9-1-2003

Toward Building Conversational Spoken-Language Interfaces: Acknowledgment Use in American English and Mexican Spanish

Karen Ward
University of Texas at El Paso, kward@cs.utep.edu

Tasha Hollingsed
University of Texas at El Paso, tasha@cs.utep.edu

Javier A. Aldaz Salmon
University of Texas at El Paso, jaldaz@cs.utep.edu

Comments:
Published in *Proceedings of the Fourth Mexican International Conference on Computer Science (ENC'03)*. 8-12 Sept. 2003. pp. 10-17.

Toward Building Conversational Spoken-Language Interfaces: Acknowledgment Use in American English and Mexican Spanish

Karen Ward, Tasha Hollingsed, Javier A. Aldaz Salmon
The University of Texas at El Paso
{kward,tasha,jaldaz}@cs.utep.edu

Abstract

Should spoken-language interfaces incorporate human discourse phenomena? Acknowledgments, for example, are ubiquitous in human conversation but are 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 design of current spoken-language interfaces? We found that, given a simple spoken-language interface that responded to acknowledgments, over two thirds of subjects used acknowledgments at least once, about the same number that used more traditional commands to control the interface. These results were consistent for both Mexican Spanish and American English versions of the interface, and they suggest that it may be possible to make use of human discourse mechanisms such as acknowledgment to build more flexible spoken-language interfaces.

1. Introduction

Spoken-language interfaces, which allow one to talk to a computer application instead of typing or clicking, are beginning to move from the laboratory to the real world. In addition to offering hands-free access to those who cannot or prefer not to use a traditional keyboard interface, spoken-language systems have gained popularity in telephone-based applications such as airline information systems. The quality of interaction offered by these interfaces, however, is far from that offered by a human operator.

Current-generation interfaces are still relatively fragile. To reduce errors, designers of spoken-language systems create prompts that guide the user toward short, focused, in-vocabulary responses, e.g., [2], [6]. Typically, the system prompts the user through a series of questions to get the information needed to complete the task. One result of this approach is the deliberate suppression of dialogue behaviors that, in human conversation, manage and coordinate the conversation. In essence, the computer is always in charge of the flow of the conversation. While this may be adequate, even appropriate, for limited tasks such as checking for flight delays, we would like to move toward more sophisticated spoken-language interfaces capable of handling more complex tasks.

Before we build such systems, though, we should have a better understanding of what we should be building. To what extent should such interfaces accommodate or even mimic human conversational behavior? Do users want to speak to computers using the casual interaction style that one might use with a friend, or are computers mere tools to be controlled with brusque and business-like voice commands? Work addressing this issue shows mixed results. Many studies have shown that people alter their speaking style when they believe that they are talking to a computers, e.g., [3]. It is not clear, however, whether they do so because they would prefer to talk to computers differently or because they believe that computers cannot understand their regular speech.

In this study we examine people's willingness to use a particular kind of dialogue behavior, acknowledgment, when interacting with a spoken-language computer interface.

1.1. Acknowledgments in human speech

The term "acknowledgment" is from Clark and Schaefer [5], 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 English as "uh-huh" and in Spanish as "ajá." Acknowledgments, also called "back-channels" by some researchers (e.g., [4]), are one of several meta-dialogue behaviors that people use to control the flow of conversation by signalling that they understand what has been said and that they wish the conversation to continue.

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. These often are used when conveying complex information, such as when copying an address or telephone number. Repetitions are also

Proceedings of the Fourth Mexican International Conference on Computer Science (ENC'03)
0-7695-1915-6/03 $17.00 © 2003 IEEE

termed “paraphrases” [17], “echoing” [15], and “demonstration” [5], and some researchers classify repetition as a type of back-channel [8]. In this paper, repetitions are considered a kind of acknowledgment behavior.

Acknowledgments are ubiquitous in human conversation, though the frequency varies by language and culture [8]. In a corpus of American English problem-solving dialogues, for example, Traum and Heeman [17] found that 51% of turns began with or consisted of an explicit acknowledgment.

1.2. Acknowledgment in spoken-language interfaces

Meta-dialogue behaviors such as acknowledgment are of interest to the spoken-language interface community because of their role in managing turn-taking: although acknowledgments may preface a new contribution by the same speaker [11], often they occur alone as a single-phrase turn that appears to serve the purpose of explicitly declining an opportunity to take a turn [13]. If acknowledgment behavior is incorporated in spoken-language systems, it may offer a more fluid and adaptable means of controlling turn-taking and pacing in human-computer interaction.

Although some research systems are beginning to incorporate acknowledgments, e.g., [1], [7], [10], [19], real-world spoken-language interfaces generally don’t allow acknowledgments to serve their turn-taking purpose. Turn-taking is completely controlled by one conversant, usually the system. In many systems, the use of barge-in defeats the common interpretation of an acknowledgment: if the user speaks, the system quits speaking and begins interpreting the user utterance. If that utterance was intended to signal that the system should continue, the effect is exactly the opposite of the one the user intended.

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? Or would they prefer to use commands?

In this study, we confirm and quantify a limited pilot study done in American English using West Coast speakers [18], and we extend the results to Mexican Spanish and American English speakers from the Mexico-USA border region.

2. Experiment

This study is part of a larger investigation of the effects of incorporating more sophisticated models of turn-taking in spoken-language interfaces by both responding to and producing acknowledgments. Before we attempted to compare interfaces with and without acknowledgement behavior, though, we wanted to understand to what extent people are willing to use this sort of dialogue behavior when interacting with a computer.

2.1. Approach

In designing the study, we assumed that it would not immediately occur to subjects that they could use acknowledgments to a computer; as discussed in the Introduction, interfaces that allow for acknowledgment behavior are largely confined to research labs. 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 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.

As in our pilot study, we focused on a somewhat narrow use of acknowledgments. Conversants are likely to offer acknowledgments and repetitions when complex information is being presented, especially when copying the information. While this is certainly explainable in terms of ensuring 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 might be a 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 E-mail. Subjects were told that the computer system would read E-mail messages to them over the telephone and that their task was to locate and transcribe particular items of information contained in the messages, e.g., “How do you get to the coffee house?” 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 both “interesting” information that was to be copied and “uninteresting” information that was not, so that subjects would want to move through the “uninteresting” material more

quickly. In this way we hoped to motivate subjects to try to control the pace at which information was presented.

The E-mail was presented in segments roughly corresponding to a long phrase (Figure 1). 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 reduce this delay by acknowledging the contribution, e.g., "okay" or by commanding the system to continue, e.g., "go on" or "continuar." The system signalled the possibility of controlling the delay by asking the subject the question "Are you ready to go on" or "Estas listo(a) para continuar" 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 without explicitly telling them so. Also, we made the tone of the messages deliberately informal to suggest E-mail that students might exchange with friends; we hoped that this style would further license the use of acknowledgment behavior.

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. 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, or commanding the system to continue, or acknowledging the presentations in a fashion more typical of 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 were to talk to it using normal, everyday speech.

We tested a total of 40 subjects, balanced for gender and language. Subjects were solicited from the University of Texas at El Paso campus. They ranged in age from 18 to 44 with most being between 20 and 25. Each subject was paid $10.00 for participating in the study.

Many subjects spoke both Spanish and English, so subjects were assigned to the English or Spanish-language conditions based on their reports of their first language and the language they spoke at home. Most of the Spanish subjects reported growing up in Juarez, Chihuahua or other northern Mexico cities. Most of the English subjects were from the west or southwest regions of the United States.

2.4. Interface

As mentioned earlier, one difficulty with using acknowledgements in spoken-language interfaces is that current system designs tend to discourage their use, especially when they occur in overlapped speech. We used a Wizard of Oz protocol as a way to allow the system to respond to such utterances and to provide robustness in handling repetitions. A Wizard of Oz protocol is one in which the subject is presented with what purports to be a fully-computerized system; in reality, a human "wizard" working behind the scenes provides at least part of the functionality, typically functioning as a highly robust and accurate speech understanding module. The "wizard" listens to the subject utterance and selects the appropriate system response.

The wizard's interface was constructed using the Rapid Application Developer in the Center for Spoken Language Understanding Toolkit [14]. 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. The wizard could 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 what to do next.
- End the program.
- Play one of several error and help messages.

The texts of the E-mail messages were presented in segments of varying lengths, with each segment 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 conversation, so we hoped that this lengthy pause would encourage the subjects to take the initiative in controlling the pace of the interaction.

Synthesized speech from the Festival speech synthesizer [16] was used throughout the interface. The message texts were presented in a synthesized male voice, while the control portions of the interface used a synthe-

- El mensaje numero 5 es de rosario@utep.edu. Sobre juntarse después de clase
- Miguel y yo nos vamos a juntar hoy en la cafetería después de clases
- Si quieres acompañarnos nos vemos en la cafetería, a las 5:00 p.m.
- Las direcciones son
- Saliendo de la escuela, te vas hacia arriba por la Avenida Universidad
- das vuelta a la izquierda en la calle Mesa
- cuentas 5 cuadras
- y das vuelta a la derecha en la calle Cincinnati.
- La cafetería estará a tu derecha
- Esperamos verte ahí

- Message 5 is from rosario@utep.edu about meeting after class
- Mike and I are meeting at the coffeehouse after class today
- If you want to meet us there at 5:30

- The directions coming from school are:
- Go up University Avenue

- Turn left on Mesa
- Go down Mesa for five blocks
- Turn right on Cincinnati

- The coffeehouse will be on the left side.
- Hope to see you there

Figure 1. One of the messages, Spanish and English versions, showing the message segments

sized female voice. Default pronunciations were used except when the default was incorrect, e.g., in English the word "read" defaulted to the past-tense pronunciation in all contexts. To improve the understandability, voices were slowed slightly to 90% of the default speaking rate.

2.5. Measures

Subjects were able to use one of three strategies in response to the system's pauses: they could wait, making no attempt to control the pacing of the information presentation; they could use commands to control the pacing, or they could use acknowledgments to control the pacing.

In comparing the strategies used by the subjects, the dependent variable was the number of times each strategy was used to control the pacing of the interface. The total number of turns varied between subjects because some subjects listened to each message only once while others went through messages multiple times. We therefore normalized the counts by dividing the number of times each strategy was used by the number of turns where the subject had had an choice of strategies.

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 acknowledgment, e.g., [4]. For example, "sí" or "yes" in the context of a system inform (presentation of a segment of an E-mail message) was considered an acknowledgment, but in the context of a system question such as "Are you ready to go on?" the same words would be considered an answer to a question. Immediately following a system inform, the words "yes," "sí," "uh-huh," "ajá," and "okay" or a repetition of part or all of the system inform were considered acknowledgments. Phrases such as "go on," "continue," "next," "continuar," or "siguiente" following an inform were considered commands. The interpretation was confirmed during the post-experiment interview by questioning the subjects about their word choice. Transcriptions and categorizations of the subject utterances were checked by a second person for accuracy.

Some subjects (2 in the Spanish-language condition and 10 in the English-language condition) combined acknowledgments and commands in a single utterance, e.g., "okay, go on." These are not included in the figures reported in Table 1 nor in the analysis because they could not be clearly assigned to any category. Most subjects did this only once (2 subjects in the Spanish-language condition and 5 in the English-language condition), and only one speaker (English) produced as many as five combined-type responses.

2.6. Post-Experiment Interview

A post-experiment debriefing was conducted to gather subject feedback, to explain the true purpose of the experiment, and to answer subjects' questions. This interview was taped and the experimenter took notes. Data from subjects who had realized that they were interacting with a human instead of a completely-automated system were excluded from the study because of the well-verified tendency for people to speak differently when they believe that they are speaking with a human instead of a computer [3].

3. Results

Three categories of results are discussed in this section: the quantitative results of subjects' choices of strategy in controlling the interface, qualitative results, which include user judgements of the system's usability and self-reports of their attitudes toward the interface, and other dialogue behaviors.

3.1. Quantitative Results

The main questions to be answered were

- Which strategy will subjects prefer to use to control the pacing of information presentation?
- Will we see a significant difference between users of the English-language interface and those of the Spanish-language interface?
- Will we see a significant difference between male and female speakers?

The most commonly-used strategy was to wait, allowing the interface to pace itself. Over 70% of all subjects used this strategy more frequently than any other strategy. Most of these subjects used at least one command or acknowledgment, but five did not use any strategy other than waiting. Three subjects used only one or two acknowledgments, and they seemed to be unaware that they had done so. Of the 40 subjects, then, 20% did not discover that they were able to control the system pacing (confirmed in the debriefing).

An interface error with 12 subjects in the English-language condition may have made it harder for those subjects to discover that the pacing could be controlled: the "Are you ready to go on?" prompt played less frequently than it should have. We compared the strategies used by those 12 subjects with the strategies used by the other English-language subjects. There was not a significant difference between strategies used by the two groups ($p = 0.67$, two-tailed t-test on the percentage of turns that the subject waited).

Commands and acknowledgments were preferred by only 13% and 15%, respectively, of all subjects. Seventy-five percent of the subjects used a command at least once, and 68% used an acknowledgement at least once. Nine subjects seemed comfortable with both commands and acknowledgments, using at least five examples of each.

When acknowledgments were used, the most common word choice was "okay" (both languages). When commands were used, the most common word choices were "go on" in English, and "continuar" in Spanish.

Comparing strategies between subjects using the Spanish-language interface and the English-language interface, we found a suggestion of a trend that, however, did not rise to the level of significance: $p = 0.16$ on a two-tailed t-test comparison of the turns that the subject controlled the interface using acknowledgment. Because the sample size is small and may not be normally distributed, we checked the finding using the non-parametric Mann-Whitney *U* test. The *U* value found was 165, which exceeds the critical value of 127. We conclude that the differences are not significant.

The figures in Table 1 suggests a difference between the strategy preferred by female Spanish-language subjects, 90% of whom preferred waiting, and that preferred by female English-language subjects, only 60% of whom preferred waiting. A two-tailed t-test comparison on the percentage turns that the subjects controlled the interface with waiting yielded $p=.06$, suggestive but not significant. The Mann-Whitney *U* value was 29, which exceeds the critical value of 23 and the hypothesis is rejected.

In comparing the strategies of male and female speakers, no significant differences were found.

3.2. Qualitative results

During the debriefing, we asked subjects to comment on the notion of using acknowledgments instead of commands to control a computer. Most subjects (16 subjects in the Spanish-language conditions, 14 in the English-language condition, 75% total) reported that they were comfortable with the idea of speaking to a computer as if it were a person. A few subjects (4 Spanish-language and 2 English-language, 15% total) reacted negatively to the idea; one subject termed the notion "weird." Other subjects reported no opinion.

It is interesting to note that subjects' expressed attitudes toward using words like "okay" and "uh-huh" to a computer did not always match their behavior. For example, one of the English-language subjects, a female, reported that she perceived that she had responded to the interface "like a machine," even though she had used relatively few commands (five occurrences of "next") compared to acknowledgments (77 occurrences of "okay" and three occurrences of "mm-hmm"). Another subject, also an English-speaking female, reported that she would "feel stupid having a conversation with one," although she used roughly equal numbers of the command "go on" and the acknowledgment "okay." Two other subjects reported surprise that they had used "okay" and "mm-hmm" to the computer.

3.3. Other dialogue behaviors

As we had seen in our previous study [18], some subjects made use of politeness behaviors such as

Table 1. Preferred strategies for controlling the pace of the system presentations

Subjects controlled the interaction using	Spanish		English		Total (40 subjects)
	Female (10 subjects)	Male (10 subjects)	Female (10 subjects)	Male (10 subjects)	
Acknowledgments, e.g., sí, okay					
Used at least once	6 (60%)	6 (60%)	8 (80%)	7 (70%)	27 (68%)
Used preferentially	0	2 (20%)	3 (30%)	1 (10%)	6 (15%)
Commands, e.g., go on, continuar					
Used at least once	8 (80%)	5 (50%)	8 (80%)	9 (90%)	30 (75%)
Used preferentially	1 (10%)	2 (20%)	1 (10%)	1 (10%)	5 (13%)
Waiting					
Used at least once	10 (100%)	10 (100%)	10 (100%)	10 (100%)	40 (100%)
Used preferentially	9 (90%)	6 (60%)	6 (60%)	8 (80%)	29 (73%)

"please" or "por favor" when issuing commands; we saw this with two males and two females in Spanish and with four females and three males in English. Politeness behaviors did not seem to be strongly associated with a willingness to use acknowledgments, however; only one of the "polite" English-speakers, a female, and one of the "polite" Spanish-speakers, a male, employed acknowledgments in preference to commands or waiting. One male subject used "perdón" to request that the system repeat a message segment; this subject used politeness behaviors more than any other subject (32 occurrences).

One subject, a Spanish-speaking male, played with the system initially. He offered responses like "Siiiiiiiii" (drawn out like a cheer) and "Si, señor," "Si, señorita" to the system voices in an obviously joking manner. During the debriefing he confirmed that he had been playing. In the latter part of his session he became more task-oriented, although he continued to use acknowledgments in preference to commands. In the debriefing, he stated that he would prefer to talk to a computer using such informal behaviors if the computer would understand them.

4. Conclusions

Subjects were provided with three methods for controlling the pace at which information was presented: waiting (accepting the system's slow default pace), or using commands or acknowledgments to move the system forward more rapidly. About 20% of subjects did not discover that they could control the pacing of the system, but many of those that did used acknowledgments as well as commands to control the system.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 such words.

In a pilot study [18], we had found that about half of the subjects offered acknowledgments at least once and nearly 30% used them extensively during the interaction. The subjects in the current study offered acknowledgments at least once at a higher rate (68%), but used them extensively at a lower rate (15%). The pilot study, however, was not balanced for gender and those subjects were from a fairly-uniform linguistic background (American English, from the Pacific Northwest region of the United States). We regard the findings of the current study as being more general.

It is interesting to consider these results in light of those reported by Okato [12]. 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, 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 strategy in human-computer dialogue.

We believe—and one subject confirmed—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. This contrasts, however, with the

findings of Nass [9] that people do offer socially-desirable behavior to computers. Perhaps one of the English-language subjects expressed this ambivalence well when she said "I guess when you like a machine you'll treat it more like a person, but when you don't like it you'll just treat it as a machine."

Would a more human-like system voice increase the incidence of acknowledgment behavior? Many subjects (11 in the English-language condition, 9 in the Spanish) thought that it would. We are continuing the study using recorded human voices in place of the synthesized voices with the hypothesis that we will see a higher rate of acknowledgment behavior.

A minority of subjects (10%) expressed reservations about using human-like conversational interaction with a computer. This suggests that some users, at least, would prefer to interact with computers as machines and tools instead of as "people." An attempt to make interfaces too "human-like" may backfire with some users. We plan to probe this issue further in future work.

Our larger goal is to compare the usefulness and user acceptability of spoken language dialogue interfaces that employ meta-dialogue control mechanisms such as acknowledgments. We hypothesize that such interfaces will offer a more fluid and "human-like" interaction that will improve the usability of spoken-language interfaces.

Acknowledgments

This work was partially supported by a gift from Microsoft Corporation and by the National Science Foundation's Model Institutions for Excellence Initiative EEC 9550502. The authors thank the anonymous reviewers for their helpful suggestions for improving this paper, and thank Christian Servin, Tyler Smith and Pauline Williamson for their assistance with the study.

References

[1] G. Aist, "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*, 1998 pp. 413-417.

[2] S. Basson, S. Springer, C. Fong, H. Leung, E. Man, M. Olson, J. Pitrelli, R. Singh, and S. Wong, "User Participation and Compliance in Speech Automated Telecommunications Applications," in *Proceedings of ICSLP 96 Fourth International Conference on Spoken Language Processing*, 1996, pp. 1676-1679.

[3] S.E. Brennan, "Conversation With and Through Computers," *User Modeling and User- Adapted Interaction*, 1991, 1:67-86.

[4] J. Chu-Carroll and M.K. Brown, "Tracking Initiative in Collaborative Dialogue Interactions," in *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, 1997, pp. 262-270.

[5] H.H. Clark and E.F. Schaefer, "Contributing to Discourse," *Cognitive Science*, 1989, 13:259-294.

[6] R.A. Cole, D.G. Novick, P.J.E. Vermeulen, S. Sutton, M. Fanty, L.F.A. Wessels, J.H. de Villiers, J. Schalkwyk, B. Hansen and D. Burnett, "Experiments with a Spoken Dialogue System for Taking the U.S. Census," *Speech Communication*s, Vol. 23, 1997.

[7] T. Iwase and N. Ward, "Pacing Spoken Directions to Suit the Listener," in *Proceedings of ICSLP 98 Fifth International Conference on Spoken Language Processing*, 1998, Vol. 4, pp. 1203-1207.

[8] B. Heinz, "Backchannel responses: a strategic responses in blingual speakers' conversations," *Journal of Pragmatics*, 2003, Vol. 35, pp. 1113-1142.

[9] C. Nass, Y. Moon, and P. Carney, "Are Respondents Polite to Computers? Social Desirability and Direct Responses to Computers," *Journal of Applied Social Psychology*, 1999, 29:5, pp. 1093-1110.

[10] H. Noguchi and Y. Den, "Prosody- Based Detection of the Context of Backchannel Responses," in *Proceedings of ICSLP 98 Fifth International Conference on Spoken Language Processing*, 1998, Vol. 2, pp. 487-490.

[11] D.G. Novick and S. Sutton, "An Empirical Model of Acknowledgment for Spoken-Language Systems," in *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, 1994, pp. 96-101.

[12] Y. Okato, K. Kato, M. Yamamoto and S. Itahashi, "System-User Interaction and Response Strategy in Spoken Dialogue System," in *Proceedings of ICSLP 98 Fifth International Conference on Spoken Language Processing*, 1998, Vol. 2, pp. 495-498.

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

[14] S. Sutton, R. Cole, J. de Villiers, J. Schalkwyk, P. Vermeulen, M. Macon, Y. Yan, E.Kaiser, B. Rundle, K. Shobaki, P. Hosom, A. Kain, J. Wouters, D. Massaro and M. Cohen,"Universal Speech Tools: the CSLU Toolkit," in *Proceedings of the International Conference on Spoken Language Processing*, 1998, pp. 3221-3224.

[15] M. Swerts, H. Koiso, A. Shimojima and Y. Katagiri, "On Different Functions of Repetitive Utterances," in *Proceedings of ICSLP 98 Fifth International Conference on Spoken Language Processing*, 1998, Vol. 2, pp. 483-487.

[16] P. Taylor, A.W. Black and R. Caley, "The Architecture of the Festival Speech Synthesis System," in *The Third ESCA/ COCOSDA Workshop on Speech Synthesis*, 1998, pp. 147-151.

[17] D.R. Traum and P.A. Heeman, "Utterance Units and Grounding in Spoken Dialogue," in *Proceedings of ICSLP 96 Fourth International Conference on Spoken Language Processing*, 1996, pp. 1884-1887.

[18] K. Ward and P.A. Heeman, "Acknowledgments in Human-Computer Interaction," in *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL 2000)*, April 29-May 4, 2000, pp. 280-287.

[19] N. Ward, "Using Prosodic Clues to Decide When to Produce Back-Channel Utterances," in *Proceedings of ICSLP 96 Fourth International Conference on Spoken Language Processing*, 1996, pp. 1724-1727.