Acknowledgments in Human-Computer Interaction

Karen Ward and Peter A. Heeman

Center for Spoken Language Understanding
Oregon Graduate Institute of Science & Technology
20000 NW Walker Road
Beaverton, Oregon 97006
`wardk@cse.ogi.edu, heeman@cse.ogi.edu`

Abstract

Acknowledgments are relatively rare in human-computer interaction. Are people unwilling to use this human convention when talking to a machine, or is their scarcity due to the way that spoken-language interfaces are designed? We found that, given a simple spoken-language interface that provided opportunities for and responded to acknowledgments, about half of our subjects used acknowledgments at least once and nearly 30% used them extensively during the interaction.

1 Introduction

As our ability to build robust and flexible spoken-language interfaces increases, it is worthwhile to ask to what extent we should incorporate various human-human discourse phenomena into our dialogue models. Many studies have shown that people alter their dialogue techniques when they believe that they are talking to a computer (e.g., Brennan, 1991), so it is not clear that observations of human-human conversation will provide us with the guidance we need. At the same time, we cannot always look to current systems to determine which discourse phenomena should be supported. Current-generation interfaces are still relatively fragile, and so designers of spoken-language systems go to some effort to structure dialogues and create prompts that guide the user toward short, contentful, in-vocabulary responses (e.g., Basson et al, 1996; Cole, et al, 1997; Oviatt et al, 1994). One result of this approach is the suppression of meta-dialogue acts such as acknowledgement and repetition.

The term "acknowledgment" is from Clark and Schaefer (1989), who describe a hierarchy of methods by which one conversant may signal that another's contribution has been understood well enough to allow the conversation to proceed. Acknowledgments often appear in American English conversation as an "okay" or "uh-huh" that signals understanding but not necessarily agreement. These are also called "back channels" or "prompts" (e.g., Chu-Carroll & Brown, 1997),

Closely related to acknowledgments are repetitions, in which the conversant provides a stronger signal that a contribution has been understood by repeating part or all of the other's contribution. Repetitions are also referred to as "paraphrases" (Traum & Heeman, 1996), "echoing" (Swerts et al, 1998), and "demonstration" (Clark & Schaefer, 1989). Repetitions are often seen when one is conveying complex information, such as when one copies an address or telephone number.

Neither acknowledgments nor repetitions contribute new domain information to the conversation, but they serve to assure the speaker that information has been conveyed successfully. Acknowledgments also play a role in managing turn-taking in mixed-initiative dialogue; although acknowledgments may preface a new contribution by the same speaker (Novick & Sutton, 1994), often they occur alone as a single-phrase turn that appears to serve the purpose of explicitly declining an opportunity to take a turn (Sacks et al, 1974).

Acknowledgments and repetitions are ubiquitous in many types of human-human conversation. In a corpus of problem-solving spoken dialogues, for example, Traum and Heeman (1996) found that 51% of turns began with or consisted of an explicit acknowledgment. Given this, one would expect that acknowledgments should be modeled in dialogue models for spoken-language systems, and

indeed some research models are beginning to incorporate acknowledgments, e.g., Kita et al (1996), Aist, (1998), Iwase & Ward (1998).

Typical human-computer dialogue models are structured in ways that suppress the use of acknowledgments. In many systems turn-taking is completely controlled by one conversant, e.g., the user responds to system prompts, which tends to eliminate the need for acknowledgments as a turn-taking mechanism. In other systems, the use of barge-in defeats the common interpretation of an acknowledgment; if the user speaks, the system contribution is cut off before the user utterance is interpreted. If that utterance was intended to signal that the contribution should continue, the effect is exactly the opposite of the one desired.

Thus, current design practices both discourage and render meaningless the standard uses of acknowledgments. If these impediments were removed, would people choose to use acknowledgments when interacting with a computer interface?

2 Experiment

This study was designed as a pilot to our larger investigation into the effects of incorporating acknowledgement behavior in dialogue models for spoken-language interfaces. Before we attempted to compare interfaces with and without acknowledgement behavior, we wanted to understand whether people are willing to use this sort of meta-dialogue behavior when interacting with a computer.

2.1 Approach

In this study we hypothesized that subjects will choose to use acknowledgments in human-computer interaction if they are given an interface that provides opportunities for and responds to acknowledgments.

In designing the study, we assumed that it would not immediately occur to subjects that they could use acknowledgments to a computer. At the same time, we did not want to explicitly instruct or require subjects to use acknowledgment behavior, as that would tell us nothing about their preferences. We therefore decided against a comparison/control-group experimental design for this initial study and instead focused on creating a situation in which subjects would have a reason to use

acknowledgments, perhaps even gain an advantage from doing so, while still keeping the behavior optional.

We decided to focus on a somewhat narrow use of acknowledgments. Conversants are especially likely to offer acknowledgments and repetitions when complex information is being presented, especially when the conversant is copying the information. While this is certainly explainable in terms of mutuality of understanding, this particular use of acknowledgment may be viewed from a more mechanical standpoint as regulating the pace at which information is presented. This insight suggested to us that a fruitful task for this study might be one in which the subject is asked to write down verbally-presented information, as when taking messages over the telephone.

2.2 Task

We selected the domain of telephone interface to email and designed a task in which subjects were asked to transcribe items of information from the messages. Writing is slow in comparison to speaking, so we anticipated that subjects would require a slower pace of information presentation when they were writing. The messages included information not asked for on the question list to simulate "uninteresting" material that the subject would want to move through at a faster pace. In this way we hoped to motivate subjects to try to control the pace at which information was presented.

The email was presented in segments roughly corresponding to a long phrase. After each segment, the system paused to give the subject time to make notes. If the subject said nothing, the system would continue by presenting the next message segment. Subjects could accept—and perhaps make use of—this delay, or they could reduce it by acknowledging the contribution, e.g., "okay," or by commanding the system to continue, e.g., "go on." The system signalled the possibility of controlling the delay by prompting the subject "Are you ready to go on?" after the first pause. This prompting was repeated for every third pause in which the subject said nothing. In this way we hoped to suggest to the subjects that they could control the wait time if desired without explicitly telling them how to do so.

On the surface, there is no functional difference in system behavior between a subject's use of

a command to move the system onward (e.g., "go on," "next", "continue") and the use of an acknowledgment ("okay," "uh-huh", or a repetition). In either case, the system responds by presenting the next message segment, and in fact it eventually presents the next segment even if the subject says nothing at all. Thus, the design allows the subject to choose freely between accepting the system's pace (system initiative), or commanding the system to continue (user initiative), or acknowledging the presentations in a fashion more typical of mixed-initiative human conversation. In this way, we hoped to understand how the subject preferred to interact with the computer.

2.3 Subjects

Subjects were told that the study's purpose was to assess the understandability and usability of the interface, and that their task was to find the answers to a list of questions. They were given no instructions in the use of the program beyond the information that they could talk to it using normal, everyday speech.

The 14 volunteers were native speakers of North American English, and most were staff at a research university. Ten were female, four were male. Ages ranged from 13 to 57. All used computers, typically office software and games, but none had significant programming experience. Each session lasted about 45 minutes total, and each subject was paid $10.00. One subject declined payment.

2.4 Interface

As mentioned earlier, one difficulty with recognizing acknowledgements in spoken-language interfaces is that the use of barge-in tends to defeat the purpose of acknowledgments when they occur in overlapped speech. We used a Wizard of Oz protocol as a simple way to allow the system to respond to such utterances and to provide robustness in handling repetitions.

The wizard's interface was constructed using the Rapid Application Developer in the Center for Spoken Language Understanding Toolkit (Sutton, et al, 1998). A simple button panel allowed the wizard to select the appropriate response from the actions supported by the application. The application functionality was deliberately limited to suggest realistic abilities for a current spoken-language interface. Using messages pre-recorded

in a synthesized voice, the wizard was able to direct the system to:

- Read a list of all messages.
- Begin reading a particular message.
- Read the next message segment.
- Repeat the current message segment.
- Repeat the previous message segment.
- Ask the subject whether the program should continue reading the current message.
- Ask the subject to what to do next.
- End the program.
- Play one of several error and help messages.

The texts of the email messages were presented in phrases of varying lengths, with each phrase followed by a pause of about five seconds. Preliminary tests showed that the combined response time of the wizard and the interface was between one and two seconds, and that pauses of less than five seconds were not obviously different from the normal pace of system response. Five seconds is a long response time, uncomfortably so for human-human conversation, so we hoped that this lengthy pause would encourage the subjects to take the initiative in controlling the pace of the interaction.

The messages were divided into segments by hand. The divisions were intended to simulate a phrase-level presentation, although some short phrases were combined to make the presentation less choppy. An example of one message and its division into phrases may be seen in Figure 1.

Synthesized speech from the Festival speech synthesizer (Taylor, et al, 1998) was used throughout the interface. The message texts were presented in a synthesized male voice, while the control portions of the interface used a synthesized female voice. Default pronunciations were used except when the default was incorrect, e.g., "read" defaulted to the past-tense pronunciation in all contexts. Also, there was minor use of the SABLE markup language (Wouters, et al, 1999) to flatten the pitch range at the end of phrases in list items; the intent was to suggest the prosody of list continuation rather than the default sentence-final drop.

Message six is from Jo at teleport dot com, about, please stop by store on your way home.
I'm going to be late getting home tonight, so would you please stop by the store on your way home?
We need milk,
eggs,
a bunch of spinach,
fresh ginger,
green onions,
maple syrup,
a pound of coos-coos,
mild curry powder,
a pound of coffee,
and a package of seventy five watt light bulbs.
Thanks! See you tonight.

Figure 1. Text of a sample message.

The subject's list of questions included "What items are you supposed to pick up at the store?"

To improve the understandability, both voices were slowed slightly to 90% of the default speaking rate.

2.5 Measures

The central question to be answered is: will the subject use acknowledgments in interacting with the program? A subject can show one of several patterns of response:

- The subject may make no attempt to control the pacing of the interface, instead allowing the interaction to proceed via time-outs.
- The subject may use only commands to control the pacing.
- The subject may use only acknowledgments to control the pacing.
- The subject may use a mixture of commands and acknowledgments.

The determination as to whether a particular utterance constituted an acknowledgment or a command was based primarily on word choice and dialogue context; this approach is consistent with definitions of this behavior, e.g., Chu-Carroll and Brown (1997). For example, "yes" in the context of a system inform (a segment of an email message) was considered an acknowledgment, but "yes" in the context of a system question was not. The words "okay," "uh-huh," and "yes" (immediately following an inform) were taken as evidence of acknowledgments, and phrases such as "go on," "continue," "next" following an inform were taken as evidence of commands. The interpretation was confirmed during the post-experiment interview by questioning the subjects about their word choice.

2.6 Post-Experiment Interview

A post-experiment interview was conducted to gather subject feedback and to answer subjects' questions. The experimenter took notes and thus could have introduced bias in the record of responses. No tape recording was made.

The subject was first invited to comment on the interface and the interaction in an open-ended fashion. When the subject had finished, the experimenter asked several specific questions to assess their understanding of the interface functionality. During this time, the experimenter reminded the subjects of the words that they had used most frequently to prompt the system to continue during pauses and asked the subjects why they had selected those words.

Finally, the experimenter explained the true purpose and hypothesis of the experiment, verified that the subject was unaware that they had been interacting with a Wizard-of-Oz interface, and asked the subject to comment on the notion of using acknowledgments when interacting with a computer. The responses to this question, especially, must be assumed to be somewhat optimistic, as it is likely that at least some subjects would be reluctant to disagree with the experimenter.

3 Results

Results are summarized in Table 1.Because the subject pool was not balanced for gender, results for male and female subjects are reported separately. Due to the small number of male subjects in

283

Table 1. Summary of Acknowledgment Behavior

Behavior	Subjects		Total (14)
	Female 10 subjects	Male. 4 subjects	
Used acknowledgment/repetition at least once	4 (40%)	4 (100%)	8 (57%)
Used acknowledgment/repetition more than command	3 (30%)	1 (25%)	4 (29%)
Used acknowledgment but no commands	1 (10%)	0	1 (7%)
Described acknowledgment to computer as strange	2 (20%)	0	2 (14%)

this pilot study, no tests of statistical significance of differences in the rates of acknowledgment behavior were made.

Eight of the fourteen subjects used an acknowledgment or repetition at least once, and four used acknowledgment/repetitions more frequently than they used commands. Only one subject used acknowledgments exclusively, while five subjects never used acknowledgments. No subject relied exclusively on time-outs to allow the system to proceed at its own pace, although one subject did use that as her predominant method (42 times, while using acknowledgments only six times and commands three times). Only one subject used repetition, and he reported during the interview that he was unaware of having done so.

It is interesting to note that while all of the male subjects in this sample exhibited acknowledgment behavior at least once, only one preferred acknowledgment over command. One of the male subjects used acknowledgments only three times, in all cases as prefaces to commands. Conversely, although a lower percentage of women used acknowledgments (40%), a higher percentage of them (30%) used acknowledgments in preference to commands. Because of the small numbers of subjects, however, we do not conclude that these differences are significant.

During the post-experiment interview, two subjects (both female) described the idea of using acknowledgments to the computer as strange and stated that they didn't feel that they would do this unless directed to—and even then, they would regard it as simply an alternate command. Two other subjects, both females who had used acknowledgments 2-6 times during the task, each reported that she had felt silly when she had caught

herself saying "please" and "okay" to a computer but had been pleased when it had worked. The remainder of the subjects either expressed no strong opinion (two, both female) or expressed a positive attitude toward being able to use acknowledgments when interacting with a computer. Two subjects who had not used acknowledgments commented that they would probably be more likely to use human-like conversation if the synthesized voice were more human-like.

Again, this report of the subjects' attitudes should be interpreted with caution; at this point in the interview they knew the experimenter's hypothesis and so may have been reluctant to disagree.

3.1 Other Dialogue Behaviors

Although we had not formed any hypothesis about other dialogue behaviors, we noticed several interesting dialogue behaviors that we had not anticipated.

We were surprised at the number of subjects who exhibited politeness behavior toward the interface, either saying "please" when issuing commands to the computer or responding to the program's "good-bye" at the end of the session. One subject used "please" throughout the interaction, but a more common pattern was to use "please" at the beginning of the session and to drop the behavior as the interface became more familiar. Politeness did not seem to be strongly associated with a willingness to use acknowledgments, however; four of the nine subjects who exhibited politeness did not use any acknowledgments in their interaction.

Despite the deliberately-artificial interface, several subjects responded at least once to the message content as if they were talking to the message

284

System: I could come to your office now or at any of the following times. one thirty
SUBJECT: continue
System: three o clock
SUBJECT: continue
System: or five fifteen
SUBJECT: continue
System: thank you. I look forward to your prompt reply
SUBJECT: thank you- uh ((laugh)) continue

Figure 2. Excerpt of transcript.
Subject thanks the interface. The system is reading the text of one of the messages.

sender. In the excerpt shown in Figure 2., for example, the subject replied "Thank you" to the message text's "thank you." This did not appear to be a matter of misunderstanding the capabilities of the interface; the subject later reported that despite the synthesized voices she had briefly forgotten that she wasn't talking to her secretary.

Three subjects also made one or more meta-comments, e.g., "ah, there it is" when finding a particular piece of information. These may have been at least partially an artifact of the "treasure hunt" nature of the task. When questioned in the post-experiment interview, subjects didn't seem aware that they'd made these comments. All but one of these instances were followed immediately by a command, so the wizard responded to the command and ignored the meta-comment. The one stand-alone meta-comment was treated as an unrecognized command (an error message was played).

4 Discussion

Subjects were provided with three methods for controlling the pace at which information was presented: silence, command, or acknowledgment/ repetition. The majority of the subjects used commands more than they used acknowledgments, but over one half used an acknowledgment or repetition at least once during their interaction and nearly 30% used acknowledgments in preference to commands. This occurred despite the fact that subjects were given no reason to think that this behavior would be effective: the interface was deliberately limited in functionality, and voice synthesis was used instead of recorded voice to emphasize the artificial nature of the interaction.

Furthermore, the interface did not offer acknowledgments to the subjects, and the subjects were given no instructions suggesting that the interface understood acknowledgments. In fact two subjects who did use acknowledgments expressed surprise that they had worked, and two who had not used acknowledgments reported that they would probably have used them if they had known it would work.

It is interesting to consider these results in light of those reported by Okato et al (1998). They describe a Japanese-language Wizard-of-Oz study in which the subjects were given some instruction on using the system, and in which the system both presented and accepted back-channel feedback. They found that even when the interface offered back channels itself the rate of subject back-channels was somewhat lower in human-computer interaction than in comparable human-human conversation. This makes the fact that our interface elicited acknowledgments without offering them even more encouraging. Clearly, some people are willing to utilize this human conversational convention in human-computer dialogue. Our post-experiment interviews suggest, however, that some people will find the use of acknowledgements strange or uncomfortable in human-computer interaction. While self-reports of attitudes toward hypothetical situations must be treated with some caution, it seems reasonable to assume that even when such interfaces become available there will be users who will prefer to interact with computers using commands.

Will attitudes and conversational behavior change as people gain experience with more advanced spoken-language interfaces? Despite the

285

relatively short duration of this test—most subjects completed the task itself in 15-20 minutes—some changes in behavior could be observed over the course of the dialogue. In particular, politeness behaviors were likely to be seen early in the dialogues and then diminish as the subjects became more comfortable with their interaction. We speculate that the use of politeness words did not reflect a strong underlying politeness toward the computer so much as a falling back on human conventions when faced with an unfamiliar dialogue situation. One subject who had used "please" 21 times during the interaction, for example, simply hung up without warning when she had finished. This contrasts, however, with the findings of Nass et al (1999) that people do offer socially-desirable behavior to computers.

Would a better voice increase the incidence of acknowledgment behavior? Several subjects thought it would, and even with the current synthesized voices we saw several examples of subjects seemingly forgetting briefly that they were not talking to a human. We plan to explore this question in future work.

4.1 Conclusions and Future Work

We conducted a preliminary study to examine the willingness of subjects to use a particular dialogue act, acknowledgment, in human-computer interaction. Although the number of subjects was small, we saw that about half of our subjects used acknowledgements or repetition at least occasionally to control the pace at which information was presented, and about 29% used acknowledgments more frequently than they used commands for that purpose.

Our immediate plans include extending this study to a larger and gender-balanced group of subjects so that we can draw firmer quantitative conclusions about the percentage of people who are likely to prefer this style of interaction. In particular, we cannot conclude from the current study's small sample how strong the preference for using acknowledgment might be, especially among male subjects. Also, in our current study the subject achieved no functional benefit in using acknowledgments. With better estimates of subject preferences, we can then proceed to our larger goal of comparing the usefulness and user acceptability of spoken language dialogue models with and without

acknowledgment behavior (c.f. Walker, 1993). We also plan to explore the effect of the quality of the synthesized voice on the incidence of acknowledgment behavior.

References

Gregory Aist. 1998. Expanding a Time-Sensitive Conversational Architecture for Turn-Taking to Handle Content-Driven Interruption. In *Proceedings of ICSLP 98 Fifth International Conference on Spoken Language Processing*, pages 413-417.

Sara Basson, Stephen Springer, Cynthia Fong, Hong Leung, Ed Man, Michele Olson, John Pitrelli, Ranvir Singh, and Suk Wong. 1996. User Participation and Compliance in Speech Automated Telecommunications Applications. In *Proceedings of ICSLP 96 Fourth International Conference on Spoken Language Processing*, pages 1676-1679.

Susan E. Brennan. 1991. Conversation With and Through Computers. *User Modeling and User-Adapted Interaction.* 1:67-86.

Jennifer Chu-Carroll and Michael K. Brown. 1997. Tracking Initiative in Collaborative Dialogue Interactions. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pages 262-270.

Herbert H. Clark and Edward F. Schaefer. 1989. Contributing to Discourse. *Cognitive Science*, 13:259-294.

Ronald A. Cole, David G. Novick, Pieter J. E. Vermeulen, Stephen Sutton, Mark Fanty, L. F. A. Wessels, Jacques H. de Villiers, Johan Schalkwyk, Brian Hansen and Daniel Burnett. 1997. Experiments with a Spoken Dialogue System for Taking the U.S. Census. *Speech Communications*, Vol. 23.

Peter A. Heeman, Michael Johnston, Justin Denney and Edward Kaiser. 1998. Beyond Structured Dialogues: Factoring out Grounding. In

Proceedings of ICSLP 98 Fifth International Conference on Spoken Language Processing, pages 863-867.

Tatsuya Iwase, and Nigel Ward. 1998. Pacing Spoken Directions to Suit the Listener. In *Proceedings of ICSLP 98 Fifth International Conference on Spoken Language Processing*, Vol. 4, page 1203-1207.

K. Kita, Y. Fukui, M. Jagata and T. Morimoto. 1996. Automatic Acquisition of Probabilistic Dialogue Models. In *Proceedings of ICSLP 96 Fourth International Conference on Spoken Language Processing*, pages 196-199.

Clifford Nass, Youngme Moon and Paul Carney. 1999. Are Respondents Polite to Computers? Social Desirability and Direct Responses to Computers. *Journal of Applied Social Psychology*, 29:5, pages 1093-1110.

H. Noguchi and Yasuharu Den. 1998. Prosody-Based Detection of the Context of Backchannel Responses. In *Proceedings of ICSLP 98 Fifth International Conference on Spoken Language Processing*, Vol. 2, pages 487-490.

David G. Novick and S. Sutton. 1994. An Empirical Model of Acknowledgment for Spoken-Language Systems. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pages 96-101.

Yohei Okato, Keiji Kato, Mikio Yamamoto and Shuichi Itahashi. 1998. System-User Interaction and Response Strategy in Spoken Dialogue System. *Proceedings of ICSLP 98 Fifth International Conference on Spoken Language Processing*, Vol. 2, pages 495-498.

S. L. Oviatt and P. R. Cohen and M. Wang. 1994. Toward Interface Design for Human Language Technology: Modality and Structure as Determinants of Linguistic Complexity. In *Speech Communication*, 15:3-4, pages 283-300.

H. Sacks, E. Schegloff and G. Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking in Conversation. *Language*, 50:696-735.

Stephen Sutton, Ronald Cole, Jacques de Villiers, Johan Schalkwyk, Pieter Vermeulen, Mike Macon, Yonghong Yan, Ed Kaiser, Brian Rundle, Khaldoun Shobaki, Paul Hosom, Alex Kain, Johan Wouters, Dominic Massaro and Michael Cohen. 1998. Universal Speech Tools: the CSLU Toolkit. In *Proceedings of the Inter-national Conference on Spoken Language Processing*, pages 3221-3224.

Marc Swerts, Hanae Koiso, Atsushi Shimojima and Yasuhiro Katagiri. 1998. On Different Functions of Repetitive Utterances. In *Proceedings of ICSLP 98 Fifth International Conference on Spoken Language Processing*, Vol. 2, pages 483-487.

Paul Taylor, Alan W. Black and Richard Caley. 1998. The Architecture of the Festival Speech Synthesis System. In The Third ESCA/COCOSDA Workshop on Speech Synthesis, pages 147-151.

David R. Traum and Peter A. Heeman. 1996. Utterance Units and Grounding in Spoken Dialogue. In *Proceedings of ICSLP 96 Fourth International Conference on Spoken Language Processing*, pages 1884-1887.

Marilyn A. Walker. 1993. *Informational Redundancy and Resource Bounds in Dialogue*. Doctoral dissertation, University of Pennsylvania.

Johan Wouters, Brian Rundle and Michael W. Macon. 1996. Authoring Tools for Speech Synthesis using the Sable Markup Standard. In *Proceedings of Eurospeech '99*.

Nigel Ward. 1996. Using Prosodic Clues to Decide When to Produce Back-Channel Utterances. In *Proceedings of ICSLP 96 Fourth International Conference on Spoken Language Processing*, page 1724-1727.