The Feeling of Another's Knowing: Prosody and Filled Pauses as Cues to Listeners about the Metacognitive States of Speakers

Susan E. Brennan

and

Maurice Williams

State University of New York at Stony Brook

Address for correspondence:

Susan E. Brennan
Department of Psychology
State University of New York
Stony Brook, NY 11794-2500

Subject Index: Paralinguistic cues; filled pauses; question-answering; feeling of knowing

Running head: Paralinguistic cues in question-answering

Abstract

In question-answering, speakers display their metacognitive states using filled pauses and prosody (Smith & Clark, 1993). We examined whether listeners are actually sensitive to this information. Experiment 1 replicated Smith and Clark's study; respondents were tested on general knowledge questions, surveyed about their FOK (feeling-of-knowing) for these questions, and tested for recognition of answers. In Experiment 2, listeners heard spontaneous verbal responses from Experiment 1 and were tested on their feeling-of-*another*'s-knowing (FOAK) to see if metacognitive information was reliably conveyed by the surface form of responses. For answers, rising intonation and longer latencies led to lower FOAK ratings by listeners. For nonanswers, longer latencies led to *higher* FOAK ratings. In Experiment 3, electronically edited responses with one-second latencies led to higher FOAK ratings for answers and lower FOAK ratings for nonanswers than those with five-second latencies. Filled pauses led to lower ratings for answers and higher ratings for nonanswers than did unfilled pauses. There was no support for a *filler-as-morpheme* hypothesis, that "um" and "uh" contrast in meaning. We conclude that listeners can interpret the metacognitive information that speakers display about their states of knowledge in question-answering.

To communicate successfully, people need to coordinate their individual knowledge states. This requires accurately estimating and monitoring not only their own knowledge states, but also the knowledge states of others and the knowledge they have in common, or mutual knowledge (Clark & Marshall, 1981; Clark & Wilkes-Gibbs, 1986; Krauss & Fussell, 1991). Speakers design utterances taking into account the mutual knowledge they share with their listeners (Clark & Wilkes-Gibbs, 1986) and they are able to continually monitor their listeners for evidence of understanding (Brennan, 1990); meanwhile, listeners interpret utterances by taking into account the common ground they share with speakers (Clark, Schreuder, & Buttrick, 1983). Speakers can intentionally express confidence in what they are saying (Scherer, London, & Wolfe, 1973). It is plausible that listeners estimate speakers' knowledge about, commitment to, and confidence in what they are saying by monitoring not only *what* speakers say, but *how* they say it.

Paralinguistic features of utterances such as pauses, intonation, and interjections have played almost no role in psycholinguistic studies of how listeners interpret utterances. Interjections such as "um," "uh," "hmm," clicks, and other noises that have been called *filled pauses* (Maclay & Osgood, 1959) or *fillers* are usually dismissed as disfluencies that listeners ignore. Yet listeners may continually process paralinguistic information along with other levels of linguistic information. For example, the filler "um" in spontaneous speech seems to help listeners recognize an upcoming target word faster; fillers may cause listeners to pay more attention to what is to follow (Fox Tree, 1993). Paralinguistic cues may directly affect pragmatic interpretation as well, such as when intonation helps a listener identify what speech act a speaker is trying to initiate, or when the manner in which an utterance is produced leads a listener to conclude that the speaker is confident or tentative. Consider this hypothetical interchange:

Lyn: Can I borrow that book?

Steve: ……………um…………………all right.

It is likely that Steve's delay and interjection will lead Lyn to different inferences than if Steve had responded "all right" immediately. However, most of the research on paralinguistic features of speech has examined what speakers produce, but not what listeners perceive (Levin, Schaffer, & Snow, 1982; O'Connell & Kowal, 1983).

In three experiments, we focus on question-answering in order to examine the contribution, if any, of fillers, latency to response, and rising or falling intonation to listeners' interpretations of utterances. The interpretation of an utterance, we assume, includes listeners' estimates of speakers' confidence about or commitment to what they are saying. This kind of metacognitive information is likely to be important in spontaneous conversations, where language is used in the service of goals and social relationships.

Estimating One's Own and Other's Knowledge

People's ability to accurately assess and monitor their own knowledge has been called the "feeling of knowing" or FOK (Hart, 1965). FOK is important in models of question-answering, which typically include a search component, in which a respondent tries to generate candidate answers, and a monitoring component, in which the respondent selects an answer (Glucksberg & McCloskey, 1981; Reder, 1979; 1987). To answer a question, then, a respondent must be able to verify whether information retrieved from memory is likely to be relevant or correct. To respond "I don't know," she must decide when to stop searching (Williams & Hollan, 1981; Glucksberg & McCloskey, 1981). FOK and latency to response are negatively correlated when retrieval is successful and positively correlated when it is not (Glucksberg & McCloskey, 1981; Smith & Clark, 1993). That is, the higher the reported FOK, the faster people retrieve answers from memory, but the longer they search for elusive answers before giving up. People's FOK is remarkably accurate, even for retrieval failures (Blake, 1973; Hart, 1965).

A speakers' knowledge, as displayed in her[1] answer to a question, could be evaluated by a listener using several different sources of information. First, the listener can rely on his own knowledge. If he happens to know the correct answer to a question, he can easily judge whether someone else's answer is correct; if he can't retrieve the answer, he can use his own FOK to predict whether another person would be likely to know the answer (Jameson, Nelson, Leonesio, & Narens, 1993). People tend to overestimate the likelihood that what they know is also known by others (Fussell & Krauss, 1991; Nickerson, Baddeley, & Freeman, 1987). Second, independent of whether they themselves know the answer, listeners can try to assess the difficulty of a question for the average person or for the typical member of a particular community and use that information to judge a speaker's confidence; this amounts to making an actuarial judgment of the question's difficulty (Jameson et al., 1993). People are reasonably good at assessing the knowledge of others in some domains, using, when relevant, characteristics such as gender (Fussell & Krauss, 1992) and expertise about a geographic area (Isaacs & Clark, 1987), although their estimates are still biased in the direction of their own knowledge (Fussell & Krauss, 1991; Nickerson, Baddeley, & Freeman, 1987). Third, they can base their judgments on what is mutually evident to their partner and themselves; this includes information from their shared physical environment and from immediately previous conversation (dubbed *visual copresence* and *linguistic copresence* by Clark and Marshall, 1981). Estimates of mutual knowledge sometimes appear to be biased in the direction of people's own knowledge; in a study of "mutual memory," people appeared to rely on their own memory for a past event they had experienced together in order to estimate whether their partner would remember it (Isaacs, 1990). Fourth, people can base their estimates of others' knowledge on information about the other person's ability or previous performance (Vesonder & Voss, 1985). And finally, they may be able to rely on the paralinguistic information displayed in the surface features of speakers' responses, including the intonation of answers, latency to response, and accompaniments to answers and nonanswers such as hedges and interjections.

Some have tried to assess, more or less explicitly, what we will call the "feeling-of-*another's*-knowing" or *FOAK*. In a study by Vesonder and Voss (1985), respondents experienced trials in which they were presented with 48 sentences (one at a time) and were asked to predict whether they would be able to recall each one when cued with its subject phrase. After the first cued recall test, there were three more trials with the same sentences. Predictions about each respondent's recall for each sentence were made by trios consisting of the respondent, a "listener" who heard the respondent's attempts to recall the same sentence aloud previously, and a judge who could see the items to be recalled but who heard white noise during the respondent's previous recall attempts. Results were that the listener had no advantage over the respondent; both respondent and listener predicted the respondent's recognition memory with equal accuracy, more accurately than did the judge. This led the authors to conclude that both the respondent's and the listener's predictions were based on the characteristics of the items to be recalled and on the respondent's previous performance, and not on the kind of privileged information about self-knowledge that is presumed to be involved in FOK (available only to the respondent).

It is possible that in Vesonder and Voss's study, the respondent's FOK was not really tapped because the task involved memorizing sentences instead of being truly knowledge-based. Another possible explanation for the results is that the respondent's observable behavior led to a "feeling of another person's knowing" on the part of the observer that happened to be as reliable a predictor of the respondent's performance as the respondent's own FOK (Jameson et al., 1993). Using similar respondent, observer, and judge roles along with a more knowledge-based task, Jameson et al. did find a difference between respondents and observers; when an observer had seen a respondent trying to recall the answers to general knowledge questions presented on a video monitor, the observer was less accurate in predicting the respondent's future recognition performance than was the respondent herself.

The judge, who saw only the knowledge questions and who was not present at the same time as the respondent (and thus did not witness her behavior during recall), made predictions about the respondent's recognition performance that were above chance, but these predictions were less accurate than the observer's predictions. Observers appeared to use specific cues in the respondents' observable behavior, including respondents' latency to response (Jameson et al., 1993).

Neither Jameson et al.'s study nor Vesonder and Voss's allowed any overt verbal communication between respondents and observers during the question-answering task. In Jameson et al.'s study, the respondents were not allowed to speak, but typed their answers to questions at a computer terminal while the observers watched over their shoulders (in experiment 1) or through a one-way mirror (experiment 2). So these studies do not address the question of how people interpret paralinguistic information during conversation.

A study by Smith and Clark (1993) investigated FOK in a conversational setting and proposed an interactive model of question-answering. This model adds a third component to the search and monitoring components of question-answering -- *producing a response* -- that is driven by social goals such as self presentation. The claim is that speakers deliberately display their metacognitive states to listeners using prosodic cues such as rising intonation and filled pauses. Smith and Clark's study followed the standard FOK paradigm, where people tried to recall the answers to general knowledge questions, then estimated their FOK about each question, and then were tested for their recognition of the correct answers. Smith and Clark found that FOK was positively correlated with recognition and with response latency when retrieval failed and negatively correlated when retrieval succeeded and an answer was produced. In addition, Smith and Clark examined the surface forms of responses, finding that people used rising intonation, fillers such as "uh," and hedges more often with incorrect than correct answers, and that they reported experiencing a lower FOK after answers they had produced with rising intonation than those with falling intonation. Smith and Clark also looked at the forms of nonanswers; FOK was higher and latency to response was longer for "I don't know" responses than for "I have no idea" responses, and even more so for "I don't remember" responses.

Smith and Clark's study focused only on the metacognitive information displayed in speakers' responses, and so their interactive model of question-answering lacks supporting data on two counts. Their study did not examine (1) whether listeners are actually sensitive to this information about speakers' metacognition, or (2) whether speakers' metacognitive displays are deliberately intended to be recognized by listeners. The present investigation aims to answer the first question (since intonation, latency, and interjections are not effective signals for a listener unless the listener is actually capable of recognizing and interpreting them). Experiment 1 had two purposes: to replicate relevant findings from Smith and Clark's study of metacognitive displays by speakers and to generate a set of answers that we could use as stimuli for listeners in Experiments 2 and 3. Experiment 2 examined whether listeners were sensitive to the intonation of answers, latencies to responses, and the form of nonanswers. Experiment 3 replicated Experiment 2, but with the additional control afforded by digital editing. In addition, it directly compared the effects of filled pauses that contain the interjections "uh" and "um" with silent pauses equivalent in length to test whether these interjections are simply associated with different amounts of delay, or whether listeners interpret them as contrasting indicators of the speaker's metacognitive state.

Experiment 1

The first experiment used the feeling of knowing paradigm (Hart, 1965; Wilkenson & Nelson, 1984; Smith & Clark, 1993). Respondents were tested orally on general knowledge questions, then surveyed about their FOK (feeling-of-knowing), and finally, tested for recognition of the correct answers to the question.

Method

Subjects. Fourteen students volunteered to participate in exchange for research credit in a psychology course at the State University of New York at Stony Brook. They were asked to answer some general knowledge questions. All of them gave permission for the tape recordings of their voices to be used later as stimuli in other experiments.

Materials. Twenty general knowledge questions were chosen from the same database of 300 questions from which Smith and Clark selected their questions (Nelson & Narens, 1980). Since these questions and their probabilities of correct recall were developed by Nelson and Narens over a dozen years ago, it seemed likely that today's college students would differ in their knowledge of the same items. So, using the norms only as a rough guide, we chose a set of 20 questions that we expected would vary in difficulty for our subjects. These questions are listed in Appendix A.

Procedure. Smith and Clark's three-part procedure was used. First, the experimenter read the questions aloud one at a time, waited until the subject answered, and then went on to the next question. During this part, the experimenter did not tell subjects whether or not their answers were correct, and he faced away from them in order to avoid providing any visual cues. This session was tape-recorded. Second, subjects completed a feeling of knowing questionnaire where they rated each of the 20 questions as to whether they thought they would recognize the answer in a multiple choice test, from 1 ("definitely not recognize") to 7 ("definitely recognize"). Finally, subjects took a multiple choice test, choosing from four alternatives to answer each of the 20 questions.

Analysis

The 14 respondents' 280 tape-recorded responses to the knowledge questions were transcribed in detail by the second author, including fillers, self-speech, punctuation ("." or "?") to differentiate final-rising intonation from flat or falling intonation, and notation as to whether the latency to the final answer was short or long. Any unclear speech was marked as such, using parentheses. Six individual responses were discarded because the respondent didn't hear the question and asked the experimenter to repeat. The remaining 274 responses were digitized on a Macintosh IIci computer using MacRecorder. Latencies to responses were measured from the offset of the final word of the experimenter's utterance to the onset of the phrase the respondent used to express an answer or nonanswer (not including any filled pauses that preceded the answer or nonanswer), using the software SoundEdit Pro 1.0. To examine the relationships between FOK and latency to response, we computed correlation coefficients for these two variables for each of the 14 respondents, transformed the individual correlations into Fischer's z_r scores in order to average them, and compared the mean z_r score to zero (following Smith & Clark, 1993). The correlations we report have been transformed back from mean z_r scores.

Results and Discussion

FOK ratings, recall, and recognition. Respondents' mean FOK ratings were higher when they were able to produce an answer than when they were not, 5.9 to 4.0 (with subjects as a random factor, $F1(1,13) = 62.12$, $p < .001$; with items as random, $F2(1,19) = 40.90$, $p < .001$). Their mean FOK ratings were higher for their correctly recalled answers than for their incorrect ones, 6.6 to 5.1 ($F1(1,13) = 28.73$, $p < .001$; $F2(1,19) = 28.77$, $p < .001$). Respondents also reported higher FOK for the items that they later recognized correctly on the multiple choice test, 5.7 to 4.1 ($F1(1,13) = 47.17$, $p < .001$; $F2(1,19) = 33.64$, $p < .001$). So respondents' FOK was an accurate predictor of recall and recognition performance.

Correlation of FOK and latency to final response. We measured latency to final response from the offset of the question asked by the experimenter to the onset of the phrase the respondent used to express the answer or nonanswer (excluding any preceding fillers or self-speech). Respondents produced their answers more quickly when they had a higher FOK. This was the case both for correct answers (r = -.69, p < .01) and for incorrect answers (r = -.50, p <.06). When people could *not* produce an answer, the higher their FOK, the longer they searched before giving up (r = .65, p < .01). These results are consistent with standard findings on metamemory and latency to response (e.g. Glucksberg & McCloskey, 1981).

Intonation. Respondents used rising intonation with incorrect answers 64% of the time and with correct answers only 33% of the time ($F1(1,13) = 20.09$, $p < .001$; $F2(1, 19) = 8.83$, $p < .005$). This result is consistent with Smith and Clark's; their subjects used rising intonation twice as often with incorrect answers as with correct answers, 44% to 21%. Answers produced with rising intonation had a mean FOK rating of 5.47, marginally lower than the FOK rating of 6.63 for answers with falling intonation ($F1(1,136) = 14.25$, $p < .001$; $F2(1,19) = 2.57$, $p = .11$). These FOK ratings for answers with rising and falling intonation are similar to those in Smith and Clark's study (5.61 to 6.04).

Our results replicate Smith and Clark's and support the conclusion that in conversational settings, people display metacognitive information about their knowledge states in their latency to response, the intonation of their answers, and the form of their nonanswers. The question remains: Can listeners interpret the information in these displays?

Experiment 2

In judging whether another person is likely to know the answer to a question (FOAK), listeners can base their judgments on actuarial information such as the probability that a particular fact would be known by the average person or by a person in the speaker's social or gender category (Fussell & Krauss, 1992; Jameson et al., 1993). Listeners can also use their own knowledge or FOK about a question (Fussell & Krauss, 1991; Jameson et al., 1993). But in order to focus only on listeners' sensitivity to the metacognitive information displayed in the surface features of the answers, and not on actuarial information or on listeners' own knowledge, we edited an audio tape of the answers generated by respondents in Experiment 1 without including the questions that had elicited them. To preserve each answer's pragmatic status as part of a question-answer pair, we preceded it with a generic question, e.g. "What is the answer to Question *N*?" where *N* referred to the answer's order on the audio tape.

Method

Subjects. Subjects were 48 undergraduates at the State University of New York at Stony Brook who volunteered in exchange for research participation credit in a psychology class. None had participated in Experiment 1. They did the task in a room with the experimenter and up to two other subjects, seated in carrels where they could not see one other. Data from ten subjects were not analyzed but were replaced with data from ten additional subjects for these reasons: four subjects were not native English speakers, two had heard details about the experiment from other students, two conspired as they filled out their test booklets, and twice there was equipment failure.

Materials. We chose 32 answers out of the 280 responses from the textual transcripts of the tapes generated during the general knowledge test in Experiment 1. Responses from two of the 14 respondents were eliminated because the ambient quality of the recordings did not match that of the others, and six individual responses were discarded because the respondent didn't hear the question and asked the experimenter to repeat. This left 234 responses to choose from, including 137 answers (73 correct and 64 incorrect), and 97 nonanswers (where the respondent failed to come up with an answer). From these, we chose 60 responses, consisting of 32 answers and 28 nonanswers. Responses were chosen randomly from the transcript, but were replaced until we had iteratively satisfied all the following constraints: No two answers gave the same content word. Half of the items in the answer and nonanswer sets were preceded by a short pause and half by a long pause (as had been categorized subjectively by the second author and noted in the transcripts). Of the answer set, half had been spoken with rising intonation and half with falling intonation. Of the nonanswers, in half the respondent claimed to know the answer but to be unable to produce it (e.g. "I can't remember" or "I forget"), and in the other half the respondent claimed not to know the answer (e.g. "I don't know" or "I have no idea").

These 60 responses were ordered randomly and edited manually onto the analog audio track of a videotape, using the insert controls on a videocassette recorder. The videotape image contained visible SMPTE time code, accurate to 1/30 sec.; this was used to help preserve the latency from the end of the original question to the onset of the response as closely as possible (to the nearest 1/30 of a second). Each response was preceded by the first author's voice asking a generic question, e.g. "What is the answer to question one?" Each response was followed by approximately 10 seconds of silence, in order to allow listeners time to rate what they had heard.

Procedure. People were told that they would be hearing other people's responses to trivia questions. They were asked to listen carefully to the recording and estimate how likely it was that the listener knew the correct answer. These feeling-of-another's-knowing (FOAK) judgments were made in a booklet of items numbered from 1-60. For each item that was an answer, the booklet asked "Do you think this was the correct answer to this question?" and provided a rating scale from 1 ("definitely incorrect") to 7 ("definitely correct"). For each item that was a nonanswer, the booklet asked "How likely do you think this person would be to recognize the correct answer to this question on a multiple choice test?" and provided a rating scale from 1 ("definitely not recognize") to 7 ("definitely recognize"). The questions in the booklet were worded in order to make the scale for answers as comparable as possible to that for nonanswers. The task took about 20 minutes to complete, and people had no difficulty making the ratings.

Results

FOAK, as measured on the two seven-point scales for answers and nonanswers, was higher overall for answers than for nonanswers, 5.02 to 2.73 ($F1(1,47) = 248.17$, $p < .001$; $F2(1,58) = 102.84$, $p < .001$). FOAK ratings for answers were affected by the latency to and the intonation of those answers; FOAK ratings for nonanswers were affected by the latency to and the form of the nonanswers. As predicted, the effects of latency were different for answers than for nonanswers. We will consider these effects for answers and nonanswers in turn.

Answers. For items that were answers, listeners rated FOAK 1.0 point higher on average when the answer followed a pause categorized as short than when it followed a pause categorized as long ($F1(1,47) = 229.4$, $p < .001$; $F2$(1,28 = 14.6, $p < .001$).

Listeners also rated FOAK marginally higher when the answer was in the falling intonation category than when it was in the rising intonation category, by .6 points ($F1(1,47) = 30.4$, $p < .001$; $F2(1,28) = 2.8$, $p = .11$). The interaction of latency and intonation was reliable by subjects ($F1(1,47) = 12.6$, $p < .001$), but not by items ($F2$(1,28 = .50, *n.s.*). So for these particular items, intonation appeared to make more of a difference when the answer followed a short pause than when it followed a long pause (but since this result is not reliable by items, we cannot generalize to other answers). For mean FOAK ratings, see Table 1.

Insert Table 1 about here

Nonanswers. The phrasing of nonanswers made a reliable difference in FOAK; listeners' mean rating for responses in the form "I can't remember" was 1.22 points higher than for "I don't know" responses ($F1(1,47) = 75.73$, $p < .001$; $F2(1,24) = 45.68$, $p < .001$).

Listeners made marginally higher FOAK judgments when the latency to nonanswer was in the long category than when the latency was in the short category, by .29 points. This difference was reliable by subjects ($F1(1,47) = 14.25$, $p < .001$) but only marginally by items ($F2(1,24) = 2.5$, $p = .127$). For mean FOAK ratings for nonanswers, see Table 2.

Insert Table 2 about here

Discussion

As predicted by the interactive model of question-answering, listeners' FOAK judgments were sensitive to the surface features of speakers' responses. Long pauses led to lower FOAK ratings of answers and higher FOAK ratings of nonanswers than did short pauses. These results for listeners mirror the FOK results for speakers obtained in Experiment 1; that is, in question-answering, both FOK and FOAK are negatively correlated with latency to answers and positively correlated with latency to nonanswers.

Recall that our FOAK scales for answers and nonanswers asked about the likelihood that the speaker had correct knowledge. By removing the questions that elicited the responses, we intended to measure the listener's response to the confidence the speaker displayed paralinguistically in her answer, and not the listener's own FOK. However, it is possible that the disembodied answers may still have elicited FOK about those topics in the listeners. So we tested another group of 40 SUNY undergraduates for their FOK by presenting each answer in a text questionnaire (e.g. "How likely is it that you would know the answer to a trivia question if the answer was 'dermatologist'?"). Then we correlated the mean FOK rating for each answer with its mean FOAK from Experiment 2. FOK and FOAK were uncorrelated (r = .06, *n.s.*).

Another possible confound would occur if FOAK ratings were affected by social stereotypes evoked by different voices. Post hoc, we examined FOAK ratings to see if there were any differences by speaker. There were no differences among individual speakers, nor were there differences by speaker's gender. So it appears that listeners' judgments about speakers' knowledge were based on speakers' displays of confidence.

A difficulty with Experiment 2 is that the latencies to speakers' responses were categorized subjectively by the experimenter as either long or short, rather than being measured. It is possible that some pauses may have been miscategorized, adding noise to the results. In addition, although speakers' responses were selected from text transcripts on the basis of latency and rising or falling intonation, uncontrolled factors may have contributed to the impressions these responses made on listeners. For example, sometimes the interval before the final response contained fillers such as "um" and "uh," clicks, sighs, and other speech (e.g. "...oh no... ooy... I can't think of it... epidermis?"). In Experiment 3, we aimed to test specific hypotheses about the content of these intervals and also to replicate Experiment 2 while exerting more control over items.

Experiment 3

According to the interactive model of question-answering, a filler is a signal that a speaker can use to display to a listener that she is actively engaged in searching for information. Fillers are strongly associated with delay; in Smith and Clark's experiment, the longer a speaker paused before producing an answer, the more likely the pause was to contain one or more fillers. In their study, speakers produced correct responses after an average silent pause of .97 second and after an average filled pause of 5.00 seconds. Speakers produced nonanswers after an average silent pause of 3.39 seconds and after an average filled pause of 10.30 seconds (Smith & Clark, 1993).

What do listeners make of filled pauses? Experiment 3 unconfounded latency and fillers to test whether the presence or absence of filled pauses affects FOAK judgments. We predicted that when a speaker produces a nonanswer, a listener should rate FOAK higher when there is a filled pause (because the speaker has signaled that she is just having trouble retrieving the information) than when there is an unfilled (silent) pause of equal length. And when the speaker produces an answer, the listener should rate FOAK lower when the pause is filled (again, because the speaker has signaled trouble) than when the pause is silent and of equal length.

The particular form of a filled pause may also lead to a contrast in meaning. James (1972, 1973) proposed that interjections such as "uh," "oh," and "ah" differ semantically and that "uh" is a signal that the speaker is having retrieval trouble. In Smith and Clark's study, for answers and nonanswers, the filler "um" both reacted to and predicted longer pauses than the filler "uh." That is, "um" was both preceded by and followed by longer silences than "uh." The mean latency to response (for both answers and nonanswers) was 3.53 seconds when "uh" preceded the response and 7.82 seconds when "um" preceded the response. Smith and Clark concluded that "um" and "uh" are not interchangeable, and that speakers "can choose between 'um' and 'uh' to signal the depth of their ongoing retrieval problem" (p. 37). They presented the alternation between these two interjections as a distinct choice and argue that these interjections should also serve distinct purposes, consistent with E. Clark's Principle of Contrast (1987). "Our proposal is that respondents use 'uh' both to signal a delay and to offer a brief account for it. In this view, 'uh' is not merely a filled pause -- which is an oxymoron anyway. It is a deliberate signal chosen from a range of interjections that include 'uh,' 'um,' 'hm,' 'mm,' and the tongue click 'ts'" (Smith & Clark, 1993, p. 27).

But does the form of a filler make any difference to listeners? In a study by Fox Tree (1993) on monitoring words in spontaneous speech, listeners were marginally faster at recognizing a target word when it followed a filled pause containing "um" than when the filled pause was replaced with a silent pause. There was no difference between "uh" and a silent pause, leading to the proposal that "uh" and "um" may have different effects on the listener, like different discourse markers (Fox Tree, 1993). Discourse markers such as "oh," "well," and "so" are metalinguistic particles that appear to be used by speakers to signal and manage the information structure of spontaneous discourse (Schiffrin, 1987). Discourse markers are used in different contexts from one another and do not appear to be interchangeable. If "um" and "uh" are really like discourse markers and mark a morpheme-like contrast in meaning, then listeners should be able to interpret this contrast. We will call this the *filler-as-morpheme* hypothesis. In concert with the interactive model of question-answering, the filler-as-morpheme hypothesis predicts that before an answer, "uh" should lead listeners to estimate a higher FOAK and "um," a lower FOAK. In the context of a nonanswer, the prediction is reversed: "uh" should lead to a lower FOAK and "um," a higher FOAK. Another version of the filler-as-morpheme hypothesis is that fillers could be like morphemes is if listeners treat "um" and "uh" as contrasting but with fixed meanings, regardless of the context. That is, the presence of "um" vs. "uh" before any response (answer or nonanswer) may lead to a difference in FOAK rating in the same direction.

A final possibility is that listeners may not distinguish at all between "um" and "uh" in their FOAK ratings. Although speakers may produce these fillers in distinct circumstances (e.g. as reactions to pauses or to FOK), their alternation may not be morphemic, but simply an emergent consequence of producing an answer. After all, "um" and "uh" are phonetically very similar; a speaker can turn an "uh" into an "um" by closing the lips to produce a bilabial nasal consonant [m]. In Experiment 3, we tested the filler-as-morpheme hypothesis. We used the same task as in Experiment 2, varying latency and fillers within items by editing them digitally.

Method

Subjects. Subjects were 72 undergraduates at the State University of New York at Stony Brook who volunteered in exchange for research participation credit in a psychology class. None had participated in Experiments 1 or 2, and all identified themselves as native speakers of English. They did the task in a room with the experimenter and up to two other subjects, seated in carrels that faced away from each other.

Materials. Our design called for half nonanswers and half answers, with half of the answers having rising intonation and half, falling. Within each response item there were four types of pauses: a one-second unfilled pause, a five-second unfilled pause, a five second pause containing "uh," and a five second pause containing "um." We selected responses using the text transcripts from Experiment 1. First we screened the 234 responses that had provided the selections for Experiment 2, looking for the fillers "um" and "uh." Ten speakers produced one or more distinct "um"s (for a total of 29), while nine produced one or more distinct "uh"s (for a total of 31). Eight speakers produced one or more instances of *both* kinds of fillers. We considered responses from only those eight speakers in constructing the stimuli.

Answers were chosen in multiples of four from the same speaker and in the same intonation category, so that we could vary pause length and filler within answers. Of the eight speakers who produced both "um" and "uh" fillers, only six produced at least four answers with the same category of intonation (rising or falling). We chose answers from the text transcripts, replacing individual answers to satisfy these constraints: No two answers contained the same content word, half had rising intonation, and half had falling intonation. The maximum set of answers we could find to satisfy these constraints was twenty four, and these were drawn from four speakers.

For the 24 *non*answers, we did not have intonational or content word constraints, so we chose them randomly from the text transcripts of six speakers who had produced an "um," an "uh," and at least four nonanswers (these speakers included the same four speakers who contributed the 24 answers). There were not enough "I forget" responses for us to compare "I forget" with "I don't know" responses as in Experiment 2, so these were combined. In 22 of the nonanswers, speakers claimed to not know the answer, and in two, they claimed to have forgotten it.

Finally, we selected one distinct "um," one distinct "uh," and one unfilled pause (longer than one second) from the transcripts of the six speakers who contributed responses. When a speaker's transcript offered several fillers to choose from, we randomly chose an "uh" or "um" whose length was mid-range for that speaker. Then we located the appropriate filled pauses, unfilled pauses, and responses on the audiotapes and used MacRecorder to digitize these on a Macintosh IIci. We used SoundEdit Pro 1.0 to digitally edit four templates from this material for each of the six speakers: (1) *short*, an unfilled pause of one second, (2) *long*, an unfilled pause of five seconds (edited from several copies of the shorter unfilled pause), (3) *um*, an unfilled pause of three seconds followed by "um" and more silence (five seconds total), and (4) *uh*, an unfilled pause of three seconds followed by "uh" and more silence (five seconds total).

We made four digital copies of each answer and nonanswer and concatenated each one electronically onto a copy of one of the four templates produced by the same speaker who had originally produced the answer or nonanswer. Finally, we constructed four stimulus lists (A, B, C, D) such that a particular answer or nonanswer appeared once per list, accompanied by a different kind of pause on each list. A single random order was used for the responses on all four lists. For example, "Charles Lindbergh" appeared on all four lists as item #20, but on List A it was preceded by "um," on List B by "uh," on List C by a short unfilled pause, and on List D, by a long unfilled pause. An audiotape was made of each of the four lists. In sum, each list contained three instances of each of eight different kinds of answers (two kinds of intonation by four kinds of pause templates: short, long, um, or uh) for a total of 24 answers, plus six instances of each of four different kinds of nonanswers (using the four kinds of pause templates) for a total of 24 nonanswers. The first two items per list (an answer and a nonanswer) were practice items, for a total of 50 items per list.

Procedure. Each subject heard only one response tape. As in Experiment 2, each response was preceded on the tape by a generic question and followed by approximately 10 seconds of silence during which listeners rated what they had heard. After listening to a response, subjects made a feeling-of-another's-knowing (FOAK) judgment about it in a booklet of items numbered from 1-50. As in Experiment 2, the booklet asked "Do you think this was the correct answer to this question?" for answers and "How likely do you think this person would be to recognize the correct answer to this question on a multiple choice test?" for nonanswers, with rating scales from 1-7. Subjects were debriefed after the task; none of them appeared to be aware that the responses had been electronically constructed. None of the subjects appeared to find the task difficult or the stimuli unnatural.

Results

Mean FOAK was higher for answers than for nonanswers, by 1.91 points. As predicted, intonation, latency, and filled vs. unfilled pauses affected listeners' FOAK ratings. Figure 1 shows the overall pattern of results for answers and nonanswers.

Insert Figure 1 about here

Answers. Listeners' FOAK ratings were .75 higher with falling intonation than with rising intonation ($F1(1,71) = 70.95$, $p < .001$; $F2(1,22) = 16.50$, $p < .001$).

Listeners' FOAK ratings also depended on whether the pause before an answer was long, short, filled, or unfilled ($F1(3,213) = 129.39$, $p < .001$; $F2(3,66) = 96.57$, $p < .001$). Answers with one-second pauses were rated 1.59 points higher than answers with five-second pauses (means combined for filled and unfilled pauses; we will report all planned comparisons by subjects as $t1$ and by items as $t2$: $t1(213) = 28.26$, $p < .001$; $t2(66) = 16.39$, $p < .001$). This replicates Experiment 2, where short pauses before answers led to higher FOAK ratings than long ones. Filled five-second pauses before answers led to lower FOAK ratings than did unfilled five-second pauses, by .52 points ($t1(213) = 5.80$, $p < .001$; $t2(66) = -14.56$, $p < .001$). So not only do speakers rate their own FOK as lower in cases where they have answered with a filled pause than with an unfilled pause (in Experiment 1), but also listeners rate FOAK lower for answers articulated after filled pauses. The interaction between the four types of pause and the two types of intonation was reliable by subjects ($F1(3,213) = 3.95$, $p < .01$) but not by items ($F2(3,66) = 1.64$, *n.s.*; see Figure 1).

As for the filler-as-morpheme hypothesis, we found no support for it. The fillers "uh" vs. "um" did not contrast in FOAK ratings ($t1(213) = .6$, *n.s.*; $t2(66) = .5$, *n.s.*). Means for each kind of answer appear in Table 3.

Insert Table 3 about here

Nonanswers. Listeners' FOAK ratings were affected by the type of interval before a response ($F1(3,213) = 13.30$, $p < .001$; $F2(3,69) = 6.63$, $p < .001$). In particular, five-second pauses (whether filled or unfilled, means combined) led to .39 higher FOAK ratings than one-second pauses ($t1(213) = 4.88$, $p < .001$; $t2(69) = 3.38$, $p < .002$). When a speaker delayed five seconds before producing an "I don't know" response, listeners appeared to notice this and rated the speaker as slightly more likely to recognize the answer on a multiple choice test than a speaker who answered "I don't know" after only one second.

Filled five-second pauses before nonanswers led to .31 higher FOAK ratings than unfilled five-second pauses ($t1(213) = 3.65$, $p < .001$; $t2(69) = 2.54$, $p < .02$). So when a speaker displays her effort by producing a filled pause before giving up the search, a listener can interpret this as a signal that the speaker is, literally, less certain about her "I don't know." These effects of latency and fillers support the interactive model of question-answering.

There was a .20 difference in FOAK ratings of pauses containing "uh" and "um" before nonanswers; this planned comparison was reliable by subjects but not by items ($t1(213) = 2.02$; $p < .05$; $t2(69) = 1.42$, *n.s.*; see Table 3).

As with Experiment 2, we sought to eliminate any influence of listeners' FOK by correlating each disembodied answer's mean FOAK rating in Experiment 3 with its FOK rating from the text questionnaire administered to 40 SUNY undergraduates (see discussion, Experiment 2). FOK and FOAK were uncorrelated (r = .07, *n.s.*). We also examined FOAK ratings from Experiment 3 for differences among the six voices (four speakers were male and two were female). No differences existed among individuals nor between male and female voices.

To what extent were people aware of the strategies they used to make FOAK ratings? The 72 subjects were interviewed individually after completing the experiment and before being debriefed. In explaining their strategies, 20.8% explicitly mentioned that they listened for either how "confident" or how "uncertain" the speaker sounded. Most of the subjects (76.3%) mentioned speakers' hesitations or pauses before answers, but only 2.8% mentioned pausing in the context of nonanswers; 38.9% mentioned noticing either "tone of voice" or "questioning" intonation; 23.6% mentioned using tone of voice (or questioning intonation) in addition to latency. Only 6.9% claimed to have paid attention to either "um" or "uh" (though no one mentioned noticing both of these fillers).

Three of the subjects who mentioned fillers (4.2%) said they made their judgments based on the *length* of the filler. This was not systematically varied in the stimuli. Post hoc, we compared the lengths of the 12 different fillers inserted within the five-second pauses (the "um" and the "uh" from each of the six speakers who contributed stimuli for Experiment 3); "um" was longer than "uh," 527 to 344 msec. ($t(5) = 3.28$, p = .02). So not only is "um" strongly associated with a longer latency to response than "uh" (Smith & Clark, 1993), but "um" takes up more speaking time than "uh." There were no reliable correlations between listeners' FOAK ratings and the lengths of fillers within the five-second pauses; for answers, r = .09 for "uh" and r = .22 for "um", and for nonanswers, r = -.07 for "uh" and r = -.26 for "um" (all *n.s.*).

Discussion

Answers produced more quickly or with falling intonation sounded more likely to be correct than ones produced more slowly or with rising intonation. Nonanswers after longer pauses produced higher FOAK ratings than after shorter ones. These findings for listeners mirror Smith and Clark's (1993) associations between FOK, recall, and recognition for speakers. Listeners were sensitive to the presence or absence of fillers; they rated answers preceded by "um" or "uh" as less likely to be correct than those preceded by unfilled pauses of the same length, and they gave nonanswers preceded by a filler higher FOAK ratings than nonanswers preceded by unfilled pauses. So interjections within a pause appear to intensify the effect of that pause.

To what extent are speakers' displays during question-answering deliberate? That is, do they count as symbolic (non-natural) meaning in Grice's sense (1957), where a speaker intends to produce some effect in an addressee by means of the addressee's recognizing her intention? Or are they just natural symptoms of a speaker's effort that a listener can interpret? The length of the pause before an answer or before "I don't know" is not likely to be something ordinarily under a speaker's conscious control, to the extent that an answer cannot be displayed until it has been retrieved, although a speaker could purposely *delay* a response. However, the speaker may well be purposeful about the intonation she uses, and whether and how she fills the pause before her response.

A related question is whether speakers consciously control their interjection rates. In one study (Broen & Siegel, 1972), speakers produced more interjections in conversation than when speaking into a tape recorder while alone; supposedly, speakers tried harder to be fluent with the tape recorder because the monologue was considered more "important" than informal conversation. But content factors beyond the speaker's control may also influence interjection rates. Instructors giving humanities lectures used more filled pauses than those giving social science lectures, who in turn used more filled pauses than those giving lectures in the natural sciences; there were no differences in base rates of filled pauses when these instructors were interviewed about the same neutral topic (Schacter, Christenfeld, Ravina, & Bilous, 1991). Different interjection rates were attributed to the structure of knowledge in these three kinds of disciplines; e.g., when there are more choices to make in describing a topic, speakers use more filled pauses (Schacter et al., 1991). Whether they do so deliberately, however, is an open question. Smith and Clark proposed that fillers save face while displaying speakers' active engagement in searching for an answer or their lack of commitment to a particular answer. Whether the speaker's paralinguistic behavior is intentional (and intended to be face-saving) is beyond the scope of the current study. We have limited our perspective to that of the listener in order to determine whether a speaker's metacognitive displays can be easily interpreted.

Smith and Clark also raised the possibility that a speaker's choice between "um" and "uh" is a deliberate one. Although there is evidence that speakers produce these two fillers in different contexts, what the listener does with "um" vs. "uh" is another issue. If these fillers contrast as morphemes do, we would expect listeners to be able to distinguish between them in their interpretations. This was not the case; we found no clear support for the filler-as-morpheme hypothesis. There was no difference in FOAK ratings between "um" and "uh" in the context of answers. The marginal difference (n.s. by items) that we found in the context of nonanswers did go in the direction predicted by the filler-as-morpheme hypothesis. It is possible that the measures in Experiment 3 were not sensitive enough, or that our "um"s and "uh"s were not articulated distinctly enough to yield reliable differences. Alternatively, perhaps "um" and "uh" must each appear within a characteristic latency context to have a contrasting effect. Experiment 3 used an intermediate delay of five seconds, chosen to sound natural with both "um" and "uh." Recall that Smith and Clark found "um" to be associated with longer delays than "uh." When these fillers are spontaneously produced by speakers, they are confounded with latency, and the information that they may bear may be redundant.

From Experiment 3, it seems likely that a speaker's production of "um" or "uh" is not so much a premeditated choice as an emergent production. After all, a speaker can turn an "uh" into an "um" as she searches for an answer. Listeners appeared not to distinguish them; indeed, these interjections are much more similar than most other potentially contrasting morphemes. What may be deliberate on the part of the speaker is the overall *length* of the filler. A longer filler may react to a longer previous pause and predict a longer delay to response. And the longer the filler, the more likely the speaker may be to end in a bilabial nasal, changing an "uh" into an "um." This is consistent with Smith and Clark's conclusion that "um"s and "uh"s function both reactively and predictively.

General Discussion

The three experiments reported here support the interactive model of question-answering. Experiment 1 replicated Smith and Clark's study of question-answering, showing that while speakers search memory and monitor their search, they also display their metacognitive states. Experiments 2 and 3 showed that listeners can use this display as a cue to speakers' metacognitive states. A listener's FOAK, based on a speaker's display of confidence in or commitment to an answer, was affected by the intonation of answers, the form of nonanswers, the latency to response, and the presence of fillers. Even though "um" and "uh" are used by speakers in different contexts, we found no evidence that listeners interpret them as contrasting, nor was FOAK affected by length of filler (at least not for fillers used in Experiment 3). That listeners are sensitive to filled vs. unfilled pauses shows that paralinguistic displays can be used to estimate other people's knowledge, a necessity for coordinating individual mental states during communication.

We have investigated paralinguistic effects in an extremely limited context, that of answering trivia questions. In Experiment 1, the interaction between speakers and listeners was atypical of everyday conversation in that subjects tacitly assumed that the experimenter knew the answers to the questions but was unable to give them feedback about the correctness of their answers. However, the task *was* conversational in that the parties were copresent and the experimenter did interact with the subjects when they asked him to repeat a question or when they asked a question about the task. It remains to be shown whether displays in question-answering would change when there is no interactive partner, or when the partner is not presumed to know the answer. In Experiments 2 and 3, our listeners indicated their interpretations along a single dimension. In everyday conversation, we would expect the paralinguistic features of utterances to be context dependent, enabling speakers and listeners to accomplish a wide variety of pragmatic effects. Such effects might include conveying lack of compliance or enthusiasm (as in our first example), creating an impression of candidness or of evasiveness (Kraut, 1978), or conveying spontaneity (Kowal, Bassett, & O'Connell, 1985; Levin, Schaffer, & Snow, 1982).

Acknowledgments

Susan E. Brennan, Department of Psychology; Maurice Williams, Department of Linguistics.

Experiments 1 and 2 were conducted by the second author as part of an undergraduate research project under the supervision of the first author. This research was supported by NSF grant # 92-02458 to the first author and by the Minority Research Apprenticeship Program (Department of Education Grant # P202A30062), which funded the second author. We thank Michele Serpico for conducting Experiment #3, and Herbert H. Clark, Michael F. Schober and three anonymous reviewers for their insightful comments.

Address correspondence and reprint requests to Susan E. Brennan, Department of Psychology, State University of New York, Stony Brook, NY 11794-2500.

Notes

[1]To simplify matters, we assume the generic speaker and question-answerer to be female and the generic question-asker and listener to be male.

References

Blake, M. (1973). Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. Journal of Verbal Learning and Verbal Behavior, **12**, 311-319.

Brennan, S. E. (1990). Seeking and providing evidence for mutual understanding. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Broen, P. A. & Siegel, G. M. (1972). Variations in normal speech disfluencies. Language and Speech, **5**, 219-231.

Clark, H.H. & Marshall, C.R. (1981). Definite reference and mutual knowledge. In A.K. Joshi, B. Webber, & I.A. Sag (Eds.), Elements of discourse understanding (pp. 10-63). Cambridge, England: Cambridge University Press.

Clark, H. H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. Cognition, **22**, 1-39.

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), Mechanisms of language acquisition (pp. 1-33). Hillsdale, NJ: Erlbaum.

Clark, H. H., Schreuder, R. & Buttrick, S. 1983). Common ground and the understanding of demonstrative reference. Journal of Verbal Learning and Verbal Behavior, **22**, 245-258.

Ekman, P. & Friesen, W. V. (1975). Unmasking the face: A guide to recognizing emotions from facial cues. Englewood Cliffs, NJ: Prentice Hall, Inc.

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

Fussell, S. R. & Krauss, R. M. (1991). Accuracy and bias in estimates of others' knowledge. European Journal of Social Psychology, **21**, 445-454.

Fussell, S. R. & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. Journal of Personality and Social Psychology, **62**, 378-391.

Glucksberg, S. & McCloskey, M. (1981). Decisions about ignorance: Knowing that you don't know. Journal of Experimental Psychology: Human Learning and Memory, **7**, 311-325.

Grice, H. P. (1957). Meaning. Philosophical Review, **66**, 377-388.

Grice, H. P. (1975). Logic and conversation (from the William James lectures, Harvard University, 1967). In P. Cole and J. Morgan (Eds.), Syntax and semantics 3: Speech acts. (pp. 41-58). New York: Academic Press, .

Grice, P. (1989). Meaning revisited. In: Studies in the way of words. Cambridge, MA: Harvard University Press, pp. 283-303.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. Journal of Educational Psychology, **56**, 208-216.

Isaacs, E. A. & Clark, H. H. (1987). References in conversation between experts and novices. Journal of Experimental Psychology: General, **116**, 26-37.

Isaacs, E. (1990). Mutual memory for conversation. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

James, D. (1972). Some aspects of the syntax and semantics of interjections. Papers from the Eighth Regional Meeting. Chicago Linguistic Society.

James, D. (1973). Another look at, say, some grammatical constraints on, ok, interjections and hesitations. Papers from the Ninth Regional Meeting. Chicago Linguistic Society.

Jameson, A., Nelson, T. O., Leonesio, R. J. & Narens, L. (1993). The feeling of another person's knowing. Journal of Memory and Language, **32**, 320-335.

Kowal, S., Bassett, M. R., & O'Connell, D. C. (1985). The spontaneity of media interviews. Journal of Psycholinguistic Research, **14**, 1-18.

Krauss, R. M. & Fussell, S. R. (1991). Perspective-taking in communication: Representations of others' knowledge in reference. Social Cognition, **9**, 2-24.

Kraut, R. E. (1978). Verbal and nonverbal cues in the perception of lying. Journal of Personality and Social Psychology, **36**, 380-91.

Levin, H., Schaffer, C. A., & Snow, C. (1982). The prosodic and paralinguistic features of reading and telling stories. Language and Speech, **25**, 43-54.

Maclay, H. & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. Word, **15**, 19-44.

Nelson, T. O. & Narens, L. (1980). A new technique for investigating the feeling of knowing. Acta Psychologica, **46**, 69-80.

Nickerson, R. S., Baddeley, A., & Freeman, B. (1987). Are people's estimates of what other people know influenced by what they themselves know? Acta Psychologica, **64**, 245-259.

O'Connell, D., & Kowal, S. (1983). Pausology. In W. A. Sedelow & S. Y. Sedelow (Eds.), Computers in Language Research: Part 1. Formalization in literary and discourse analysis. Berlin: Mouton.

Reder, L. (1979). The role of elaboration in memory for prose. Cognitive Psychology, **11**, 221-234.

Reder, L. M. (1987). Strategy selection in question answering. Cognitive Psychology, **19**, 111-138.

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

Scherer, K. R., London, H., & Wolf, J. J. (1973). The voice of confidence: Paralinguistic cues and audience evaluation. Journal of Research in Personality, **7**, 31-44.

Schiffrin, D. (1987). Discourse Markers. New York: Cambridge University Press.

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

Vesonder, G. T., & Voss, J. F. (1985). On the ability to predict one's own responses while learning. Journal of Memory and Language, **24**, 363-376.

Wilkinson, T. S. & Nelson, T. O. (1984). FACTRETRIEVAL2: A PASCAL program for assessing someone's recall of general-information facts, confidence about recall correctness, feeling-of-knowing judgments for nonrecalled facts, and recognition of nonrecalled facts. Behavior Research Methods, Instruments, & Computers, **16**, 486-488.

Williams, M. D., & Hollan, H. D. (1981). The process of retrieval from very long-term memory. Cognitive Science, **5**, 87-119.

Table 1

Effects of Intonation and Latency to Answer on Mean FOAK Rating, Experiment 2

		Latency		
		Short	Long	Overall
Intonation	Falling	5.87	4.63	5.25
	Rising	5.21	4.36	4.79
	Overall	5.54	4.49	5.02

Table 2

Effects of Form and Latency to Nonanswer on Mean FOAK Rating, Experiment 2

		Latency		
		Short	Long	Overall
Form	"Don't know"	2.00	2.24	2.12
	"Can't remember"	3.17	3.50	3.34
	Overall	2.58	2.87	2.73

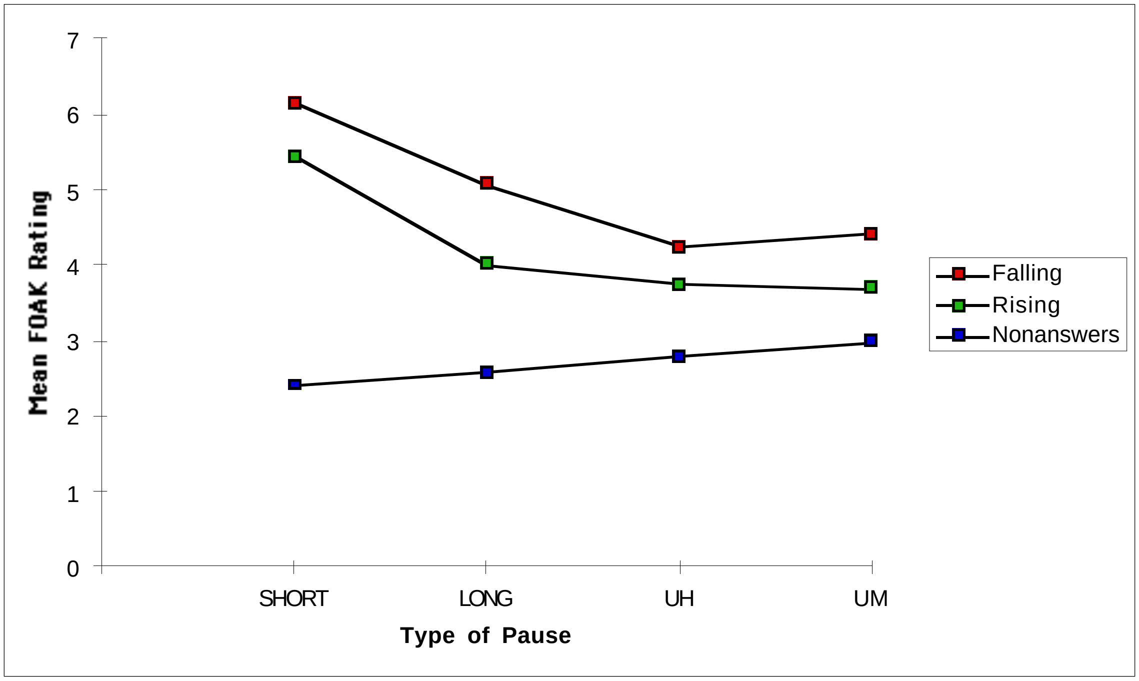
Table 3

Effects of Intonation, Latency, and Fillers on FOAK, Experiment 3

	Type of pause before response				
	Short, unfilled	Long, unfilled	Long, "uh"	Long, "um"	Overall
Answers					
Falling intonation	6.15	5.07	4.26	4.44	4.98
Rising intonation	5.44	4.03	3.75	3.69	4.23
Answers, overall	5.80	4.55	4.01	4.07	4.60
Nonanswers					
	2.42	2.60	2.81	3.01	2.71

Figure heading:

Figure 1. FOAK ratings for answers with rising and falling intonation and for nonanswers, as a function of pause type: Short (unfilled, one-second) pauses, long (unfilled, five-second) pauses, and the fillers "uh" and "um" (appearing within five-second pauses).


Mean FOAK Rating
7
6
5
4
3
2
1
0
SHORT
LONG
UH
UM
Type of Pause
Falling
Rising
Nonanswers

Appendix A - General knowledge questions, Experiment 1

1. What is the last name of the man who showed that lightning is electricity?
2. What is the name for a medical doctor who specializes in diseases of the skin?
3. What is the last name of the playwright who wrote "A Street Car Named Desire" ?
4. What was the name of the Zeppelin that exploded in Lakehurst, New Jersey in 1937?
5. What is the last name of the author of the book "1984"?
6. What is the last name of the first person to complete a solo flight across the Atlantic Ocean?
7. What is the last name of Dagwood's boss in the comic strip "Blondie"?
8. Of which country is Buenos Aires the capital?
9. What was the name of the Apollo Lunar Module that landed the first man on the moon?
10. What is the last name of the actress who received the best actress award for the movie "Mary Poppins"?
11. What is the name of the submarine in Jules Verne's "20000 Leagues Beneath the Sea"?
12. What is the last name of the second US president?
13. What is the last name of the author who wrote under the pseudonym of Mark Twain?
14. Of which country is Budapest the capital?
15. In what park is "Old Faithful" located?
16. What is the name of the Roman Emperor who fiddled while Rome burned?
17. What is the name of the organ that produces insulin?
18. What is the last name of the woman who founded the American Red Cross?
19. What is the capital of Canada?
20. What is the kind of cat that spoke to Alice in the story "Alice's Adventures in Wonderland"?