Acceptance and Usability of a Relational Agent Interface by Urban Older Adults

Timothy W. Bickmore
Medical Information Systems Unit
Boston University School of Medicine
720 Harrison Ave. #1102, Boston, MA
bickmore@bu.edu

Lisa Caruso, Kerri Clough-Gorr
Geriatrics
Boston University School of Medicine
88 E Newton St. B-2103, Boston, MA
lisa.caruso@bmc.org, kmclough@bu.edu

ABSTRACT
This study examines the acceptance and usability of an animated conversational agent designed to establish long-term relationships with older, mostly minority adult users living in urban neighborhoods. The agent plays the role of an exercise advisor who interacts with subjects daily for two months on a touch-screen computer installed in their homes for the study. Survey results indicate the eight subjects who completed the pilot study (aged 62-82) found the agent very easy to interact with, even though most of them had little or no previous experience using computers. Most subjects also indicated strong liking for and trust in the agent, felt that their relationship with the agent was more similar to a close friend than a stranger, and expressed a strong desire to continue working with the agent at the end of the study. These results were also confirmed through qualitative analysis of post-experiment debrief transcripts.

Author Keywords
Relational agent, embodied conversational agent, social agent, older adult, health, longitudinal study.

ACM Classification Keywords
H5.2 [Information Interfaces and Presentation]: User Interfaces---Evaluation/methodology; Graphical user interfaces; Interaction styles; Natural language; Theory and methods; Voice I/O.

INTRODUCTION
Automated systems are being increasingly used to deliver tailored health information using a variety of media including printed materials, telephony, video, and interactive computer-based systems, in order to effect health behavior change [7]. Although some researchers have found that many older adults readily accept new technologies such as computers, this segment of the population lags behind all other age groups with respect co computer ownership (only 25.8% of senior households have a computer) and Internet access (14.6% of all senior households have Internet access) [2]. Further, more than one-third of U.S. adults over 65 have inadequate or marginal functional health literacy, and among indigent and minority patients in urban areas this number rises to over 80% [9].

While older adults are disadvantaged with respect to their access to advanced computing and communication technologies, they arguably are in the most need of them when it comes to delivery of health information and automated interventions. For example, only 12% of adults over 75 get the minimum level of physical activity currently recommended by the Centers for Disease Control and Prevention, and 65% report no leisure time activity [1].

Relational agents—computational artifacts designed to build long-term social-emotional relationships with users using simulated face-to-face conversation [4]—may provide an accessible user interface for much of the older adult population, and an especially effective channel for health communication and behavior change interventions. The interface relies only minimally on text comprehension and uses the universally understood format of face-to-face conversation, thus making it less intimidating and more accessible for patients with low literacy skills. The agent's use of certain nonverbal conversational behaviors—such as hand gestures that convey specific information through pointing or through shape or motion—also provides redundant channels of information for conveying semantic content also communicated in speech, thus enhancing the likelihood of message comprehension. Further, by using appropriate social behavior to establish trust and social bonding with the user, the agent is more effective at gaining user adherence to prescribed behavior change regimens, and more effective at keeping users engaged with the communication channel itself over the months or years it takes to carry out the intervention [3,4].

In this paper we describe the design and evaluation of a relational agent that plays the role of an exercise adviser, designed to be used by older adults in their homes daily for two months to motivate them to walk more.
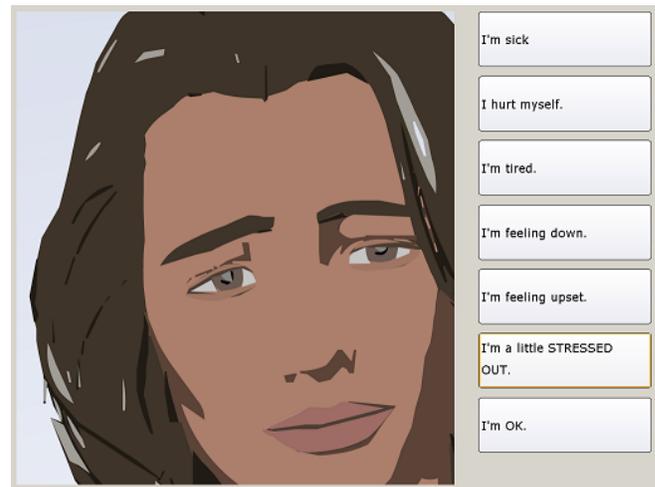
PREVIOUS WORK

There is a growing body of work on accessible technologies for older adults. However, very little has been done on evaluating the use of embodied conversational interfaces for this purpose. Perhaps the only such agent developed to date specifically for older users is the GrandChair system [8]. In this system the agent had the appearance of a 6-year-old child, and was designed to elicit narratives from older adult users. The system used postural and intonational input to provide continuation prompts at appropriate intervals ("tell me more"). In a study involving 15 grandmothers (aged 55 to 65) the agent was found to elicit significantly more and longer stories than a text-prompt control condition. Fourteen of the 15 users enjoyed the experience, and all users automatically adopted a child-directed speaking style appropriate for a 6-year-old child.

Similarly, few automated health behavior change interventions have been designed specifically for the older adult population. One study of note used an automated telephone system (recorded speech output, touch-tone input) to promote exercise adoption through weekly calls over a three month period to 41 older adults (average age 66) [5]. Compared to a control group, those receiving the intervention and who walked 15 minutes or less at baseline significantly increased the number of minutes walked at the end of the intervention. In addition, users were very satisfied with the system, giving it a score of 8.6 (where 1 was very dissatisfied and 10 was very satisfied), and 74% of the users rated their satisfaction as 10 of 10.

Finally, the FitTrack system involved the use of a relational agent who played the role of an exercise advisor designed to motivate users to exercise more, although it was designed for use by college-aged adults [4]. In a one-month daily contact study the system demonstrated the ability of the agent to establish a social bond ("working alliance") with users, compared to the same agent with relationship-building behaviors ablated.

A RELATIONAL AGENT FOR OLDER ADULTS

The relational agent used in the FitTrack study [4] was modified for use by the new target population. In conversation, the agent uses synchronized synthetic speech and animation and the user chooses their responses (see Figure 1). The new interface uses large buttons with enlarged text to allow for easy readability and touch screen input. A numeric keypad (used in conjunction with the touch screen) was implemented for participants to enter their pedometer readings (used to track amount of walking). The system was designed to be used stand-alone—since we could not assume our participants had Internet connectivity—with participants provided with a dedicated-use PC, 17" color touch screen monitor (no keyboard or mouse) and table for use during the study. Participants only need to push the start button on the PC and it automatically runs the relational agent interface, conducts the 5-10 minute daily interaction, then automatically shuts down. In order to prevent loss of data due to system crashes or theft (a

Figure 1. Relational Agent Interface

concern given the neighborhoods we were targeting) we originally designed the system to walk the user through a weekly floppy disk backup procedure.

The conversational scripts used by the agent (augmented transition networks) were designed to be used daily over a two-month period of time. The daily conversation involved having participants enter the number of steps they had walked on the previous day, showing them a self-monitoring graph of their progress, providing positive feedback (when warranted), discussing obstacles to exercise, and negotiating a walking goal for the following day ("shaped" to gradually work them from their baseline level up to their two month goal). Since there was very wide variability in the amount of walking our participants may be performing (hundreds of steps per day to tens-of-thousands of steps per day) we set different two-months goals for participants based on their first week's baseline data (10,000 steps/day if their baseline was over 5,000, otherwise 5,000 steps/day). In addition, the agent used all of the relational behaviors used in FitTrack, including social dialogue, empathy exchanges, meta-relational communication, and nonverbal immediacy behavior.

Pre-Test Study

Prior to the intervention study, the interface was pre-tested with five participants in the Geriatric Ambulatory Practice (GAP) clinic at Boston Medical Center, the primary safety net hospital in the Boston area. All participants found the system easy to use with the exception of the floppy-disk backup procedure (two participants accidentally re-booted the PC while attempting to eject the disk). Four of the five reacted positively to the agent and indicated that they would be interested in having the agent in their homes to talk to them daily about exercise.

Based on these results we eliminated the floppy-disk backup procedure from the protocol (note that no loss of data actually occurred during the intervention study).

ACCEPTANCE STUDY

To evaluate the acceptance and efficacy of the exercise advisor agent by older adults we conducted a randomized trial comparing participants who interacted with the agents daily in their homes for two months with a standard of care control group who were only given pedometers and print materials on the benefits of walking for exercise. In this paper we will focus on the intervention group and their use of the system.

Experimental Methods

Subjects. Ten participants were recruited into the intervention condition, based on referrals from the GAP clinic. Participants ranged in age from 62 to 82 (mean 72.8), were all female, 90% were African American, and 80% were single, widowed or divorced. Half reported never having used a computer before, and another 30% reported having used one "a few times". Nine of the 10 participants were scored as having low reading literacy [6]. Participants were compensated for their time.

Procedure. Candidates were contacted by phone, the study explained to them (involving the use of a "small television with a cartoon character") and an intake meeting scheduled. A researcher traveled to the home of each participant at intake and again at follow-up two months later. At intake, consent was obtained, the researcher set the computer system up, participants filled out a questionnaire, were showed how to use the pedometer, and conducted an initial training session with the agent. At follow-up, participants filled out another questionnaire, and their written step records and the computer were collected.

Measures.

Interaction History was recorded in log files that kept track of all actions participants took with their system.

Steps Walked was measured by a pedometer. Participants were instructed to write their steps down each day on a log sheet, and they were also prompted to enter these into the computer during conversations with the agent.

Satisfaction with, Repetitiveness, Friendliness, Informativeness, Interestingness, Liking of, and Trust in the agent were measured by single items on seven-point Likert scales, as were *Ease of Use* of the system, *Desire to Continue* using the system, and *Relationship* to the agent.

In addition, semi-structured interviews were held with each participant during the follow-up meeting.

Results: Retention, Usage and Engagement

Two participants withdrew from the study before the end of the two month intervention (one at day 21 the other at day 52), both reportedly due to health problems with themselves or a family member (for comparison, 2 of 11 subjects in the control group withdrew for the same reason). In addition, one participant was found to have not turned the system on once following the intake meeting, even though she claimed to have used it for awhile. Data from the participant who

withdrew at day 21 and the participant who did not use the system at all are not included in the following analyses.

Participants were asked to try to use the system daily, but that it was OK if they needed to miss a day or two. Actual use during the 60-day intervention ranged from 10 to 54 interactions, with a mean of 25.1 (42%). The typical usage pattern was daily during the first week (4.6 contacts/wk average), tapering off to once or twice a week by the end (2.3 contacts/wk average). When asked if they looked forward to the interactions with the agent, subjects gave a range of responses, with 5/8 responding positively:

- Yes, I... yes. Because those two nights I forgot--I think, maybe I had been out late or whatever--but I was really surprised, I was like "Oh I forgot Laura". Then I'd turn the light on and talk to her. But it was something I looked forward to, I'd say it was my little night cap.

- I can't say that I looked forward to it. If I did I would have called in every day.

Participants also indicated that they would like to continue using the system, giving this an average rating of 6.4 (with 1="not at all" and 7="very much").

Results: Ease of Use

All subjects found the system easy to use, rating this an average of 1.9 on a 1 ("easy") to 7 ("difficult") scale. Except for some problems entering pedometer steps (described below) and a few other minor problems, none of the subjects reported having any problems using the system:

- That is *so* easy. That is so good. Regular computers I don't do. But, that was so easy, even a baby could do that.

A number of participants did have problems using the touch screen keypad to enter their pedometer steps. Compared to their written records they made errors in data entry 49% of the time, often by dropping or duplicating a digit, causing an order-of-magnitude error. This was a major problem if it occurred during the first week when the system was collecting baseline data, as it caused the agent to set an unrealistically high goal for two participants. Two participants also reported problems with touch screen alignment that made selection difficult.

Participants felt that the simulated conversation worked reasonably well. When asked if they felt that they and the agent understood each other, participants rated this at 5.4 (with 1="not at all" and 7="very much"). Only one instance of a problem talking to Laura was reported: the participant thought the agent asked her "are you tired?" when she was really asking "are you retired?" causing the agent to ask her a series of inappropriate follow-up questions. Several participants mentioned that they could not express themselves completely using the constrained interaction:

- When she ask me questions ... I can't ask her back the way I want.

- I felt that she was programmed to answer. She was programmed to listen to the questions that you put on the screen. She would ask a question and I would have a choice, one, two, three, four. … But I could never explain. Or she could never follow up, or follow through with another question.

Results: User Acceptance and Satisfaction

Satisfaction with the overall intervention was very high, with most participants acknowledging that it was for their benefit:

- It was the best thing that happened to me, to have something that pushed me out and get me walking.

- I appreciated having that kind of a reminder, because I don't have anybody who will tell me what to do, to remind me, you know, to get up, get out and get some fresh air.

When asked specifically about the agent, reviews were a little more mixed, as shown in Table 1:

- She's nice. She's really good. Really good. She asks you the right questions. She tells if you if you're not doing up to par, you know, and all that. And if you're doing good, she'll tell you. If you're not she'll tell you. And it's honest. And it works. It really does. I like it. I like talking to her.

- You'd be talking to her and sometimes you forget and think she's a real person.

- I liked the … fact that it wasn't cut and dried every single night. …

- Um, she took getting used to. But, um, by maybe the third or the fourth night she appeared to be familiar, you know…

- It was an animation. You know those questions [on the follow-up questionnaire] were inappropriate.

	Min (1)	Max (7)	Mean
Satisfaction with	Not at all	Very	5.4
Liking of	Not at all	Very much	6.3
Trust in	Not at all	Very much	6.4
Relationship with	Stranger	Close friend	6.8
Friendly	Not at all	Very	6.7
Informative	Not at all	Very	6.5
Repetitive	Not at all	Very	4.8
Interesting	Boring	Interesting	6.4

Table 1. Ratings of Relational Agent

Results: Efficacy

Although full results on efficacy must await the comparative analysis with the control group, all 8 participants increased their number of daily walking steps between the baseline week and the last week they kept records for, with an average increase of 215%.

CONCLUSION

Relational agents are an especially effective modality for delivering health communication and health behavior change interventions to older adults, especially those with low functional health, reading, or computer literacy. Participants in our study had very few problems using the overall system or talking to the agent; the only notable exception was the part of the interaction that was non-conversational (pedometer step entry). Future work in this area should focus on more sophisticated models of dialogue planning—so that users can engage in richer conversations and can more freely express themselves—and on automated upload of any sensor readings (e.g., from pedometers) to eliminate user data entry errors.

REFERENCES

1. *Healthy People 2010,* Office of Disease Prevention and Health Promotion, US Dept of Health and Human Services, Washington, DC, 2001.
2. *Falling through the Net: defining the digital divide. A report on the telecommunications and information technology gap in America,* National Telecommunication and Information Administration, Washington, DC, 1999.
3. Bickmore, T. and Cassell, J., Relational Agents: A Model and Implementation of Building User Trust, *Proc. CHI 2001*, ACM Press (2001), pp. 396-403.
4. Bickmore, T. and Picard, R. Establishing and Maintaining Long-Term Human-Computer Relationships. *ACM Transactions on Computer Human Interaction,* (to appear).
5. Jarvis, K., Friedman, R., Heeren, T., and Cullinane, P. Older women and physical activity: using the telephone to walk. *Womens Health Issues,* 7 (1997) 24-29.
6. Lobach, D., Hasselblad, V., and Wildemuth, B., Evaluation of a Tool to Categorize Patients by Reading Literacy and Computer Skill to Facilitate the Computer-Administered Patient Interview, *Proc. American Medical Informatics Association Symposium* (2003), pp. 391-395.
7. Revere, D. and Dunbar, P. Review of Computer-generated Outpatient Health Behavior Interventions: Clinical Encounters "in Absentia". *JAMA,* 8 (2001) 62-79.
8. Smith, J., *GrandChair: Conversational Collection of Grandparent's Stories,* MS Thesis, MIT, 2000.
9. Williams, M., Parker, R., Baker, D., Parikh, N., Pitkin, K., Coates, W., and Nurss, J. Inadequate Functional Health Literacy Among Patients at Two Public Hospitals. *JAMA,* 274, 21 (1995), 1677-1720.