

Running Head: INTERPRETING PAUSES AND UMS AT TURN EXCHANGES

To cite, please refer to: Fox Tree, J. E. (2002). Interpreting pauses and ums at turn exchanges. *Discourse Processes*, 34(1), 37-55.

Interpreting Pauses and Ums At Turn Exchanges

Jean E. Fox Tree

University of California, Santa Cruz

Abstract

In three experiments, I compared how overhearers interpreted second speakers' contributions to a conversation depending on whether the second speaker responded to a first speaker immediately, paused and responded, said um and responded, or said um, paused, and then responded. The conversational snippets tested were unscripted and diverse; an example of one exchange is "Are you here because of affirmative action? (pause or um or both) It helped me out a little bit." Overhearers thought speakers had more production difficulty, were less honest, and were less comfortable with topics under discussion when speakers either said um or paused, and even more so with both. The best explanation for the data is that overhearers are judging, for each question asked, what it means for speakers to produce an anticipated or an unanticipated delay.

Interpretations of Pauses and Ums At Turn Exchanges

Pauses, ums, and uhs are popularly understood as meaningless or as hindrances to good communication. One common conception of pauses, ums, and uhs is that they are all versions of the same thing. The traditional label for ums and uhs in the research literature, filled pauses, emphasizes this seeming interchangeability. In this view, whether a person uses an um or an uh or a pause would be a matter of personal style or habit, like speaking quickly or speaking softly. But researchers have shown that they do have meaning; when estimating respondents' knowledge of the answers to questions, overhearers interpreted ums and uhs as amplifying what pauses did (Brennan & Williams, 1995). For example, overhearers judged an immediate "I don't know" to "What is the answer to Question 4?" as reflecting more commitment to not knowing the answer than "[pause] I don't know," which in turn reflected more commitment than "um I don't know." Similarly, overhearers thought an immediate answer, such as "Ottawa," reflected more commitment to the answer than "[pause] Ottawa," which reflected more commitment than "um Ottawa." So, it mattered whether overhearers heard a pause or an um or an uh, although ums and uhs had similar effects (Brennan & Williams, 1995). In other work, researchers have suggested further that ums and uhs are used to signal upcoming delays, with ums signalling major delays and uhs minor (Clark & Fox Tree, submitted; Clark & Wasow, 1998; Fox Tree, 2001; Smith & Clark, 1993).

There has been some experimental evidence to support the delay-signalling view. Smith and Clark (1993) found that in answering factual questions like "What is the capital of Canada?," people reliably used um before a long delay and uh before a short delay in the amount of time it took for them to answer. Fox Tree (2001) found that people recognized words in a speech stream more quickly after uh than when the uh was digitally excised. In the word monitoring task used to test processing, people saw a word on a computer screen, and then listened to an utterance. If they heard the word they saw, they pressed a button as quickly as possible. Their pressing speed, which is related to

their ability to integrate information (Fox Tree, 1995; Fox Tree & Schrock, 1999; Marslen-Wilson & Tyler, 1980), was faster after an uh was heard; it was not affected by hearing um. These findings replicated cross-linguistically, in English and Dutch (Fox Tree, 2001). A possible explanation is that as indicators of minor delay, paying attention after uh may be worth the listener's while. But as indicators of major delay, paying attention after um may not pay off, because the listener cannot gauge how long it will be before the speaker will resume.

In the research described here, I tested the hypothesis that delays that are signalled by um differ from delays that are not signalled. I tested overhearers' interpretations of speakers' speech production difficulty, honesty, and comfort with topic because many researchers have identified links between these variables and um use (Brotherton, 1979; Christenfeld & Creager, 1996; Fox Tree & Clark, 1997; Hosman & Wright, 1987; Jefferson, 1974; Kasl & Mahl, 1987; Lalljee & Cook, 1969, 1973; Martin, 1967; Reynolds & Paivio, 1968; Siegman, 1979; Schachter, Christenfeld, Ravina, & Bilous, 1991; Tannenbaum, Williams, & Hillier, 1965; Vrij & Winkel, 1991). I selected um to study because (1) prior research has suggested that um and uh be treated as distinct elements, and (2) prior research has found effects for uh but not for um, inviting testing with alternative tasks to see if um effects can be found (Fox Tree, 2001). I investigated the role of ums at turn exchanges because turn exchanges seem to be a salient part of speakers' utterances. Ums are also more likely to occur at the beginning of intonation units as compared to positions later in intonation units (Clark & Fox Tree, submitted; an intonation unit is a segment of talk produced as a prosodic whole; the beginning of a turn is always also the beginning of an intonation unit).

Attributions Tested

In the current research, I explore the attributions made by overhearers about speakers depending on the type of turn exchange. I measured the following types of attributions:

Speaker's speech production difficulty. Forewarning delays may help listeners anticipate what will come next, or, more broadly, cue listeners to be attentive because the upcoming talk may be difficult to process. If ums are linked to increased speech production difficulty, and if overhearers are sensitive to this, then overhearers should rate speakers as having more production difficulty with ums than without. This was measured with the question "How difficult do you think it was for the respondent to put their response into words?" (1=not difficult, 7= very difficult).

Speaker's honesty or evasiveness. One study found that ums and uhs, in combination with pauses, made courtroom defendants appear more guilty than they would have appeared without them (Hosman & Wright, 1987). The researchers did not tease apart the contributions of ums or uhs versus pauses. Another study found that people who were instructed to deceive police officers in a mock interrogation used more ums and uhs than those telling the truth (Vrij & Winkel, 1991). If ums are linked to increased evasiveness, and if overhearers are sensitive to this, then overhearers should rate speakers as more evasive with ums than without. I tested overhearer's interpretations of speaker's honesty with "How close do you think the respondent's answer is to what the respondent truly thinks?" (1=not close (deceptive), 7= very close (honest)).

Speaker's comfort with the topic. Research findings have generally not supported a correlation between ums or uhs and anxiety (Mahl, 1987), and at least one controlled study has shown that ums or uhs can actually make speakers sound more relaxed (Christenfeld, 1995). Nonetheless, laypeople tend to believe that speakers who use ums or uhs are "inarticulate, uninteresting, ill-prepared, [and] nervous" among other negative things (Christenfeld, 1995 p. 173). The controlled study is worth replicating. My studies differ from the earlier studies in that I test more stimuli, and I test the effects of an individual um on judgements, instead of a group of ums and uhs taken as a whole. If ums are linked to increased discomfort at turn exchanges, and if overhearers are sensitive to this, then having ums should make overhearers interpret speakers as being less

comfortable with the topic. I measured overhearer's interpretations of speaker's comfort with the question "How comfortable do you think the respondent feels with the topic of discussion?" (1= not comfortable, 7=very comfortable). I chose to investigate comfort with topic because it was more relevant to the task at hand -- judgements of brief turn exchanges -- than the more trait-directed query about a speaker's anxiety level.

The links researchers identified between ums or uhs and speech production difficulty, honesty, and comfort with topic reflect the deeper link between ums or uhs and delays in speaking (Clark & Fox Tree, submitted). For example, pauses increase with greater upcoming syntactic complexity (Clark & Wasow, 1998; Ferreira, 1991) and speech production trouble (Goldman-Eisler, 1968; Levelt, 1989, Maclay & Osgood, 1959). People also take longer to start talking when telling a sensitive story than when telling a less personal story (Horowitz, Weckler, Saxon, Livaudais, and Boutacoff, 1977). In all these cases, an increased need for planning time, (because of production difficulty, figuring out how to talk about an uncomfortable topic, or some other reason) will lead to delays, and speakers will recognize those impending delays and mark them with ums or uhs (Clark & Fox Tree, submitted). In this way, ums may be linked to speech production trouble, honesty, and comfort with topic even though their underlying function is to indicate upcoming delays in speech.

Turn Intervals Tested

To hold as many spontaneous speech variables constant as possible, the same snippets of conversation were edited to create different types of turn exchanges, with different overhearers hearing different versions of each snippet.

An approximately 1 s pause is the amount of silence comfortably sustained before a conversational participant tries to fill the gap, either by the current speaker continuing or by the next speaker beginning (Jefferson, 1989). For example, when searching for a word, speakers make an attempt to continue after about 1 s of hiatus, as in "Harry uh [1 s pause] Schirmer?" (adapted from Jefferson, 1989, p. 191). As if to highlight the normality

of this base metric of 1 s, delays several seconds long are often accompanied by expletives (Jefferson, 1989, p. 183, p. 191). Very long pauses, such as those over 6 s, can often be attributed to nonverbal activities interlocutors are engaged in at the time that they are talking (Jefferson, 1989). Pauses over 3 s long, without ongoing activities to attribute them to, make interlocutors uncomfortable (McLaughlin & Cody, 1982), and people probably try to avoid silences of this length; in one study of conversations, almost all pauses were under 2 s in length, and none were over 3 s (Goldman-Eisler, 1968).

In Experiment 1, I compared the three turn intervals of a 1 s pause, a 3 s pause, and a .5 s pause plus um plus 1 s pause (the um condition). The um was preceded by a .5 s pause to make the turn-exchange sound natural, as an um coincident with the offset of the last speaker's turn sounded premature. The 1 s versus 3 s pause conditions should replicate earlier research that uncomfortably long inter-turn pauses yield negative attributions (McLaughlin & Cody, 1982).

The outcome of the um condition is less predictable. If ums are filtered out and ignored (the filter hypothesis), we might expect attributions for the um condition to be more similar to the 1 s condition than the 3 s condition (ignoring um, there is a .5 s difference between the um and 1 s conditions, compared to 1.5 s difference between the um and 3 s conditions). If ums and pauses contribute to attributions based on their lengths in time (the duration hypothesis), we might expect attributions for the um conditions to be halfway between attributions for the 1 s and 3 s conditions (ums average .5 s long; so the durations of the inter-turn intervals would be 1 s, ~2 s, 3 s). Ums may also affect interpretations independently from pauses (the separate contributions hypothesis). Because there are a variety of ways of thinking about separate contributions, I will discuss this hypothesis in the next section.

In Experiment 2, I compared the three turn intervals of a .5 s pause, a 3 s pause, and a .5 s pause plus um plus 3 s pause (the um 3 s condition). This test once again provides end-points for measuring the effects of an um on the interpretations of pauses:

the .5 s versus 3 s pause conditions should replicate earlier research that uncomfortably long inter-turn pauses yield negative attributions (the 1 s inter-turn pause was reduced to .5 s to more closely match the pause preceding the um in the um 3 s condition and to assure the perception of a smooth turn exchange in the .5 s condition). Unlike in Experiment 1, in Experiment 2 the um precedes an usually long pause.

Like in Experiment 1, the outcome of the um 3 s condition is less predictable. According to the filter hypothesis, we might expect attributions for the um 3 s condition to be more similar to the 3 s condition than the .5 s condition (ignoring um, there is a 3 s difference between the um 3 s and the .5 s conditions, compared to .5 s difference between the um 3 s and the 3 s conditions). According to the duration hypothesis, we might expect attributions for the um 3 s condition to exceed attributions for .5 s and 3 s (durations of the inter-turn intervals would be .5 s, 3 s, ~4 s). Possible outcomes predicted by the separate contributions hypothesis will be discussed in the next section.

Experiment 3 controlled for the time elapsed between the main messages of the first and second speakers, to directly test the duration hypothesis. In Experiment 3, I compared the turn intervals of a .5 s pause, a .5 s pause plus um plus 3 s pause, and a long pause equivalent in length to the .5 s pause plus um plus 3 s pause. That is, in Experiments 1 and 2, the time elapsed in the long pause condition is 3 s, but in Experiment 3, the time elapsed in the long pause condition is about 4 s, with the time varying depending on the length of the um for each stimulus. This ~4 s equals the amount of time between the first and second speakers in the um plus 3 s pause conditions of Experiments 2 and 3. According to the filter hypothesis, we might expect attributions for the um 3 s condition to be closer to the ~4 s condition than the .5 s condition (ignoring um, there is a ~.5 s difference between the um 3 s and ~4 s conditions, compared to a 3 s difference between the um 3 s and .5 s conditions). According to the duration hypothesis, we might expect attributions for the um 3 s condition to equal attributions for the ~4 s condition (durations of the inter-turn intervals would be .5 s, ~4 s, ~4 s). Once again,

possible outcomes predicted by the separate contributions hypothesis will be discussed in the next section.

The Separate Contributions Hypothesis

There are at least three subhypotheses of the separate contributions hypothesis, and they predict different things. In the expectation subhypothesis, listeners expect ums to forewarn major delays. According to this subhypothesis, (a) ratings for the um condition will be similar to the 3 s condition in Experiment 1, (b) ratings for the um 3 s condition will be similar to the 3 s condition in Experiment 2, and (c) ratings for the um 3 s condition will be similar to the ~4 s condition in Experiment 3. In each of these cases, hearing an um (and expecting a major delay) and experiencing a major delay (in the form of a 3 s or ~4 s pause) would lead to the same changes in attributions as compared to a short inter-turn interval (.5 s or 1 s). It wouldn't matter whether the ums were actually followed by long silent pauses.

In the responsibility subhypothesis, ums play a role in assigning responsibility for long inter-turn pauses. With a 3 s pause alone, part of the responsibility for the long inter-turn interval could be carried by the first speaker. Long pauses that follow ums, however, can be unambiguously attached to the um-producer. That is, the second speaker says um, gains the floor, and therefore takes the responsibility for a pause following um. Ums can be used to give up the floor as well (see Clark & Fox Tree, submitted), but in this case, their positions in the utterances would likely be different; for example, they might be turn-final, which was never the case for the current materials. According to this subhypothesis, (a) ratings for the um condition will be similar to the 1 s condition in Experiment 1, (b) ratings for the um 3 s condition will be more negative than the 3 s condition in Experiment 2, and (c) ratings for the um 3 s condition will be more negative than the ~4 s condition in Experiment 3.

In the interpreting-in-context subhypothesis, the attributions made when hearing um will be related to what it means to indicate advance knowledge of an upcoming delay

for that attribution with this task. I say “with this task” because having a pause means different things in different situations. When driving, a pause can mean “I am paying attention to the road now” (Fox Tree, 2000; Jefferson, 1989), but when talking face-to-face it could mean “I am having trouble saying what I want to say.” With respect to attributions of speech production difficulty, overhearers may reason that to have advance knowledge of a delay, speakers must have had a greater awareness of their production difficulty, and therefore be having a more serious problem than a silent pause alone would indicate. Similar reasoning may apply for attributions of honesty and comfort with topic. Long silent pauses may suggest speakers are planning what to say, perhaps in the service of deception or figuring out how to talk about an uncomfortable topic. Awareness of the upcoming silence may indicate even more serious planning problems. According to this subhypothesis, (a) ratings for the um condition will be more negative than the 1 s condition in Experiment 1, (b) ratings for the um 3 s condition will be more negative than the .5 s condition and the 3 s condition in Experiment 2, and (c) ratings for the um 3 s condition will be more negative than the .5 s condition and possibly the ~4 s condition in Experiment 3. This subhypothesis does not make any particular predictions about the relative lengths of long silent pauses on attributions, or the relative contribution of ums versus long silent pauses to attributions.

Experiment 1

Method

Participants. Forty-two UCSC students participated in exchange for course credit. All were native English speakers. No student participated in the other experiments reported here.

Materials. Stimuli came from recorded conversations of 11 pairs of UC Santa Cruz students who had a conversation in exchange for course credit. Some pairs were acquainted, others were not. Speakers were instructed to engage in a conversation on any topic for about half an hour. If they needed help finding something to discuss, they could

use a topic list that a research assistant had prepared in advance. These topics were considered to be of interest to UC Santa Cruz students, and included questions such as Are you for or against the ban on assault weapons?, What crime, if any, warrants the death penalty?, and Do you believe in aliens? ghosts? Elvis? Speakers knew that they were being recorded and consented to allow their speech to be used in future experiments.

Thirty turn exchanges from these spontaneous conversations were investigated (see Appendix). In each segment, one speaker began talking, followed by a turn exchange in which a second speaker responded, with one exception: in addition to the response, one second speaker also produced the backchannel uh huh before the end of the first speaker's turn. Most exchanges were question-answer pairs. The three that were not included two statements of opinion followed by reactions, and one description of an event followed by a reaction. Twenty-three of the questions could be interpreted as yes-no questions, although no responses were in fact only yes or no. Each segment of conversation was edited to create three versions, as in the following example:

- (1) A: Are you here because of affirmative action?
B: (1 s pause) It helped me out a little bit.
- (2) A: Are you here because of affirmative action?
B: (.5 s pause + um + 1 s pause) It helped me out a little bit.
- (3) A: Are you here because of affirmative action?
B: (3 s pause) It helped me out a little bit.

Any nonpropositional information between the first and second speakers' propositional statements, such as nonword vocalizations, were removed. Pauses were created by copying background noise from elsewhere in the recording, and ums that were not spontaneously produced at the beginning of the turn were digitally spliced in. In stimuli with spontaneous ums, pauses surrounding the ums were edited. This means that editing took place in all conditions, which distinguishes the experiments described in this paper from other experiments using the

editing technique to test the role of ums (Fox Tree, 2001). In the um condition, a .5 s pause was added before the um to make the stimuli sound more natural. Ums averaged 505 ms long (s.d. 112), with a range from 309 ms to 803 ms. The editing procedures yielded 90 stimuli.

In addition to the critical stimuli, there were also ten lure stimuli. Lure stimuli also consisted of one speaker followed by a second speaker, but the turn exchanges did not contain ums or pauses. Instead, turn exchanges contained discourse markers such as like, I mean, I guess, well, and oh, overlaps, or false starts. The purpose of lure stimuli was to mask the manipulation. No participant in the three experiments detected the manipulation, although this also might have occurred without the lures.

To prevent the overhearers from focussing on the turn exchange, they were told that the experiment was about how people get acquainted with one another, and that they would hear recorded samples of conversations followed by questions about the interactions. The questionnaire was numbered and each exchange was labelled by a topic phrase, such as Affirmative Action for the above example. After hearing each trial, participants responded to five questions using a 7-point Likert scale, of which three questions and their ratings will be discussed here (the other two questions supported the cover story; they were (a) How well do you think these two people know each other? and (b) How likely do you think it is that the respondent will seek out further contact with the other person?). The questions of interest were (1) How difficult do you think it was for the respondent to put their response into words? (1=not difficult, 7=very difficult), (2) How close do you think the respondent's answer is to what the respondent truly thinks? (1=not close (deceptive), 7=very close (honest)), and (3) How comfortable do you think the respondent feels with the topic of discussion? (1=not comfortable, 7=very comfortable). Questions were always presented in the same order. This held constant any potential order effects across conditions.

Design. Three counterbalanced lists were created for the 90 critical stimuli. Each list contained 10 stimuli with a short pause between turns, 10 stimuli with a long pause, and 10 stimuli with an um such that a particular version of a stimulus only occurred once per list. The 10

lures were constant across lists, yielding 40 stimuli per list. Stimuli conditions were presented in a random order. The item order was constant across lists.

Procedure. Participants were randomly assigned to one of the three lists. After reading instructions, participants sat in front of a tape player with a questionnaire before them, put on headphones, and pressed play. They first heard a practice trial, after which they stopped the tape and consulted with the experimenter to make sure they understood the instructions. Thereafter, the experiment was self-paced. After each trial, a voice on the tape said “Please stop the tape now and answer the five questions.” When they had finished answering the questions, participants were instructed to press play to hear the next trial. Participants were told that it was important to stop the tape in order to answer the questions, and to keep in mind that the questions always referred to the second speaker of the dyad. Participants were not allowed to rewind the tape.

Results

Results will be discussed by question asked. Means and standard deviations, averaged across subjects and items, are presented in Table 1.

Insert Table 1 About Here

(1) How difficult do you think it was for the respondent to put their response into words? Both ums and 3 s pauses made a respondent appear to have speech production difficulty, although 3 s pauses suggested more difficulty than ums (contrast 1 s to um 1 s: $F_1(1,41) = 27.64$, $p < .001$; $F_2(1,29) = 36.37$, $p < .001$; contrast 1 s to 3 s: $F_1(1,41) = 72.15$, $p < .001$; $F_2(1,29) = 90.64$, $p < .001$; contrast um 1 s to 3 s: $F_1(1,41) = 12.56$, $p < .001$; $F_2(1,29) = 13.83$, $p < .001$).

(2) How close do you think the respondent’s answer is to what the respondent truly thinks? Either an um or a 3 s pause made respondents appear less honest, and to the same extent (contrast 1 s to um 1 s: $F_1(1,41) = 21.06$, $p < .001$; $F_2(1,29) = 36.48$, $p < .001$; contrast 1 s to 3 s: $F_1(1,41) = 10.09$, $p < .01$; $F_2(1,29) = 17.44$, $p < .001$; contrast um 1 s to 3 s: $F_1(1,41) = .45$, $p =$

.50; $F_2(1,29) = .82, p = .37$).

(3) How comfortable do you think the respondent feels with the topic of discussion?

Either an um or a 3 s pause made respondents appear less comfortable with the topic, and to the same extent (contrast 1 s to um 1 s: $F_1(1,41) = 15.68, p < .001$; $F_2(1,29) = 30.11, p < .001$; contrast 1 s to 3 s: $F_1(1,41) = 37.2, p < .001$; $F_2(1,29) = 40.86, p < .001$; contrast um 1 s to 3 s: $F_1(1,41) = 2.89, p = .10$; $F_2(1,29) = 3.02, p = .10$).

Summary

As expected, the 1 s interval between turns was always rated more positively than the 3 s interval. With a shorter interval, overhearers rated second speakers as having less production difficulty, being more honest, and being more comfortable with the topic discussed. The um plus 1 s interval always fell within the boundaries of the other two intervals. It was never more positive than the 1 s interval, nor more negative than the 3 s interval. With respect to judgments of honesty and comfort with the topic, it was similar to the 3 s interval. With respect to judgements of speech production difficulty, um plus 1 s fell between the other two, similar to neither.

Experiment 2

Method

Participants. Thirty UCSC students participated in exchange for course credit. All were native English speakers. No student participated in the other experiments reported here.

Materials. The materials were the same as in Experiment 1 except that the 1 s turn exchanges were reduced to .5 s, and that the 1 s pause after um was increased to 3 s.

Design and Procedure. The design and procedure were the same as in Experiment 1.

Results

As before, results will be discussed by question asked. Means and standard deviations, averaged across subjects and items, are presented in Table 2.

Insert Table 2 About Here

(1) How difficult do you think it was for the respondent to put their response into words? As in Experiment 1, the short um-free turn interval was associated with the least difficulty in speech production. But whereas in Experiment 1 an um 1 s pause entailed less production difficulty than a 3 s pause, in Experiment 2 an um 3 s pause entailed more production difficulty than a 3 s pause (contrast .5 s to 3 s: $F_1(1,29) = 57.02, p < .001$; $F_2(1,29) = 65.01, p < .001$; contrast .5 s to um 3 s: $F_1(1,29) = 113.24, p < .001$; $F_2(1,29) = 255.98, p < .001$; contrast um 3 s to 3 s: $F_1(1,29) = 42.25, p < .001$; $F_2(1,29) = 29.99, p < .001$).

(2) How close do you think the respondent's answer is to what the respondent truly thinks? As in Experiment 1, the short um-free turn interval was considered the most honest. But whereas in Experiment 1, an um 1 s pause entailed the same honesty level as a 3 s pause, in Experiment 2, an um 3 s pause entailed less honesty than a 3 s pause (contrast .5 s to 3 s: $F_1(1,29) = 22.09, p < .001$; $F_2(1,29) = 15.86, p < .001$; contrast .5 s to um 3 s: $F_1(1,29) = 49.6, p < .001$; $F_2(1,29) = 36.37, p < .001$; contrast um 3 s to 3 s: $F_1(1,29) = 7.07, p = .01$; $F_2(1,29) = 8.81, p < .01$).

(3) How comfortable do you think the respondent feels with the topic of discussion? As in Experiment 1, the short um-free turn interval was associated with the most comfort. But whereas in Experiment 1, an um 1 s pause entailed the same comfort as a 3 s pause, in Experiment 2, an um 3 s pause entailed less comfort than a 3 s pause (contrast .5 s to um 3 s: $F_1(1,29) = 49.62, p < .001$; $F_2(1,29) = 99.84, p < .001$; contrast .5 s to 3 s: $F_1(1,29) = 46.2, p < .001$; $F_2(1,29) = 74.79, p < .001$; contrast um 3 s to 3 s: $F_1(1,29) = 8.37, p < .01$; $F_2(1,29) = 12.34, p = .001$).

Summary

As with Experiment 1, the .5 s interval between turns always yielded the most positive judgements. Unlike Experiment 1, the um interval did not fall within the boundaries of the other two intervals for any question. Instead, when followed by a 3 s pause, the um interval made all interpretations more negative than a 3 s pause alone. With either an um or a 3 s pause,

interlocutors were thought to have more speech production difficulty, to be less honest, and to be more uncomfortable with the topic, and even more so with both.

Experiment 3

Method

Participants. Forty-two UCSC students participated in exchange for course credit. All were native English speakers. No student participated in the other experiments reported here.

Materials. The materials were the same as in Experiment 2 except that the 3 s turn exchanges were increased in length. Each stimulus was increased by exactly the amount of time that the um took up in the um version of each stimulus. So, the inter-turn intervals varied across stimuli, averaging about 4 s.

Design and Procedure. The design and procedure were the same as in Experiment 1.

Results

As before, results will be discussed by question asked. Means and standard deviations, averaged across subjects and items, are presented in Table 3.

Insert Table 3 About Here

(1) How difficult do you think it was for the respondent to put their response into words? As in Experiment 2, overhearers thought respondents had least speech production difficulty with .5 s turn exchanges, more with ~4 s turn exchanges, and most with um 3 s turn exchanges (contrast .5 s to ~4 s: $F_1(1,41) = 18.85$, $p < .001$; $F_2(1,29) = 135.95$, $p < .001$; contrast .5 s to um 3 s: $F_1(1,41) = 84.69$, $p < .001$; $F_2(1,29) = 188.70$, $p < .001$; contrast um 3 s to ~4 s: $F_1(1,41) = 6.74$, $p = .01$; $F_2(1,29) = 15.34$, $p = .001$).

(2) How close do you think the respondent's answer is to what the respondent truly thinks? Like Experiment 2, overhearers thought respondents were most honest in the .5 s turn exchanges. But unlike Experiment 2, they were less honest in either the um 3 s exchanges or the ~4 s exchanges (contrast .5 s to ~4 s: $F_1(1,41) = 12.48$, $p = .001$; $F_2(1,29) = 22.00$, $p < .001$;

contrast .5 s to um 3 s: $F_1(1,41) = 25.40, p < .001$; $F_2(1,29) = 42.03, p < .001$; contrast um 3 s to ~4 s: $F_1(1,41) = 1.96, p = .17$; $F_2(1,29) = 1.89, p = .18$).

(3) How comfortable do you think the respondent feels with the topic of discussion? Like Experiment 2, overhearers thought respondents were most comfortable in the .5 s turn exchanges. But unlike Experiment 2, they were less comfortable in either the um 3 s exchanges or the ~4 s exchanges (contrast .5 s to ~4 s: $F_1(1,41) = 20.89, p < .001$; $F_2(1,29) = 70.20, p < .001$; contrast .5 s to um 3 s: $F_1(1,41) = 63.89, p < .001$; $F_2(1,29) = 103.20, p < .001$; contrast um 3 s to ~4 s: $F_1(1,41) = .04, p = .84$; $F_2(1,29) = .05, p = .82$).

Summary

As with Experiments 1 and 2, the .5 s interval between turns always yielded the most positive judgements. Like Experiment 2, more production trouble was attributed to speakers in the um 3 s condition than in the ~4 s condition; the length of the silence (3 s in Experiment 2 and ~4 s in Experiment 3) didn't matter. Unlike Experiment 2, the same honesty and comfort was attributed to speakers in the um 3 s condition as in the ~4 s condition; the extra ~1 s pause over Experiment 3's materials increased negative attributions as much as ums did.

General Discussion

Delays in spontaneous speech are unavoidable, and speakers have many choices about how to handle them. One is to leave a silent stretch of speech until the speaker is ready to continue. Another is to mark that upcoming silence with an um. Prior research has shown that the choice between one or the other can affect overhearers' interpretations of a respondent's feeling of knowing (Brennan & Williams, 1995). In the current research, I explored how three other types of judgements -- of speech production difficulty, honesty, and comfort with topic -- are influenced by ums and pauses. Each of these other judgements have been linked to ums by prior researchers, but little direct exploration of their effects on overhearers has been done.

In the context of these judgements, I tested three hypotheses: (1) the filter hypothesis, that ums are filtered out by overhearers and have no effect on judgements, (2) the duration hypothesis, that the effects of ums can be linked to their durations, and (3) the separate

contributions hypothesis, that there is an effect of the phonological form of um that is separate from the effects of silent pauses. I introduced three subhypotheses of the separate contributions hypothesis: (a) the expectation subhypothesis, that listeners hear an um and expect a major delay regardless of whether or not there is a delay, (b) the responsibility subhypothesis, that listeners attribute full responsibility for the negative effects of long silent pauses to the second speakers when second speakers begin their turns with ums, and (c) the interpreting-in-context subhypothesis, that listeners interpret ums and pauses according to what ums and pauses imply for the attributions they are making. For ease of comparison, Table 4 summarizes the predictions of each hypothesis and the results obtained.

There was overwhelming evidence against the filter hypothesis. The predicted results were obtained in only two out of nine cases (Experiment 3 honesty and comfort with topic questions). The Experiment 3 production difficulty results and all the results from Experiments 1 and 2 demonstrated that the presence of um mattered; ums were not filtered out and ignored.

In contrast, the duration hypothesis predicted six out of nine results. The Experiment 1 production difficulty results, all the Experiment 2 results, and the Experiment 3 honesty and comfort with topic results fit the duration hypothesis, but the Experiment 1 honesty and comfort with topic results and the Experiment 3 speech production difficulty results did not. The duration hypothesis is a clean and durable hypothesis; it goes a long way in explaining the current results as well as prior data. Brennan and Schober (2001) found that it was the time that uhs took up (ums were not studied) rather than their phonological forms that led to listeners' improved recoveries from speech errors. Nonetheless, for the current materials and method, the duration hypothesis comes up short, particularly because it fails at some point for each of the three types of ratings collected (otherwise, it might have been the case that, for example, judgements of honesty were sensitive to duration, but judgements of production difficulty were not).

Two of the separate contributions subhypotheses fared less well than the duration hypothesis. The expectation subhypothesis predicted four out of nine results. The Experiments 1 and 3 honesty and comfort with topic results fit the hypothesis, but the Experiments 1 and 3

production difficulty results and all the Experiment 2 results did not. The responsibility subhypothesis also predicted four out of nine results. The Experiment 2 results and the Experiment 3 production difficulty results fit the hypothesis, but the Experiment 1 results and the Experiment 3 honesty and comfort with topic results did not. As with the duration hypothesis, both the expectation subhypothesis and the responsibility subhypothesis fail at some point for each of the three types of ratings collected.

The remaining separate contributions hypothesis, the interpreting-in-context subhypothesis, can account for all the results, but this is in part because it makes looser predictions. In addition, although it can explain the main results, it does not explain all the detailed results.

With respect to the attributions of production difficulty, the data show that displaying advance knowledge of delays with um indicates a problem, having long delays indicates a problem, and having both indicates even more difficulty. The three experiments together demonstrate that ums and pauses contribute separately to attributions of speech production difficulty. It makes sense that overhearers would judge a delay that speakers were aware of in advance (i.e. signalled by um) as more noteworthy than a delay they were potentially not aware of in advance (i.e. long silent pauses). It also makes sense that both the anticipation of a silent pause and the silent pause itself would impact production difficulty judgements. But the data also lead to the conclusion that the actual experience of a 3 s silence implies more production difficulty than the expectation of a major delay signalled by um (Experiment 1), and there is no clear explanation for why.

With respect to attributions of honesty and comfort with topic, the data show that long silences indicate speakers' constructions of more carefully worded replies, and anticipated long silences indicate more serious struggles with wording. The careful wording could result from either preparing a dishonest answer or figuring out how to talk about an uncomfortable subject. Somewhat differently from the production difficulty question, the data lead to the conclusion that the actual experience of a 3 s silence implies about the same wording struggles as the expectation

of a major delay signalled by um (Experiment 1), and that these judgements of wording struggles appear to be sensitive to the lengths of pauses (judgements of dishonesty and discomfort were greater in the um 3 s condition than the 3 s condition of Experiment 2, but as great in the um 3 s condition as the ~4 s condition of Experiment 3). As before, however, there is no clear explanation for why.

The three studies presented here explored a number of ways of thinking about the role of ums and pauses at turn exchanges. There are still some questions. Why does a long silent pause lead to attributions of greater production difficulty than an um, but the same amount of dishonesty and discomfort as an um? Is it just chance that the ~1 s extra silent pausing leads to similar increases in negative ratings as an um for the honesty and comfort with topic questions of Experiment 3, or is there something systematic going on? For now, the interpreting-in-context subhypothesis of the separate contributions hypothesis explains the data best.

Prior research has also supported a separate contributions approach to pauses and ums. Brennan and Williams (1995) found that controlling for pause length, ums and uhs had an additional amplifying effect. The difference between Brennan and Williams (1995), Brennan and Schober (2001), and the current experiments may have to do with the different types of tasks used (rating speech versus following instructions in real time), whether ums or uhs were studied, whether the surrounding speech was disfluent or not, or the type of talk studied (tending towards scripted or tending towards natural dialogue).

Despite lingering questions, the experiments presented here demonstrate that ums and pauses are not the same thing, as the term filled pause implies. The presence or absence of an um has implications for how far in advance speakers were aware of their upcoming delay. With only silent pauses, overhearers don't know if speakers anticipated the pauses or not. Ums before long pauses show advance knowledge of delays. For judgements of speech production difficulty, honesty, and comfort with the topic, it matters whether speakers say um, or pause, or do both.

References

- Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. Journal of Memory and Language, *44*(2), 274-296.
- Brennan, S. E., & Williams, W. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. Journal of Memory and Language, *34*, 383-398.
- Brotherton, P. (1979). Speaking and not speaking: Processes for translating ideas into speech. In A. W. Siegman and S. Feldstein, Of Speech and Time. New York: Wiley.
- Christenfeld, N. (1995). Does it hurt to say um? Journal of Nonverbal Behavior, *19*(3), 171-186.
- Christenfeld, N., & Creager, B. (1996). Anxiety, alcohol, aphasia, and ums. Journal of Personality and Social Psychology, *70*(3), 451-460.
- Clark, H. H. (1994). Managing problems in speaking. Speech Communication, *15*, 243-250.
- Clark, H. H., & Fox Tree, J. E. (submitted). Using Uh and Um in Spontaneous Speaking.
- Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. Cognitive Psychology, *37*, 201-242.
- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. Journal of Memory and Language, *30*, 210-233.
- Fox Tree, J. E. (2000). Coordinating spontaneous talk. In L. R. Wheeldon (Ed.), Aspects of Language Production (pp. 375-406). Philadelphia: Psychology Press.
- Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. Memory and Cognition, *29*(2), 320-326.
- 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.
- Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. Cognition, *62*(2), 151-167.

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

Goldman-Eisler, F. (1968). Psycholinguistics: Experiments in Spontaneous Speech. London: London Academic Press.

Horowitz, L. M., Weckler, D., Saxon, A., Livaudais, J. D., & Boutacoff, L. I. (1977). Discomforting talk and speech disruptions. Journal of Consulting and Clinical Psychology, 45(6), 1036-1042.

Hosman, L. A., & Wright, J. W., II (1987). The effects of hedges and hesitations on impression formation in a simulated courtroom context. Western Journal of Speech Communication, 51, 173-188.

Jefferson, G. (1974). Error correction as an interactional resource. Language in Society, 3(2), 181-199.

Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation. In D. Roger & P. Bull (Eds.), Conversation: An Interdisciplinary Perspective (pp. 167-195). Philadelphia: Multilingual Matters, Ltd.

Kasl, S. V., & Mahl, G. F. (1987). Speech disturbances and experimentally induced anxiety. In G. F. Mahl (Ed.), Explorations in Nonverbal and Vocal Behavior (pp. 203-213). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Lalljee, M. G., & Cook, M. (1969). An experimental investigation of the function of filled pauses in speech. Language and Speech, 12(1), 24-28.

Lalljee, M., & Cook, M. (1973). Uncertainty in first encounters. Journal of Personality and Social Psychology, 26(1), 137-141.

Levelt, W. J. M. (1989). Speaking: From Intention to Articulation. Cambridge, MA: MIT Press.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English

speech. Word, 75, 19-44.

Mahl, G. F. (1987). Explorations in Nonverbal and Vocal Behavior. Hillsdale, N.J.: Erlbaum.

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

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

McLaughlin, M. L., & Cody, M. J. (1982). Awkward silences: Behavioral antecedents and consequences of the conversational lapse. Human Communication Research, 8(4), 299-316.

Reynolds, A., & Paivio, A. (1968). Cognitive and emotional determinants of speech. Canadian Journal of Psychology, 22(3), 164-175.

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. Journal of Personality and Social Psychology, 60(3), 362-367.

Shriberg, E. E., & Lickley, R. J. (1993). Intonation of clause-internal filled pauses. Phonetica, 50, 172-179.

Siegmán, A. W. (1979). Cognition and hesitation in speech. In A. W. Siegmán & S. Feldstein (Eds.), Of Speech and Time (pp. 151-178). Hillsdale, NJ: Lawrence Erlbaum.

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

Swerts, M. (1998). Filled pauses as markers of discourse structure. Journal of Pragmatics, 30, 485-496.

Tannenbaum, P. H., Williams, F., & Hillier, C. S. (1965). Word predictability in the environments of hesitations. Journal of Verbal Learning and Verbal Behavior, 4, 134-140.

Vrij, A., & Winkel, F. W. (1991). Cultural patterns in Dutch and Surinam nonverbal behavior: An analysis of simulated police/citizen encounters. Journal of Nonverbal Behavior, 15(3), 169-184.

Appendix

Transcript of Materials with Topic Phrases

1. Drugs. A: Should any other drugs be legalized? B: I don't think so.
2. Aliens. A: Do you believe in aliens? B: I believe there's life out there, I just don't believe it's visited here.
3. Death penalty. A: Do you feel, for any reason, do you think that there's a reason why um the state should put people to death? B: Anyone who kills someone should die.
4. Prison. A: Do you think that everybody in jail has the potential to be rehabilitated? B: I don't think that once they are past the age of twelve it's too late for them to change.
5. Weapons. A: Are you for or against the ban on assault weapons? B: I'm for it.
6. Santa Cruz. A: Do you like Santa Cruz? B: There's lots of different kinds of people here.
7. Animal test. A: Are you for animal testing? B: I think idealistically I'm against it, but I think it's benefin- benefitted modern science in a lot of ways.
8. Abortion. A: Are you for it or against it? B: I'm gonna say pro choice.
9. Weapons. A: Are you for or against the ban on assault weapons? B: It's so bad, because like the whole weapon thing's has got so out of control where just anybody can got- get one.
10. Drugs. A: Should alcohol be illegal? B: Alcohol can be controlled better than marijuana can, I think.
11. Rape. A: What do you think we should do to rapists? B: I think they should surgically remove their penis.
12. Aliens. A: Do you believe in aliens? B: I do to the extent that everybody has their own truth and so I'm not gonna doubt somebody and, if they say they've had that experience.
13. Religion. A: So do you believe in a divine being? B: I think that if there is a divine being that it's a Goddess, not a God.
- 14 Affirmative Action. A: Are you here because of affirmative action? B: It helped me out a little bit.

15. Abortion. A: Should the state fund abortions for the poor? B: It's gonna be a lot of money putting into them.
16. Aliens. A: I believe in ghosts and aliens. B: I think it's like just absurd.
17. Relationships. A: That must be hard, though, having to deal with a long distance relationship. Do you get to talk to him often? B: Before when he went to boot camp, I didn't talk to him at all.
18. Euthanasia. A: And it's just ma- it's just taking its toll on your life and you just say, hey I don't need to go through this, I just wanna die. Is that wrong? B: Technically no, but I think I'm a moral kind of guy.
19. Immigration. A: Do you think that the U.S. should keep the borders open? B: I don't think that the U.S. should keep the borders open.
20. Aliens. A: Do you believe in aliens? B: I'm into it.
21. Welfare. A: How do you feel about California's welfare system? B: I'm not that opposed to it. I'm- I'm generally more in favor of like a socialized form of government.
22. Abortion. A: So how do you feel about abortion? B: I'm pro choice.
23. Aliens. A: Do you think that aliens have come here? B: All that alien stuff is so supernatural and I don't know if life on other planets would really be like that, or I don't know if it's some image we've all created in our mind.
24. Philosophy. A: Are people inherently good? B: I think that's strictly a matter of belief. It's almost as subjective as "Do you believe in God?".
25. Religion. A: Are you religious? B: I consider myself religious.
26. Death penalty. A: So is it- you think it'll get to the point where you kill a person, you- you die? B: I think it's gonna come down to that.
27. Immigration. A: See, like I've been over the border like a lot of times *and um* B: *uhhuh*
A: all they do is make me talk. B: I just think that's racist. I don't know. (Asterisks indicate overlap.)
28. Music. A: So what kind of music you like? B: I'm pretty much like into hip-hop and rap.

29. Prison.A: The three strikes thing, I mean, I think it should focus only on violent criminals. B:

People are there for, like, selling acid.

30. Death penalty. A: Like Charles Manson, you know, if he'd I don't know if he I don't know how's his whole story is, but if he killed all the people that they say he supposedly killed, should he be killed? B: Yes. I do believe in eye for an eye. Table 1: Experiment 1 Means and Standard Deviations Averaged Across Subjects and Items			
	Question 1 1=easy to talk 7=hard to talk	Question 2 1=deceptive 7=honest	Question 3 1=uncomfortable 7=comfortable
1 s pause	2.9 (.9)	5.4 (.7)	4.8 (.8)
3 s pause	4.1 (.9)	5.0 (.8)	4.1 (.8)
um 1 s	3.6 (.8)	4.9 (.8)	4.3 (.8)

Table 2: Experiment 2 Means and Standard Deviations Averaged Across Subjects and Items			
	Question 1 1=easy to talk 7=hard to talk	Question 2 1=deceptive 7=honest	Question 3 1=uncomfortable 7=comfortable
.5 s pause	2.3 (.8)	5.7 (.7)	5.4 (.7)
3 s pause	3.6 (1)	5.1 (.9)	4.5 (.8)
um 3 s	4.4 (.8)	4.8 (.9)	4.1 (.9)

Table 3: Experiment 3 Means and Standard Deviations			
---	--	--	--

Averaged Across Subjects and Items			
	Question 1 1=easy to talk 7=hard to talk	Question 2 1=deceptive 7=honest	Question 3 1=uncomfortable 7=comfortable
.5 s pause	2.9 (1)	5.4 (.7)	5.1 (.9)
~4 s pause	4.1 (.8)	5.0 (.8)	4.2 (1)
um 3 s	4.5 (.8)	4.8 (.8)	4.2 (.8)

Table 4: Summary of Predictions and Results (the symbol “<” means “more positive than”)			
	Experiment 1	Experiment 2	Experiment 3
Predictions			
Filter Hypothesis	(1 s = um 1 s) < 3 s	.5 s < (um 3 s = 3 s)	.5 s < (um 3 s = ~4 s)
Duration Hypothesis	1 s < um 1 s < 3 s	.5 s < 3 s < um 3 s	.5 s < (um 3 s = ~4 s)
Separate Contributions Hypothesis:			

•Expectation

Subhypothesis	1 s < (um 1 s = 3 s)	.5 s < (um 3 s = 3 s)	.5 s < (um 3 s = ~4 s)
•Responsibility Subhypothesis	(1 s = um 1 s) < 3 s	.5 s < 3 s < um 3 s	.5 s < ~4 s < um 3 s
•Interpreting-in-Context Subhypothesis	1 s < 3 s 1 s < um 1 s	.5 s < 3 s < um 3 s	.5 s < ~4 s .5 s < um 3 s
Results			
Production Difficulty	1 s < um 1 s < 3 s	.5 s < 3 s < um 3 s	.5 s < ~4 s < um 3 s
Honesty	1 s < (um 1 s = 3 s)	.5 s < 3 s < um 3 s	.5 s < (um 3 s = ~4 s)
Comfort with Topic	1 s < (um 1 s = 3 s)	.5 s < 3 s < um 3 s	.5 s < (um 3 s = ~4 s)

Author Note

I thank Millicent Alvarado, Greg Bryant, Izabel Caetano, Susan Daugherty, Kenny Han, Timothy Higgins, Nicole Mamier, Charles Metzging, Kristin Mayfield, Emily Murai, Dominique Peters, Meria Rivera, Monika Rosicka, Chad Schrock, Delia Silva, Ripple Suthar, Cassidy Taylor, and Laura Underwood for help preparing and running the experiments and for inspiring discussions. I thank Nameera Akhtar and Erich Fox Tree for comments on an earlier draft of this manuscript. The research was supported by faculty research funds granted by the University of California, Santa Cruz, and by a Summer/Short-Term Research Publication Grant granted by the American Association of University Women Educational Foundation. Correspondence may be addressed to Jean E. Fox Tree, Psychology Department, University of California, Santa Cruz, Santa Cruz, CA, 95064.