An Assistive Conversation Skills Training System for Caregivers of Persons with Alzheimer's Disease

Nancy Green, William Lawton
Department of Mathematical Sciences
University of North Carolina at Greensboro
Greensboro, NC 27402 USA
nlgreen@uncg.edu

Boyd Davis
Department of English
University of North Carolina at Charlotte
Charlotte, NC 28223 USA
bdavis@email.uncc.edu

Abstract

We describe research in progress on a dialogue system that simulates social conversation between animated characters representing a person with Alzheimer's Disease and his caregiver. The goal of the research is to enable caregivers to practice techniques to improve their assistive conversational skills. These skills are expected to reduce caregiver stress and to improve the quality of life of persons with AD.

Introduction

Professional and family caregivers often find it difficult to engage in social conversation with a person with Alzheimer's Disease (AD). Improving the caregiver's skill in conversation with persons with AD is expected to reduce caregiver stress and to improve the quality of life of persons with AD. While much research on the impact of AD has focused on communication deficits, researchers have begun to study assistive conversational techniques (Moore and Davis 2002). The goal of our research is to develop an artificial intelligence-based system to enable caregivers to learn these techniques by having a social conversation with a simulated speaker with AD. We hope that through use of the system, the user will acquire techniques that transfer to conversational interaction with real persons with AD. A study of the effectiveness of an interactive multimedia computer training program for professional caregivers of persons with dementia showed that the program had a significant positive effect on caregivers' confidence and knowledge of conversational skills, and that this medium was more successful than a video-taped lecture-based training program (Irvine et al. 2003). Our working assumption is that, just as CD-ROM has advantages over videotape, use of AI in our training system will be advantageous. In this paper we describe an experimental version of the system that we are developing currently.

System Requirements

The system's interface must be able to portray two animated characters engaging in spoken conversation, one character representing a person with AD and the other a caregiver. At certain points in the simulated conversation, the user will be required to select the caregiver's next dialogue action from a list of choices, then will see and hear the caregiver perform the action, and observe the effect on the ensuing conversation. For greater social realism, the characters should communicate through body language and facial expression as well as through speech. Because the program must serve a variety of user groups, the system needs to provide animated characters whose appearances and voices represent a variety of socio-cultural groups.

The system will present the user with multiple *vignettes.* A vignette represents a conversation fragment in which part of the AD character's life story may be elicited by the caregiver. Although a vignette is not a complete story, it should have narrative properties. We believe this is necessary to create an engaging experience for the user. One reason for designing the presentation in terms of vignettes is for the sake of realism; i.e., in real life a caregiver is more likely to elicit pieces of an autobiographical story over time. Another motivation is so that the user is free to select the vignettes relevant to her training needs.

System Architecture and Components

The system is designed as a Model-View-Controller architecture, described in more detail in (Green 2004). One View is the performance of the conversation by the animated characters as audio-visual output; another View is a scrollable display of the dialogue history. The Controller is a graphical user interface for displaying and accepting the user's choices. A screen shot of the graphical user interface is shown in Figure 1.

The Model consists of four objects. The ADA ("AD Agent") object simulates relevant aspects of the mental state of a person with AD. The CG ("Caregiver Agent") object simulates relevant aspects of a caregiver's mental state. Each of the mental models is implemented in a rule-based inference system. The Director object is responsible for coordinating the dialogue, e.g., consulting the script, accessing the mental states of the ADA and CG, sending instructions to the animated characters portraying the ADA and CG, and sending requests to the Controller to present the user with a set of choices for the next dialogue action.

Lastly, the Script object provides access to scripts authored by system developers. A script is a directed graph implementing a vignette. Each arc is a tuple of the form (Agent, Agent-Precondition, Agent-Action, Agent-Effect, Beneficiary-Effect), where Agent is the character who currently has the turn and Beneficiary is the other dialogue participant. Agent-Precondition is a condition that must be true of the given agent's mental model in order for Agent-Action to be authorized. The Agent-Precondition is a goal for which a proof is attempted in the program representing the agent's mental model; a proof may have side-effects such as actions to be performed by an animated character at the beginning of the turn.

Agent-Action is the dialogue action to be performed by the given agent. In our prototype it consists of a pair, (Abstract-Action, Code). Abstract-Action describes the linguistic function of the action, such as *greeting*. Code is a sequence of text strings to be realized by text to speech synthesis (TTS) and/or hard-coded gesture annotations to be realized by the animation. (However, as we describe shortly, most non-spoken output is generated from the characters' mental models, rather than from hard-coded instructions in the script.)

Agent-Effect and Beneficiary-Effect are the effects of Agent-Action on the agent who has the turn and on the other participant, respectively. The Agent-Effect and Beneficiary-Effect are goals for which a proof is attempted in the programs representing the corresponding character's mental model; these goals may result in updates to an agent's mental model and actions to be performed by an animated character at the end of the turn.

The design is a simplification of an earlier proposal (Green 2002). The motivation for simplifying that design was to experiment with low-cost off-the-shelf components that could be deployed on personal computers. Thus, the animated characters are being implemented using Microsoft Agent[1]. Speech output is handled by a text-to-speech synthesizer[2] during development, but for deployment we plan to substitute recorded[3] for generated speech to increase social realism. Prolog was selected to implement the mental models as a rule-based system.

Dialogue Generation

In the current version of the system, the mental models of the ADA and CG include static rules and a dynamic mental state. The ADA's dynamic mental state includes a representation of emotional state and of dialogue act expectations (e.g. the expectation to answer a question). The emotional state of the ADA is represented on a five-point scale of negative to positive affect. (Many gerontologists use negative and positive affect in definitions of subjective quality of life (Ronzijan and Luszcz 2000).) Agent-Preconditions in the script may test the ADA's emotional state. For example, some CG actions may not be appropriate unless the ADA is in a neutral or positive state. Also, currently, we are using preconditions to trigger gesture and facial expression at the beginning of a turn.

Beneficiary-Effects on arcs of the script representing the CG's turn may influence the ADA's emotional state. For example, the Beneficiary-Effect *update-effect(show-regard)* on an arc for the CG action of showing regard to the ADA may trigger a rule in the ADA's mental model that increases the ADA's emotional state by one unit (unless the ADA's emotional state is already at a maximum or unless it is too low for the action to have this effect). Another common Beneficiary-Effect is the creation of an expectation that the beneficiary will respond to a question. Also, Agent-Effects on arcs of the script representing the ADA's own turn may influence the ADA's emotional state. For example, possible Agent-Effects are to update and show the ADA's emotional state resulting from remembering something pleasant or satisfying a conversational expectation.

The mental model of the CG agent includes static rules and a dynamic mental state representing the CG's persistent social goals and beliefs about the ADA's current state. Simulating social conversation requires a model of persistent social goals (Bickmore and Cassell 2000). In our application, these are social goals that the CG may adopt in response to beliefs about the emotional state of the ADA; for example, if the ADA is believed to be in the lowest state, the CG may adopt a goal to cheer up the ADA. This goal may persist for more than one turn of dialogue, i.e., until the ADA's affective state seems to have increased to neutral or higher.

The scripts in the prototype are based on a corpus of transcripts of recordings of 75 hours of conversations between twelve nursing home residents with AD and the "normal" speakers who visited them, as well as feedback

[1] www.microsoft.com/msagent

[2] Downloaded with Microsoft Agent.

[3] Using recordings of conversations between nursing home residents with AD and visitors.

from experts. For example, one script intended to teach how the phrasing of questions can help someone with AD to remain active in the conversation is based on the following excerpt from the corpus. At turns 4 and 6, the resident with AD (A) was unable to respond to the visitor's (C) questions; finally in turn 8, C rephrased the question in a way that enabled A to respond appropriately in turn 9. This illustrates the finding of our colleagues that while a speaker with AD may not be able to answer a direct question, he may be able to answer other forms such as a request for confirmation. Part of a script based on this recorded episode is shown in Figure 2.

A1: [looking down]
C2: That's a great looking shirt.
A3: [look up, smile]
C4: Where did you get it?
A5: Huh?
C6: Did your daughter give it to you?
A7: [look away, silence]
C8: Your daughter gave it to you, didn't she?
A9: [look at C] Yes.

In summary, in the current version of the system, generation of what is spoken by the participants is hard-coded in a script. However, there is more to dialogue than the words that are spoken. The prototype generates changes in the dynamic mental state of the participants throughout the dialogue, which may be exhibited in gesture and facial expression. Also, although currently the choice of which branch to follow in a script is controlled by the user's choice of the CG avatar's next action, we plan to extend the system to enable the Director to choose the next branch based on satisfaction of its preconditions. This will enable more complex conversational scenarios to be illustrated without requiring user action except at points in the dialogue that serve an educational function.

Related Work

The interaction style and internal design of the current version of the system was inspired by Carmen's Bright Ideas (CBI), an interactive pedagogical drama (IPD) instructional system (Marsella et al. 2000, 2003). CBI is intended to allow mothers of pediatric cancer patients to improve problem-solving skills by participating in a simulated counseling session between Carmen, a mother, and Gina, a counselor. The philosophy of IPD is to engage the learner through her participation in an interactive narrative. In CBI, the script was designed to be a compelling story while leading the user through the steps of a problem-solving methodology. As the story progresses, the user is able to choose some of Carmen's thoughts or dialogue actions during her conversation with Gina. Dialogue in CBI is generated from a state-machine-based script and Carmen's emotion model.

Despite the similarities with CBI, internally and externally, there are differences between CBI and both our current and planned systems. At the user-interaction level, instead of a single compelling story, our system presents the user with vignettes. Furthermore, although the story elements of a vignette are determined by the author of a script in the current version of the system, we plan to develop a model of content selection from the ADA's simulated life memories for generating the narrative elements of a vignette (Green 2002). Other differences follow from the requirement to simulate the mental state of persons with AD. In future versions of the system, we plan to simulate not only emotion but also cognitive problems. In addition, since AD is a progressive disease it may be necessary to have different mental models representing characters at different stages of AD.

Other related work includes development of training systems in which a user can interact in spoken language with virtual characters that have simulated emotions. For example, JUST-TALK is a system that enables the user to practice talking to five different 3D virtual characters simulating persons under stress or who have a mental illness (Hubal et al. 2003). The emotional state of the virtual character is represented as a set of emotion variables (anger, confusion, fear, etc.) that influence the virtual character's response to the user's input. In the MRE system (Swartout et al. 2001), which enables users to practice military decision-making while talking with virtual humans, a virtual human's emotion model is based on a psychological model of an agent's appraisal of events (Traum et al. 2004). Emotion may influence the dialogue, e.g. by motivating the virtual human to take the initiative to bring up a topic of concern.

In addition, our work is related to research on embodied conversational agents (ECAs) whose behavior is informed by models of interactional, attitudinal, and relational functions of language (Bickmore 2004). For example, Laura is an ECA designed to establish and maintain a long-term social-emotional relationship with the user in order to promote health behavior change (Bickmore 2003). Laura implemented relational functions such as establishing and maintaining the user's trust, and attitudinal functions such as signaling liking the user. Although our agents are not intended to interact directly with a user, these functions of language are important in the kinds of interaction to be simulated in our system.

Future Work

As we noted above, our future work includes using AI to generate an ADA's story, and extending the model of the ADA's mental state to simulate the cognitive problems experienced by people with AD. Other possible future work includes developing a computational model for linguistic realization of the ADA's speech (Green and Davis 2003). This could be used both to evaluate cognitive-linguistic models of AD (through simulation) and to demonstrate to caregivers the progressive effects of AD on language. Another area of future work is to mine the corpus of transcripts of conversations with nursing home residents collected by our colleagues. We would like to empirically evaluate the proposed interventions in terms of metrics such as apparent effect on the emotional state of the person with AD or the number of turns that continue the current topic following an intervention.

In addition to extending the capabilities of the current version of the system, there are many research issues to be addressed in the area of evaluation. As a first step, we plan to run a formative evaluation in the near future to get feedback from experts on the believability of the AD characters and from potential users on how they like using the system. Evaluation of JUST-TALK and similar systems (although getting positive feedback from users) pointed out problems due to shortcomings in speech recognition technology (Guinn et al. 2004). We would like to explore how to make effective use of spoken input in AD caregiver training despite the limitations in speech recognition technology.

Acknowledgments

William Lawton received funding as a Graduate Research Assistant in Fall 2003 from NSF DUE Grant 9986747.

References

Bickmore, T. and Cassell, J. 2000. "How about this weather?" Social Dialogue with Embodied Conversational Agents. In *Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents*.

Bickmore, T. 2003. Relational agents: Effecting change through human-computer relationships. *Media Arts and Sciences*, MIT, Cambridge, MA.

Bickmore, T.W. Unspoken Rules of Spoken Interaction. *Communications of the ACM* 47(4), April 2004, 39-44.

Green, N. 2002. A Virtual World for Coaching Caregivers of Persons with Alzheimer's Disease. *AAAI Workshop on Automation as Caregiver: The Role of Intelligent Technology in Elder Care*, 18-23.

Green, N. and Davis, B. 2003. Dialogue Generation in an Assistive Conversational Skills Training System for Caregivers of Persons with Alzheimer's Disease. *Natural Language Generation in Spoken and Written Dialogue: Papers from the AAAI 2003 Spring Symposium*, 36-43.

Green, N. 2004. A Software Architecture for Simulated Human-Human Conversation with User Intervention. In *Proceedings of IEEE SoutheastCon 2004*, 48-53.

Guinn, C. et al. 2004. Usability and Acceptability Studies of Conversational Virtual Human Technology. In *Proceedings of Fifth SIGDIAL Workshop on Discourse and Dialogue*.

Hubal, R.C., Frank, G.A., and Guinn, C.I. 2003. Lessons Learned in Modeling Schizophrenic and Depressed Virtual Humans for Training. In *Proceedings of IUI'03*, 85-92.

Irvine, A.B., Ary, D.V., and Bourgeois, M.S. 2003. An Interactive Multimedia Program to Train Professional Caregivers. *Journal of Applied Gerontology* 22(2), June 2003, 269-288.

Marsella, S.C., Johnson, W.L, and LaBore, C.M. 2003. Interactive Pedagogical Drama for Health Interventions. In *AIED 2003, 11th International Conference on Artificial Intelligence in Education*.

Marsella, S.C., Johnson, W.L, and LaBore, C.M. 2000. Interactive Pedagogical Drama. In *Proceedings of the Fourth International Conference on Autonomous Agents*, 301-308.

Moore, L. and Davis, B. 2002. Quilting narrative: Using repetition techniques to help elderly communicators. *Geriatric Nursing* 23:2-5.

Ranzijn, R. and Luszcz, M. 2000. Measurement of Subjective Quality of Life of Elders. *Int'l J. Aging and Human Development* 50(4): 263-278.

Swartout, W. et al. 2001. Towards the Holodeck: Integrating Graphics, Sound, Character and Story. In *Proceedings of the Fifth International Conference on Autonomous Agents*.

Traum, D.R., Marsella, S., and Gratch, J. 2004. Emotion and Dialogue in the MRE Virtual Humans. In *Proceedings of the Tutorial and Research Workshop on Affective Dialogue Systems.*

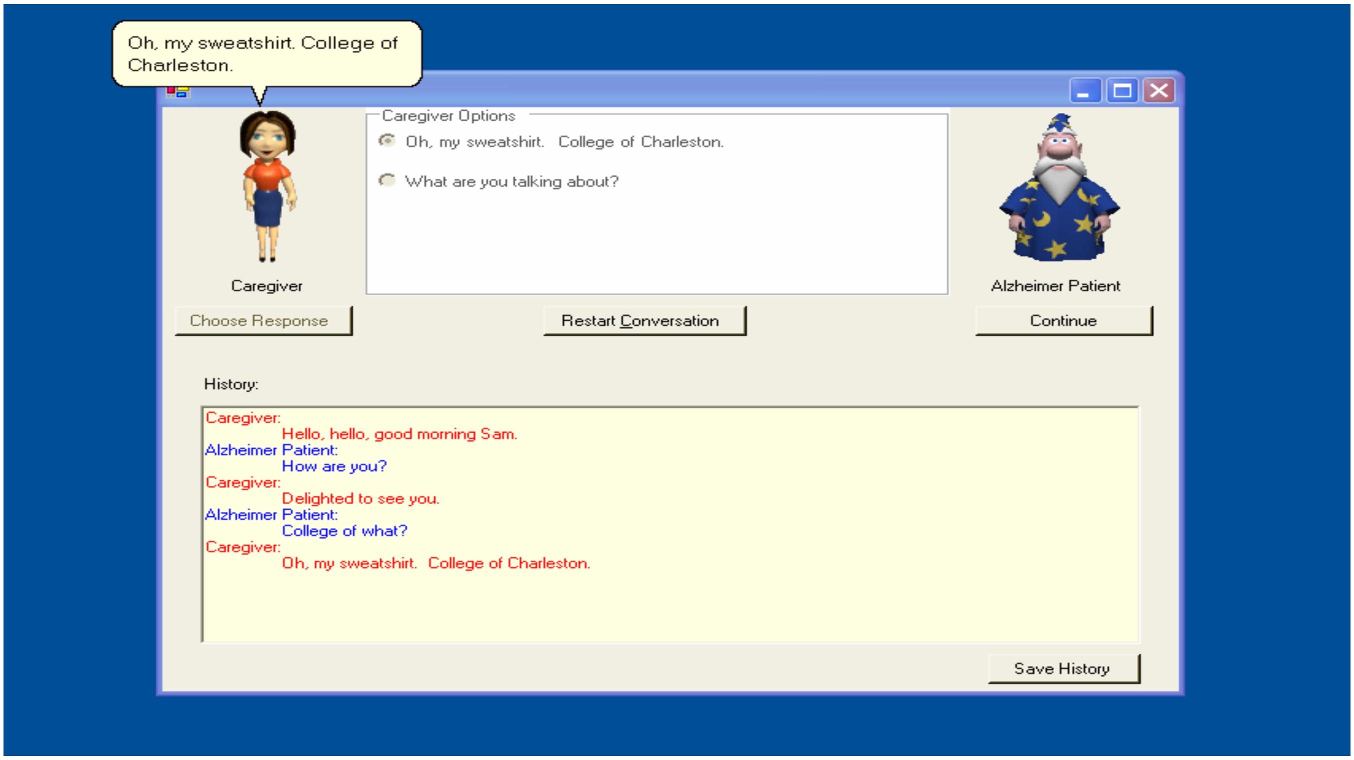

Figure 1. Screen shot of prototype system.

Arc/Agent	Precondition	Agent-action	Agent-effect	Beneficiary-effect
2/CG	goal(begin-turn)	(show-regard, "That's a great shirt!")		add-effect(show-regard)
3/ADA	goal(emote)	(silence, "")		
4/CG	goal(begin-turn)	(direct-question, "Where did you get it?")	add-expectation(ada, direct-question)	add-expectation(ada, direct-question)
5/ADA	goal(begin-turn)	(request-clarification, "Huh?)	add-expectation(cg, request-clarification)	add-expectation(cg, request-clarification)
6/CG	goal(begin-turn)	(direct-question, "Did your daughter give it to you?")	add-expectation(ada, direct-question)	add-expectation(ada, direct-question)
7/ADA	goal(emote)	(silence, "")	satisfied-expectation(ada, no)	
8/CG	goal(begin-turn)	(request-confirmation, "Your daughter gave it to you, didn't she?")	add-expectation(ada, request-confirmation)	add-expectation(ada, request-confirmation)
9/ADA	goal(begin-turn)	(answer-req-conf, "Yes")	satisfied-expectation(ada, yes)	

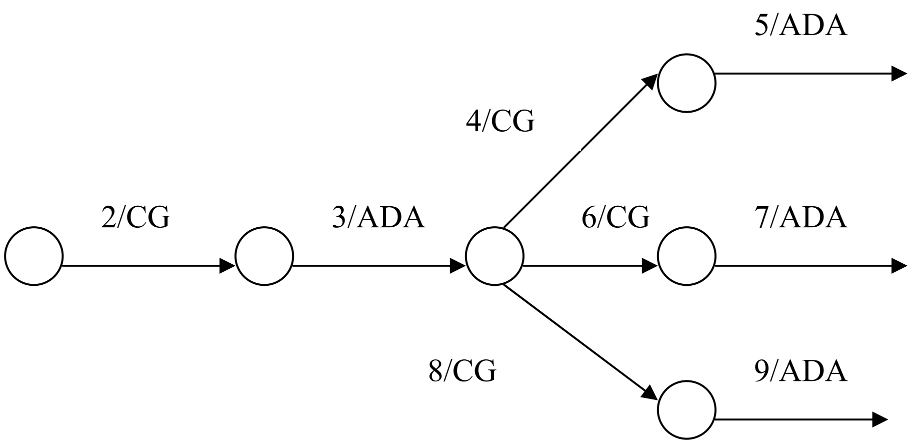

Figure 2. Part of script of a vignette that illustrates a person with AD who is unable to answer a direct question but can answer a request for confirmation. After arc 3, the user is presented a choice of three caregiver actions, represented by arcs 4, 6, and 8. The ADA's responses are given on arcs 5, 7, and 9, respectively. The precondition, *goal(begin-turn)*, is satisfied if the agent is ready and willing to take the turn, and has the side-effect of displaying the agent's current emotional state. In the case of the caregiver, this may consist of showing interest or empathetic happiness; in the case of the ADA, this consists of expressing the current degree of negative or positive emotion. Note that the emotional state of the ADA is not coded in the script. It is computed from the effects of the actions given in the script. For example, the ADA's emotion can change in the negative direction when the ADA fails to satisfy a conversational expectation such as the expectation to answer a direct question or in the positive direction when the ADA can satisfy the expectation. Thus, in the states following arc 5 and 7, the ADA's emotional state has moved in the negative direction, while after arc 9, it has moved in the positive direction.