Negotiating Task Interruptions with Virtual Agents for Health Behavior Change (Short Paper)

Timothy Bickmore, Daniel Mauer, Francisco Crespo, Thomas Brown
Northeastern University College of Computer and Information Science
360 Huntington Ave, WVH202, Boston, MA 02115
{bickmore,daniel,crespof,brown.tho}@ccs.neu.edu

ABSTRACT

Virtual health counseling agents on mobile devices need to be able to interrupt their users when it is time for them to engage in healthy behaviors, such as scheduled medication taking or exercise. However, these real-time reminders often represent task interruptions for individuals who are engaged in work activities. This paper presents the results of a study which compares four strategies used by a virtual agent on a PDA for interrupting users at work to perform a healthy behavior. We find that, among several interruption coordination strategies previously explored in the HCI literature, empathic interruptions are superior overall in gaining both short-term compliance and self-reported desire to continue working with the agent.

Categories and Subject Descriptors

H5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology, Graphical user interfaces.

General Terms

Algorithms, Measurement, Performance, Design, Experimentation, Human Factors.

Keywords

Interruption, virtual agent, embodied conversational agent, politeness, compliance, mobile computing.

1. INTRODUCTION

Poor lifestyle health behaviors, such as lack of physical activity and unhealthy dietary habits, are among the leading causes of death and chronic disease in the United States [15]. In addition, adherence to prescribed treatments — such as medication regimens — is estimated to average only 50%, and represents another significant source of morbidity, mortality and healthcare cost to the nation [10]. Each of these large classes of health behavior problems have been the targets of numerous technology-based interventions in recent years.

One of the simplest such interventions is a reminder system that alerts users when it is time to engage in a healthy behavior, such as going for a walk or taking medication. Such "cueing" or "stimulus control" techniques are used by most individuals who have successfully changed their health behavior [17]. For users who happen to be sitting idly when these alerts are triggered, the reminders may result in a relatively high compliance rate. However, as many recent studies in task interruption have shown, responsiveness to an interruption (in this case, compliance with the recommended health behavior) depends crucially on what the user is doing at the time the interruption presents itself [11] in addition to many other factors such as the emotional state of the user [12] and the modality of the interruption [1].

We are developing a mobile, PDA-based health advisor that is able to provide real-time reminders and conversational counseling to help users change their health behavior (Figure 1) [2,3]. We are exploring different interruption strategies that can be used by the advisor to persuade users to perform a healthy behavior while they are working at routine office tasks. In the current work, the advisor promotes "wrist rests" while users are typing at a desktop computer. The PDA-based advisor discusses the importance of taking frequent breaks from typing in order to avoid repetitive stress injury and other upper body musculoskeletal disorders, and interrupts users periodically with a request to rest their wrists. Prevention of such disorders is an important health problem: in 2002 they accounted for two thirds of all reported occupational illness in the US, and intervention studies that promote wrist rests for computer workers have demonstrated significant reductions in

Figure 1. PDA and Virtual Agent Interface

Cite as: Negotiating Task Interruptions with Virtual Agents for Health Behavior Change (Short Paper), Timothy Bickmore, Daniel Mauer, Francisco Crespo, Thomas Brown, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, pp. 1241-1244.

self-reported symptoms [7].

In the current study, we compare four interruption coordination strategies previously explored in the HCI literature to determine which is best at gaining immediate compliance while maintaining user desire to continue working with the advisor.

2. RELATED WORK

Although there is a significant and growing literature on technologies to promote health behavior change, very little of this literature describes real-time interventions. Several systems have been developed to provide older adults with real-time reminders to perform various activities of daily living, including medication taking (e.g., [9]). The ILSA system used automated phone calls to provide real-time reminders to older adults living alone, but the calls were not always effective and users did not like them [9]. Pollack *et al.* developed the AutoMinder system, which could reason about whether, when and how to give a reminder based on a deep understanding of the tasks involved and the user's schedule (e.g., "If you take your medicine now, I won't have to interrupt you during your favorite TV show later") [16]. Preliminary evaluation indicated that acceptance among older adult users was high, although results of any efficacy evaluations have not been reported.

Goetz, *et al*, evaluated user compliance to the requests of a "nurse" robot to perform exercise. The robot was controlled by a confederate ("wizard of oz") and spoke using synthetic speech. Study subjects performed more exercise when the robot was "serious" (emphasizing the importance of exercise) than when it was "playful" (telling jokes and making fun of exercise) [8].

There has also been considerable research done in the area of task interruption of computer users in recent years. Much of this work is primarily concerned with the impact of interruptions on task performance, while our focus is primarily on user responsiveness to the interruption. According to McFarlane, negotiation-based methods, which give users more control over the interruption process, should provide the best overall performance [14]. A number of studies have shown that such methods, in which users are alerted that there is a notification, but are able to control whether or when the full content of the notification is displayed, are preferable to simpler models in which the full notification is delivered immediately. Czerwinski *et al.* also found that delivering a pre-interruption warning prior to the delivery of the content of the interruption can also have a significant positive effect on performance [6].

There is also evidence that the use of empathy in interruptions can create a more positive user experience. Liu and Picard presented a wearable system that periodically interrupted users and asked them (via text-based prompts) to annotate whether or not it was a good time to interrupt, and to specify their current stress level and activity [13]. The use of empathic language in the system prompt was varied within subjects, who reported significantly higher desire to continue using the empathic version of the system. Additionally, subjects perceived a lower frequency of interruptions when using the empathic system. However, Liu and Picard did not investigate the impact of empathic interruption on compliance, or compare their approach to other interruption coordination strategies (users were required to either acknowledge an interruption or cancel it immediately).

3. EXPERIMENTAL PLATFORM

We have developed a general purpose virtual agent interface for use on handheld computers (see Figure 1). The animated agent appears in a fixed close-up shot, and is capable of a range of nonverbal conversational behavior, including facial displays of emotion; head nods; eye gaze movement; eyebrow raises; posture shifts and "visemes" (mouth shapes corresponding to phonemes). These behaviors are synchronized in real time with agent output utterances. Agent utterances are displayed in a text balloon rather than using speech, to avoid privacy issues. The words in the agent utterance are individually highlighted at normal speaking speed (120 words per minute) and the nonverbal behavior displayed in synchrony. The use of this form of nonverbal behavior is motivated by an earlier study which found that users who conducted relational (getting acquainted) interactions with the system rated social bonding with the agent and caring of the agent significantly higher when the agent was presented as an animated conversational agent, compared to text and static image representations of the agent [3]. User inputs are constrained to multiple choice selections at the bottom of the display.

Interaction dialogues are scripted in an XML-based state-transition network, which allows for rapid development and modification of dialogues. Scripts consist primarily of agent utterances (written in plain text), the allowed user responses to each agent utterance, and instructions for state transitions based on these responses and other system events (timers, sensor input, etc.).

Once a script is written, it is preprocessed using the BEAT text-to-embodied-speech engine [5], which automatically adds specifications for agent nonverbal behavior.

4. EXPERIMENTAL METHOD

Following McFarlane, we evaluated two negotiation methods, one which gave users control over the start of the interruption (via a "snooze" button) and one which gave them a warning that an interruption was about to occur. We are particularly interested in developing virtual agents with social competencies, so we also evaluated the impact of empathic interruptions on long-term adherence, following the results of Liu and Picard [13]. In this condition, the device apologizes for interrupting the user, assesses their emotional state at the time of the interruption and, if warranted, provides empathic feedback. Thus, the four strategies we compared were:

- BASELINE – a single audio tone, pre-rated as very polite.
- NEGOTIATED – provides users with the ability to delay the start of an interruption, by presenting them with a "snooze" button that delays the onset of the interruption for 30 seconds (they could use this feature as many times as they liked).
- FOREWARN – provides users with a very brief and polite audio tone 15 seconds before an interruption.
- SOCIAL – delivers the interruption within a social context, by apologizing for the interruption, asking how the user is feeling, and expressing empathy.

4.1 Apparatus

Study subjects were seated at a desktop (office task) computer with 17" color monitor, keyboard and mouse. The PDA with the

wrist-rest advisor agent was placed on the desk just to the left or right of the keyboard, at each subject's discretion. Three video cameras, audio, and the computer screen video were continuously recorded during the study for subsequent analysis. There was one program always running on the office task computer that sequentially posed questions to subjects during each session and provided them with a text box within which they had to type their response. A web browser, pre-loaded to a search engine page, was open on the desktop at the start of each session as well, and subjects were encouraged to use the search engine to help them answer the questions.

The PDA was running the handheld virtual agent software described above, with one of four female characters, each with a unique name and appearance. For each of the four conditions evaluated in each study, an introduction and interruption script was written. In the introduction script the character introduces itself, tells the subject a fact about upper body musculoskeletal disorders and/or how to prevent them, a statement about the importance of taking frequent wrist rests when working at a computer, and a farewell, lasting six turns of dialogue. The interruption scripts vary by study condition, but always end with the agent saying "Please rest your wrists for as long as you can." The only allowed user response is "OK, I'll rest now" (the one exception to this being the "SNOOZE" option). After the user selects this response, the PDA display turns off until the next interruption.

4.2 Experimental Design

The study is a four treatment within-subjects design experiment. Each interruption strategy was evaluated using a PDA that was presented to the user as a "different advisor agent," with the agent having a different physical appearance and name. The order of presentation of interruption methods was counterbalanced, but with a fixed order of characters and quiz questions so that different interruption methods were presented with different questions and characters for each subject.

The duration of each primary task session is approximately 10 minutes, with the interruptions timed to occur at fixed intervals, so that there are always exactly two interruption events per primary task session. Because there are only a small number of brief inter-primary-task time intervals, most interruptions occur while subjects are in the middle of performing a primary task.

4.3 Measures

Self-report. Following each condition, subjects were asked to rate the wrist-rest agent on the scales shown in Table 1, covering various ratings of the agent as well as 'desire to continue' using the agent (a proxy for long-term adherence).

Compliance Behavior. Durations of all wrist rests taken by subjects were measured based on analysis of recorded video. Rest duration was coded from the time a subject acknowledged an interruption ("OK, I'll rest now.") until they returned to work on the primary task. This second 'end of rest' time involved subjective judgment, since subjects exhibited a wide range of behavior that could be interpreted as resting. Consequently, a coding manual was written and rest times were coded in parallel by two judges, with an overall inter-rater reliability of 0.99 (intraclass correlation coefficient), and final values taken as the average of the scores by each judge.

Primary Task Impact. Productivity on the primary task was assessed by the number of questions completed per primary task session.

4.4 Procedure

Subjects were initially told that they would be testing a health advisor who will help people "avoid repetitive stress injuries, such as carpal tunnel syndrome" by reminding them to take occasional breaks when they are working at a computer. Following instruction on the primary task, subjects were told how to interact with the agent, and then were told "When the advisor wants to talk to you, the PDA will make a sound. When you hear this, you should talk to the advisor."

At the start of every session the experimenter would tell the subject "Remember, we are keeping track of how many questions

Table 1. Common Self-Report Measures for Rating Interruption Methods

Measure	Question	Anchor 1	Anchor 7
POLITE	How polite was the advisor?	not polite	extremely
SATISFIED	How satisfied were you with the advisor?	not satisfied	extremely
EFFECTIVE	How effective was the advisor at getting you to respond to her?	easy to ignore	felt I had to respond
CONTINUE	How much would you like to continue working with the advisor?	not at all	very much

Table 2. Results (Significance reported from Repeated Measures ANOVA)

Measure	BASELINE	FOREWARN	NEGOTIATED	SOCIAL	Significance
POLITE	5.38	5	5.63	6.63	p<.05
SATISFIED	3.75	3.88	4.19	4.56	n.s.
EFFECTIVE	6.0	5.81	4.94	6.63	p<.05
CONTINUE	3.75	4.13	4.56	5.19	p=.15
REST TIME	23.64	17.80	24.99	26.64	n.s.
TASKS	4.19	3.94	4.69	3.63	p<.001

you answer and we will review your performance with you at the end of the study." Importantly, however, the experimenter never said anything about whether the subject should follow the advisor's instructions or anything about the importance of wrist rests or whether the subject should take rests or not. The experimenter telling subjects the importance of the office task together with the agent telling subjects the importance of the wrist rests set up a dilemma for subjects that could often be visibly observed in their behavior following each interruption.

4.5 Participants

Sixteen subjects participated in the study: 52% female, 83% students, aged 18-30. Computer, web and search engine experience was fairly high (5.3, 5.6, 6.0 respectively on 1='never used one' to 7='expert' scales), but they had less experience using PDAs (2.4 on the same scale).

4.6 Results

Results are shown in Table 2. Regarding observed compliance, subjects actually rested the longest in the SOCIAL condition.. When asked which method they thought was most effective at getting them to rest, subjects also rated the SOCIAL condition the highest, and this difference in ratings was significant. Long-term compliance, as measured by subjects' reported desire to CONTINUE use, was also highest for SOCIAL. Subjects had the worst primary task performance in SOCIAL, but it is unclear whether this was due to SOCIAL being most effective at gaining secondary task compliance or because SOCIAL interactions with the agent took slightly longer. Interestingly, the NEGOTIATED condition was rated as the least effective, although actual rest times were almost as long as for SOCIAL. Subjects did perceive a significant difference in politeness across conditions, with SOCIAL rated as significantly more polite than the other three conditions, and FOREWARN rated as the least polite.

5. CONCLUSION

Overall, the SOCIAL condition was both the most preferred and the most effective at gaining health behavior compliance, as was also rated as the most polite. Thus, designers of systems that interrupt users to perform healthy behaviors should consider adding empathy and other forms of social interaction in order to maximize long-term compliance.

6. ACKNOWLEDGEMENTS

Thanks to Jennifer Smith for her many helpful comments on this paper. This work was supported by grant R21 LM008553 from the NIH National Library of Medicine.

7. REFERENCES

[1] Arroyo, E., Selker, T., and Stouffs., A. 2002. Interruptions as multimodal outputs: which are the less disruptive? . 4th IEEE International Conference on Multimodal Interfaces, pp. 479–482.

[2] Bickmore, T., Gruber, A., Intille, S., and Mauer, D. 2006. A Handheld Animated Advisor for Physical Activity Promotion. American Medical Informatics Association Annual Symposium.

[3] Bickmore, T. and Mauer, D. 2006. Modalities for Building Relationships with Handheld Computer Agents. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI).

[4] Brown, P. and Levinson, S. C. 1987 Politeness: Some universals in language usage. Cambridge University Press, Cambridge.

[5] Cassell, J., Vilhjálmsson, H., and Bickmore, T. 2001. BEAT: The Behavior Expression Animation Toolkit. SIGGRAPH '01, pp. 477-486.

[6] Czerwinski, M., S, C., and Schumacher, B. 1991. The effects of warnings and display similarities on interruption in multitasking environments. SIGCHI Bulletin. 23, 4, 38-39.

[7] Galinsky, T., Swanson, N., Sauter, S., Hurrell, J., and Schleifer, L. 2000. A field study of supplementary rest breaks for data-entry operators. Ergonomics. 43, 5, 622-638.

[8] Goetz, J., Kiesler, S., and Powers, A. 2003. Matching robot appearance and behavior to tasks to improve human-robot cooperation. Proceedings of the 12th IEEE Workshop on Robot and Human Interactive Communication. RO-MAN 2003.

[9] Haigh, K., Kiff, L., and Ho, G. 2006. The Independent LifeStyle AssistantTM (I.L.S.A.): Lessons Learned. Assistive Technology. 18, 87-106.

[10] Haynes, R., McDonald, H., and Garg, A. 2006. Helping Patients Follow Prescribed Treatment. JAMA. 288, 22, 2880-83.

[11] Ho, J. and Intille, S. S. 2005. Using Context-Aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices. CHI.

[12] Hudson, J. M., Christensen, J., Kellogg, W. A., and Erickson., T. 2002. "I'd Be Overwhelmed, but It's Just One More Thing to Do": Availability and Interruption in Research Management. . pp. 97-104.

[13] Liu, K. and Picard, R. 2005. Embedded Empathy in Continuous, Interactive Health Assessment. CHI Workshop on HCI Challenges in Health Assessment.

[14] McFarlane, D. C. 2002. Comparison of Four Primary Methods for Coordinating the Interruption of People in Human-Computer Interaction. Human-Computer Interaction. 17, 1, 63-139.

[15] Mokdad, A. H., Marks, J. S., Stroup, D., and Gerberding, J. 2004. Actual causes of death in the United States, 2000. JAMA. 291, 1238-45.

[16] Pollack, M. E., Brown, L., Colbry, D., McCarthy, C. E., Orosz, C., Peintner, B., Ramakrishnan, S., and Tsamardinos, I. 2003. Autominder: An Intelligent Cognitive Orthotic System for People with Memory Impairment. Robotics and Autonomous Systems. 44, 273-282.

[17] Prochaska, J. and Marcus, B. 1994. The Transtheoretical Model: Applications to Exercise. In R. Dishman, Ed., Advances in Exercise Adherence. Human Kinetics, Champaign, IL, pp. 161-180.