was held constant, whether or not there was a filler in this interval made no difference in response times or error rates. And when the interrupted word was removed and the time interval before the repair was again held constant, it again made little difference whether a filler preceded the repair. It appears, then, that with a longer editing interval (allowing more time to process the evidence that there is some trouble), listeners are better able to process the repair and select the correct target word more quickly.

As we mentioned before, disfluencies can be detected quite early-by the first word in a repair (Lickley & Bard, 1992, 1998). But what is it that serves as a cue to a listener that there is a problem with an utterance? Two pieces of evidence suggest that our listeners did not conclude that a word was necessarily being cancelled when an utterance was interrupted: First, even when the referent array was constrained to two objects, the time course of listeners' responses indicates that they waited to hear at least part of the repair after the interruption and edit interval. Second, responses to the utterances with repeated color words (e.g., *Move to the yel- yellow square*; Experiment 4) indicate that an interruption by itself did not signal the cancellation of the color word in the reparandum. If it had, then listeners should have made many errors after repetitions, and they did not. The fact that replaced color words received contrastive stress and repeated color words did not suggests that the cue that a word was being cancelled may consist of a three-part combination: the presence of the interruption,[3] the semantic content of the material after the interruption (differing from that in the reparandum), and the stress characteristics (with *F*0 either elevated or not) of the color word in the repair.

So we suggest an account that goes as follows: the parser monitors for an interruption in fluent speech, which signals that the speaker is having some difficulty. When the edit interval is

[3] While we have not manipulated the presence of phonological cues to interruptions here, others have considered them; for instance, Nakatani and Hirschberg (1994) found that 30% of interruption sites in their corpus were characterized by what they call interruption glottalization (p. 1608).

long enough, this difficulty is confirmed. When the following repair word is pronounced with the same stress as the reparandum word, a repetition is signaled; on the other hand, when it has greater stress, this helps the listener suppress the material the speaker means to cancel. When the interruption comes early enough, such as within the troublesome word, this helps prevent the listener from committing him- or herself to the wrong interpretation (as shown by the lower error rates after mid-word interruptions in all four experiments). This could happen for two reasons: either the interrupted word is a relatively poor cue for activating the (unintended) color word or the fact that this word is interrupted may help suppress any potential interference that it may cause (leading to fewer incorrect interpretations after interrupted words than after completed words). The first possibility seems somewhat unlikely; in our task, it should take very little information to activate (highly predictable) color words. The second possibility seems more likely, particularly with the help of contrastive stress on the repair word.

Our experiments raise as many questions as they answer; here are three. First, precisely what is it that announces an interruption in fluent speech? The current studies say more about what the cues are not than about what they are. What we know is that neither mid-word interruptions nor the phonological forms of fillers act as cues by themselves. The length of the editing interval (which is longer when there is a filler) plays a significant role, as does the amount of misleading information that precedes the interruption site; contrastive vs parallel stress is a promising candidate as well. Future work should focus on the role of stress in repetitions and replacements (perhaps including edited examples with appropriate and inappropriate *F*0s).

Second, by what mechanism does the information available in disfluencies aid the parsing process? As Fox Tree suggested, it may be that "listeners do not automatically enter a repair mode when they hear an incongruity. If the incongruity consists of words identical to something just heard, the repair process is inactive. The process may only begin when the incongruity consists of different words" (Fox Tree, 1995, p. 730). This idea is consistent with Lickley and Bard's (1996) finding that word recognition is hindered for words in the reparandum before a fresh start, whereas word recognition is not hindered by repetition.

We propose that listeners continuously monitor spontaneous speech for cues that the speaker is having trouble. After all, speakers monitor their own internal speech prior to articulation, which enables them to make covert repairs; they can also monitor what they say after they say it (Levelt, 1983, 1989) at the same point at which listeners have access to the utterance. Levelt (1983, p. 50) proposed for speakers that "if some mismatch is detected which surpasses certain criteria, the monitor makes the speaker aware of this, or in other words: An alarm signal is sent to working memory. The speaker can then take action on the information received." Similarly, listeners may use a continuous monitoring process to detect problems in the speech they hear. Unlike speakers, they don't have access to the intention behind the utterance; however, they do have access to the surface form and the semantic-pragmatic context. They may be able to use paralinguistic cues such as the ones we have considered as well as detectable inconsistencies with the structure of the ongoing parse to make repairs incrementally. As our data show, the cues in a disfluency can help the listener compensate for its potentially harmful effects (such as any interference, misinterpretation, or delay caused by having heard the unintended information in a reparandum).

A third question is this: What are the effects of other types of disfluencies on comprehension? For this set of studies, we have limited ourselves to looking at the informativeness of mid-word interruptions with and without fillers because we began with Levelt's suggestions and because we were able to collect sufficient tokens of these types (we wanted to avoid creating disfluent utterances by editing). We further limited ourselves to tokens in which the reparandum ended during or immediately after the color word in order to gain as much control as possible. However, speakers often back up further for their repairs, retracing material before the problem word, as in *Move to the purp- to the*

orange square. The current findings would lead us to predict that retracing part of a grammatical phrase before a problem word would establish at least as much of a disfluency advantage as the extra time that elapses during a filler.

Our finding of a disfluency advantage does *not* suggest that it is better for speakers to be disfluent than fluent. In absolute terms, comprehending Naturally Disfluent utterances in their entirety would not be faster than comprehending Fluent utterances-recall that the response time advantages were relative only to the onset of the target color word. Since this word always occurred later in Naturally Disfluent utterances than in Fluent utterances, and because Fluent utterances had lower error rates overall, fluency is still desirable from a listener's perspective. But certain features of disfluencies do appear to compensate for mishaps in speaking. That is, the earlier the speaker interrupts a reparandum, the better for the listener. And not only is pausing a bit before a repair (in the editing interval) not harmful, but it buys time for the listener to cancel the unintended part of the message. Of course, these findings are not meant to be desiderata for speakers, for we have no evidence that speakers make such choices deliberately.

We began with the observation that, although spontaneous speech contains many disfluencies, most studies of the comprehension of spoken language ignore this fact and use only idealized, constructed, or read utterances. We view the current experiments, along with those of Fox Tree (1993, 1995) and Lickley and Bard (1992, 1998), as early steps in what is largely an uncharted territory-the comprehension of spontaneous speech. Our experiments offer a methodological contribution which, we believe, has the dual benefits of increasing both control and validity for studies of spoken language processing. In addition to basing stimuli on spontaneous (not read or performed) utterances, the approach we have taken combines two additional elements: (1) digital editing to create two kinds of controls that vary disfluency within utterances (in addition to nonedited controls) and (2) a comprehension task that, while logically constrained, approximates something people do in ordinary language use: understanding references to objects.

In conclusion, listeners make fewer commitments to an unintended word in a reparandum when the utterance is interrupted earlier, when there is more time to cancel material in the reparandum, and when the repair word's stress contrasts with the reparandum's. That the information available in disfluencies is useful is consistent with an incremental interpretation process that continually monitors, recognizes, and compensates for flaws in the delivery of spontaneous spoken utterances.

REFERENCES

Bear, J., Dowding, J., & Shriberg, E. (1992). Integrating multiple knowledge sources for detection and correction of repairs in human-computer dialog. In *Proceedings, 30th Annual Meeting of the Association for Computational Linguistics*, Newark, DE (pp. 56–63).

Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, **39,** 173–194.

Bond, Z. S., & Small, L. H. (1984). Detecting and correcting mispronunciations: A note on methodology. *Journal of Phonetics*, **12,** 279–283.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (in press). Disfluency rates in spontaneous speech: Effects of age, relationship, topic, role, and gender. *Language and Speech*.

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

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

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge Univ. Press.

Cooper, W. E., Tye-Murray, N., & Nelson, L. J. (1987). Detection of missing words in spoken text. *Journal of Psycholinguistic Research*, **16,** 233–240.

Core, M. G., & Schubert, L. K. (1999). A model of speech repairs and other disruptions. In *Proceedings, AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems*, American Association for Artificial Intelligence, North Falmouth, MA (pp. 48–53).

Cutler, A. (1983). Speakers' conceptions of the functions of prosody. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 79–91). Berlin: Springer-Verlag.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, **93,** 283–321.

Ferber, R. (1991). Slip of the tongue or slip of the ear? On the perception and transcription of naturalistic slips of the tongue. *Journal of Psycholinguistic Research*, **20,** 105–122.

Fox Tree, J. E. (1993). *Comprehension after speech disfluencies*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

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

Fromkin, V. A. (Ed.) (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 133–177). New York: Academic Press.

Hindle, D. (1983). Deterministic parsing of syntactic nonfluencies. In *Proceedings, 21st Annual Meeting of the Association for Computational Linguistics*, Cambridge, MA (pp. 123–128).

Hockett, C. F. (1967). Where the tongue slips there slip I. In *To honor Roman Jakobson* (Vol. 2, pp. 910–936). The Hague: Mouton.

Howell, P., & Young, K. (1991). The use of prosody in highlighting alterations in repairs from unrestricted speech. *Quarterly Journal of Experimental Psychology*, **43A,** 733–758.

Laver, J. D. M. (1973). The detection and correction of slips of the tongue. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 132–143). The Hague: Mouton.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, **14,** 41–104.

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

Levelt, W. J. M., & Cutler, A. (1983). Prosodic marking in speech repair. *Journal of Semantics*, **2,** 205–217.

Lickley, R. (1996). Juncture cues to disfluency. In *Proceedings, International Conference on Spoken Language Processing (ICSLIP '96)*, Philadelphia (pp. 2478–2481).

Lickley, R., & Bard, E. (1992). Processing disfluent speech: Recognising disfluency before lexical access. In *Proceedings, International Conference on Spoken Language Processing (ICSLIP '92)*, Banff (pp. 935–938).

Lickley, R., & Bard, E. (1996). On not recognizing disfluencies in dialog. In *Proceedings, International Conference on Spoken Language Processing (ICSLIP '96)*, Philadelphia (pp. 1876–1879).

Lickley, R., & Bard, E. (1998). When can listeners detect disfluency in spontaneous speech? *Language and Speech*, **41,** 203–226.

Marslen-Wilson, W., & Tyler, L. (1981). Central processes in speech understanding. *Philosophical Transactions of the Royal Society London*, **P259,** 317–332.

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

Nakatani, C. H., & Hirschberg, J. (1994). A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustical Society of America*, **95,** 1603–1616.

Nooteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonologicial and lexical errors in spontaneous speech. In V. Fromkin (Ed.), *Errors in linguistic performance* (pp. 87–95). New York: Academic Press.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, **5,** 42–46.

Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language*, **9,** 19–35.

O'Shaughnessy, D. (1992). Analysis of false starts in spontaneous speech. In *Proceedings, International Conference on Spoken Language Processing (ICSLIP)*, Banff (pp. 931–934).

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Ed), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, **60,** 362–367.

Shriberg, E. E. (1999). Phonetic consequences of speech disfluency. *Proceedings, International Congress of Phonetic Sciences* (San Francisco), 619–622.

Shriberg, E., Bear, J., & Dowding, J. (1992). Automatic detection and correction of repairs in human-computer dialog. In M. Marcus (Ed.), *Fifth DARPA Speech and Natural Language Workshop* (pp. 419–424). San Mateo, CA: Morgan Kaufmann.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, **32,** 25–38.

Tent, J., & Clark, J. E. (1980). An experimental investigation into the perception of slips of the tongue. *Journal of Phonetics*, **8,** 317–325.

van Wijk, C., & Kempen, G. (1987). A dual system for producing self-repairs in spontaneous speech: Evidence from experimentally elicited corrections. *Cognitive Psychology*, **19,** 403–440.

(Received March 24, 2000)
(Revision received July 5, 2000)