

What Would Jiminy Cricket Do? Lessons From the First Social Wearable

Timothy W. Bickmore

Northeastern University College of Computer and Information Science
360 Huntington Ave WVH202, Boston, MA 02115
bickmore@ccs.neu.edu
<http://www.ccs.neu.edu/home/bickmore/>

Abstract. Work towards the development of a “wearable conscience” that helps individuals make healthy decisions in their everyday lives is described. To be effective, such a system must be portable, sense key elements of the user’s environment, have knowledge of counseling and health behavior change techniques, be able to effectively communicate with the user, and have social competencies for maintaining an ongoing working relationship. The design of a prototype system is presented, along with results from a series of design and usability studies. Current and future directions for the research are also discussed.

Keywords: social agent, relational agent, embodied conversational agent, health behavior change, wearable computer.

1 Introduction

Imagine an external conscience that goes everywhere with you, experiences everything that you do, and whispers suggestions about the “right” thing to do in your ear. Such a system could help you make healthy lifestyle decisions, such as choosing healthy over unhealthy foods, taking the stairs rather the elevator, or avoiding situations in which significant social pressure would exist to engage in unhealthy behavior such as smoking or eating or drinking too much.

Such a system could have a significant impact on society. 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 [22]. In addition, adherence to prescribed medical 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 [13]. Each of these large classes of health behavior problems could be more effectively addressed if the right information and counseling could be delivered to someone at the moment they are trying to decide whether to engage in a healthy behavior or not. Importantly, however, this positive impact will only be realized if such a conscience is carefully designed so that it is not perceived to be annoying or intrusive.

1.1 Lessons from a Cricket

Perhaps the earliest conceptualization of such a “wearable conscience” was Jiminy Cricket, who accompanied Pinocchio on many of his escapades, offering advice on the long-term consequences of Pinocchio’s actions. In the original story, Jiminy’s advice was not readily accepted; in fact Pinocchio strikes and kills the cricket with a hammer in their first meeting after he provides unwanted counsel [8]. There are important lessons to learn from this, and Jiminy can be used as a case study to understand the requirements for a digital version of the wearable conscience:

- **Social and Relational Competency.** Given the lesson above, one of the most important abilities for such a system to have is the same range of social competencies any good human coach, counselor or friend has, including the ability to establish and maintain a good working relationship with the person being helped, and the knowledge of when and how to offer suggestions in order to maximize long-term benefit. Such a conscience must interrupt and speak persuasively to have any effect at all, yet long-term effectiveness is curtailed if the user simply smashes it with a hammer after the first few pieces of advice are delivered. Beyond simply providing reminders (e.g., in the form of audio alerts), such systems must be able to counsel and persuade users in non-trivial ways, thus requiring the use of natural language. It is also important that these systems be able to use the same range of verbal and nonverbal relational behaviors that people use to build and maintain good working relationships in face-to-face interaction. These behaviors include such things as facial displays of emotion to convey empathy and the use of social dialogue for building trust. Relational agents are systems that employ such competencies to establish long-term social-emotional relationships with users [5].
- **Adeptness at Interruption.** The system must be adept at knowing not only how, but when to interrupt a user in order to maximize long-term compliance with a health regimen. Knowledge of a user’s schedule and current task context would enable the system to know the precise moment when, for example, the user should be interrupted and reminded to take a morning dose of medication or given the suggestion to take a break from work and go for a walk.
- **Portability.** Obviously, the conscience should be portable so that it can accompany users wherever they go. This is important so that: (1) it can intervene at the moment the user is making a relevant decision, wherever and whenever that occurs; (2) it is available whenever the user actively seeks counsel on a relevant topic; and (3) it can give the user the perception that it shares in a significant portion of their life, important for the establishment of credibility and a good working relationship.
- **Sensing Ability.** In order for the system to actively intervene, it must be able to detect when the user is at a point of decision making, or perhaps has just started to engage in an unhealthy behavior. Different health behaviors require different sensors, but some examples that have been developed include: accelerometers for detecting physical activity [15]; microphones for detecting chewing [1]; smoke detectors for identifying cigarette lighting; and GPS for detecting proximity to locations known to present the user with difficult health choices (e.g., stairs vs. elevator, fast food restaurant, grocery store, etc.).
- **Persistence.** In order to establish long-term working relationships with users, and to effectively change behavior over time, a wearable conscience must retain mem-

ory of past interactions with users. Persistent memory should ultimately be represented as an episodic store recording details of all (or key) past interactions with users. At a minimum, however, it can be designed to record specific facts that can be referenced in future conversations. Examples in the physical activity coaching domain include remembering the name of a user's walking buddy or favorite walking location, as well as purely social (off-task) facts, such as the user's favorite television program and whether they had any big plans for the upcoming weekend .

- **Health Behavior Change Counseling Competency.** Finally, it is not enough that the system recognize when a user is about to do something unhealthy and tell them not to. Over the last few decades the field of behavioral medicine has produced a range of theories and counseling techniques for achieving effective health behavior change over time [11], and at least some of this knowledge must be codified into the system.

2 Related Work

Although there is a significant and growing literature on technologies to promote health behavior change, very few of these technologies involve real-time or wearable interventions. One notable exception is a body of research on reminder systems for individuals with cognitive impairment using pagers [14] and PDAs [18].

Several systems have also been developed to provide older adults with real-time reminders to perform various activities of daily living (including medication taking). The ILSA system used automated phone to calls to provide real-time reminders to older adults living alone, but the calls were not always effective and users did not like them [12]. 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") [23]. Preliminary evaluation indicated that acceptance among older adult users was high, although efficacy results have not been reported.

2.1 Handheld Conversational Agent-based Health Interventions

Johnson, *et al.*, developed DESIA, a psychosocial intervention on a handheld computer that features an animated conversational agent who uses balloon text and optional recorded speech output for the agent utterances [17]. Outcome evaluations have yet to be reported.

2.2 Task Interruption Studies

There has 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 the concern here is primarily on user responsiveness to the interruption.

A number of studies have shown that, in terms of supporting human performance of all kinds, negotiation-based 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 [20,21,24]. 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 [9].

Arroyo, *et al.* found that different interface modalities (e.g. heat, light, sound, odor) carry varying degrees of “disruptiveness” [2]. However, they did not investigate different stimuli within a modality (e.g., multiple sounds), nor did they measure short or long-term compliance to an interruption-based request.

There is also evidence that the use of empathy in interruptions can create a more positive user experience. Liu and Picard developed 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 [19]. The use of empathic language in the system prompt was varied within subjects, who showed (via self-report) 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, they also 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 Prototype System

My research group at Northeastern University has been working to develop a general purpose relational agent interface for use on handheld computers that can meet the requirements set forth in Section 1.1 (see Fig. 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, which are displayed in a text balloon rather than using speech, for privacy reasons. The words in the agent utterance are individually highlighted at normal speaking speed and the nonverbal behavior is displayed in synchrony. User inputs are constrained to multiple choice selections and time-of-day specifications at the bottom of the display.



Fig 1. PDA and Relational Agent Interface

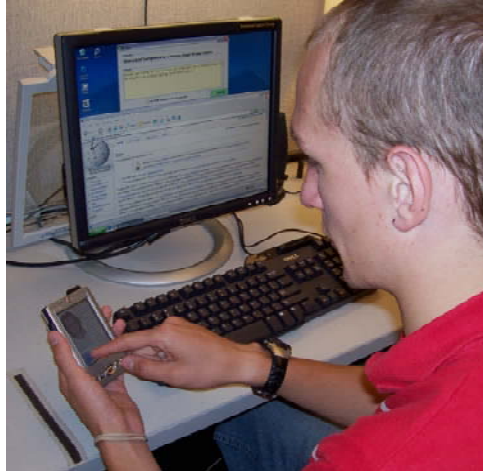


Fig 2. Experimental Setup for Interruption Studies

Interaction dialogues are scripted in an XML-based hierarchical state-transition network, which allows for rapid development and modification of system behavior. 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 [7], which automatically adds specifications for agent nonverbal behavior.

Interruption behavior can be very flexibly defined using a variety of wait states and state transitions conditioned on events. During specified wait states, the PDA's display shuts off, and the interface remains dormant until some condition is met. Example conditions include specific times of day, changes in user behavior as measured by sensor input, or other factors. The particular modality of an interruption can consist of various combinations of audio tones and/or visual cues presented on an arbitrarily complex schedule.

The run-time software was developed entirely in Macromedia Flash, and we are using Dell Axim X30 Pocket PC computers for development and experimentation. We have also developed a custom case for the PDA, which can be worn either on the waist like a large pedometer or pager, or in a shoulder harness. The initial application domain for the handheld agent is exercise promotion using an integrated accelerometer so the agent can tell whether a user is currently exercising or not.

4 Preliminary Evaluation Studies

Preliminary design and usability studies of the PDA-based system have been conducted and are described in this section. A longitudinal field study is also underway, and is described in more detail in Section 5.1.

4.1 Modality Study

In the first study, the impact of four different agent display modalities on the ability of a PDA-based agent to establish a social bond with the user was evaluated [4]. The four versions evaluated were: (FULL) the full version of the animated interface (animation, text and recorded speech consisting of backchannels and discourse markers only); (ANIM) the animated interface without the speech; (IMAGE) the interface showing only a static image of the character; and (TEXT) the interface without any character. These modalities were evaluated in a counterbalanced within-subjects experiment in which a subject conducted a brief social dialogue with each agent then filled out a self-report questionnaire evaluating the agent.

Results from the 12 subjects indicated that social bonding and perceived caring were significantly greater in the two animated conditions (FULL and ANIM) compared to the other two conditions (see Fig 2). Credibility of the information delivered by the agent was also rated higher in the FULL and ANIM conditions, but not significantly.

4.2 Task Interruption Studies

For users who happen to be sitting idly when a reminder to perform a healthy behavior is triggered, the interruption 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 a recommended health behavior) depends crucially on what the user is doing at the time the interruption presents itself [15], in addition to many other factors such as the emotional state of the user [16] and the modality of the interruption [2].

To explore the range of possible interruption strategies that a PDA-based relational agent could use to maximize long-term health regimen compliance, a series of studies were conducted in the laboratory with the prototype system. For this purpose “wrist rests” were used as the health behavior to be promoted, and web searching and typing answers to questions on a desktop computer was used as the representative office task that a user may be engaged in when a health-related interruption occurs (Fig 2 shows the experimental setup). The PDA-based agent was programmed to discuss the importance of taking frequent breaks from typing in order to avoid repetitive stress injury and other upper body musculoskeletal disorders, and to interrupt users periodically with a request to rest their wrists. Prevention of such disorders is an important problem in its own right: 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 self-reported symptoms [10].

A dilemma was presented to subjects, in which the research assistant running the study periodically reminded them to complete their work on the desktop computer as quickly as possible, while the PDA-based agent explained the importance of taking wrist rests and periodically (twice per session) asked them to take breaks from typing for as long as possible. The dilemma could often be visibly observed in subjects’ behavior following each interruption (e.g., false starts at task resumption).

Primary measures used in the interruption studies included the duration of wrist rests taken by subjects, as a measure of short-term compliance (REST1 and REST2, dual-coded from video, with inter-rater reliability of 0.99), self-reported “desire to continue use” as a proxy for long-term adherence (CONTINUE), and other self-report assessments of the agent including: politeness of the agent (POLITE), satisfaction with the agent (SATISFIED), relationship with the agent (RELATION), and how much they liked the agent (LIKE).

4.2.1 Politeness and Compliance

A very insistent or annoying interruption may be effective at gaining compliance in the short term, but individuals may be likely to use the agent less frequently in the long run (or discontinue use altogether, as Jiminy Cricket discovered), resulting in an overall loss of compliance. On the other hand, an extremely polite interruption may have the inverse effect: it may not be very effective at gaining compliance at any particular moment (e.g., if the user is engrossed in a task), though individuals may be more likely to continue use of the agent over time. Thus, a curvilinear relationship between the perceived “politeness” of interruption strategies and long-term health regimen adherence was hypothesized, holding all other factors constant.

To test this hypothesis, a range of interruption techniques were compared that were designed to vary only in their perceived politeness by just varying the audio alarm used to signal the start of each PDA interruption. Four alert sounds were selected which varied from very polite (AUDIO1, a subtle “ping”) to very impolite (AUDIO4, a loud klaxon).

The study had a counterbalanced within-subjects design, and 29 subjects completed the protocol. Participants did perceive a difference in politeness across conditions (manipulation check), which varied directly with their ratings of how much they liked the advisor, how satisfied they were with the advisor, and their relationship with the advisor (greater politeness leading to closer relationships, Fig. 3). Desire to continue using the advisor varied directly with politeness ratings, confirming part of our hypothesis.

Fig. 4 shows short-term compliance behavior (rest time) for the two interruptions in each evaluation session. Rest time following initial exposure to each stimulus

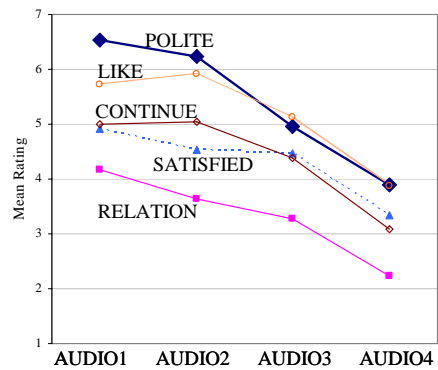


Fig 3. Self-Report Ratings

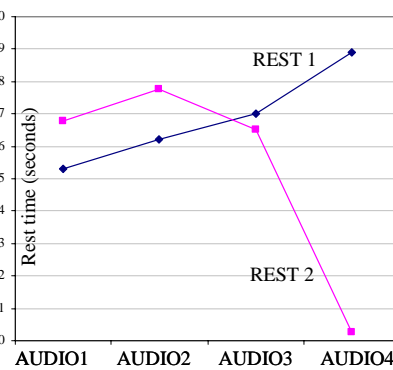


Fig 4. Compliance (Rest Times)

sound (REST1) did appear to vary according to our hypothesis (subjects rested longer for less polite sounds), but this variation was not significant. By the second exposure to each stimulus sound (REST2), the hypothesized longitudinal effects were already visible, with a curvilinear relationship between politeness and compliance, and the most annoying sound resulting in the shortest rest time.

4.2.2 Interruption Negotiation Strategies

In the second interruption study, the effect of several interruption coordination strategies on long-term health behavior adherence was compared. According to McFarlane, negotiation-based methods, which give subjects more control over the interruption process, should provide the greatest overall performance [20]. Accordingly, two negotiation methods were evaluated, one which gave users control over the start of the interruption (NEGOTIATED, via a “snooze” button) and one which gave subjects a warning that an interruption was about to occur (FOREWARN). Given that a wearable health counselor also requires social competencies, the efficacy of empathic interruptions on long-term adherence was also evaluated, following the results of Liu and Picard [19]. In this condition (SOCIAL), the agent apologizes for interrupting the subject, assesses their emotional state at the time of the interruption and, if warranted, provides empathic feedback. The final condition compared these three against a simple audio alarm from the previous study (BASELINE, identical to AUDIO3 above).

The study had a counterbalanced within-subjects design, and 16 subjects completed the protocol. Results indicated that subjects actually rested the longest in the SOCIAL condition, but the difference among conditions was not statistically significant. 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, and the pair-wise difference between BASELINE (lowest rating) and SOCIAL was significant, although the overall ANOVA was not. Interestingly, the NEGOTIATED condition was rated as the least effective, although actual rest times were second only to 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.

Thus, overall, the SOCIAL condition was both the most preferred and the most effective at gaining health behavior compliance.

5 Conclusion

The design studies completed to date indicate the importance of social and relational competencies in wearable systems that work with users to motivate health regimen adherence. Users demonstrated a preference for personified, animated counselors to less social media. Our hypothesis about the relationship between politeness and compliance in interruptions recommending a health behavior was supported, and the use of social behaviors such as empathy outperformed other interruption negotiation strategies from the recent literature on task interruption.

5.1 Future Work

There are many interesting directions of future research in this area. The next study planned involves manipulation of the perceived social distance between the user and advisor. Following Brown & Levinson's theory of politeness [6], we expect this to moderate the relationship between politeness and compliance, such that as familiarity grows, users actually expect and prefer less polite forms of interruption. Techniques from studies such as this on the *best way* to interrupt someone should be combined with techniques for determining the *best time* to interrupt someone (e.g., [15]) to design systems that have the overall best chance of promoting long-term health behavior regimen adherence.

The initial "wearable conscience" under development using the prototype system described in Section 3 is an exercise coach that is able to sense a user's walking behavior and intervene in real time in order to promote physical activity among sedentary adults [3]. Users are prompted daily to schedule times at which they commit to go for brisk walks, and if the system senses that they are not walking when they should be, it interrupts and engages them in a "problem solving" motivational dialogue. The longitudinal field study will compare this "just in time" intervention to one in which the same motivational dialogue is delivered during an end of day session. The study is intended to demonstrate the importance of the "just in time" concept, together with the importance of social and relational competencies, for health regimen adherence.

Acknowledgements. Thanks to Daniel Mauer for his work on the design and implementation of the experimental platform, and to Francisco Crespo and Thomas Brown for their help in running the evaluation studies. Jennifer Smith provided many helpful comments on the paper. This work was supported by NIH National Library of Medicine grant R21LM008553.

References

1. Amft, O., Stäger, M., Lukowicz, P., and Tröster, G.: Analysis of Chewing Sounds for Dietary Monitoring. *UbiComp* (2005) 56-72.
2. Arroyo, E., Selker, T., and Stouffs, A.: Interruptions as multimodal outputs: which are the less disruptive? . 4th IEEE International Conference on Multimodal Interfaces (2002) 479-482.
3. Bickmore, T., Gruber, A., Intille, S., and Mauer, D.: A Handheld Animated Advisor for Physical Activity Promotion. American Medical Informatics Association Annual Symposium (2006)
4. Bickmore, T. and Mauer, D.: Modalities for Building Relationships with Handheld Computer Agents. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI) (2006)
5. Bickmore, T. and Picard, R.: Establishing and Maintaining Long-Term Human-Computer Relationships. *ACM Transactions on Computer Human Interaction* 12, 2 (2005) 293-327.
6. Brown, P. and Levinson, S. C., *Politeness: Some universals in language usage*, Cambridge University Press, Cambridge (1987).
7. Cassell, J., Vilhjálmsón, H., and Bickmore, T.: BEAT: The Behavior Expression Animation Toolkit. *SIGGRAPH '01* (2001) 477-486.

8. Collodi, C.: Adventures of Pinocchio. (1881). English translation available at <http://www.gutenberg.org/etext/500> (checked 2/12/07).
9. Czerwinski, M., S. C., and Schumacher, B.: The effects of warnings and display similarities on interruption in multitasking environments. *SIGCHI Bulletin* 23, 4 (1991) 38-39.
10. Galinsky, T., Swanson, N., Sauter, S., Hurrell, J., and Schleifer, L.: A field study of supplementary rest breaks for data-entry operators. *Ergonomics* 43, 5 (2000) 622-638.
11. Glanz, K., Lewis, F., and Rimer, B., *Health Behavior and Health Education: Theory, Research, and Practice*, Jossey-Bass, San Francisco, CA (1997).
12. Haigh, K., Kiff, L., and Ho, G.: The Independent LifeStyle Assistant™ (I.L.S.A.): Lessons Learned. *Assistive Technology* 18, (2006) 87-106.
13. Haynes, R., McDonald, H., and Garg, A.: Helping Patients Follow Prescribed Treatment. *JAMA* 288, 22 (2006) 2880-83.
14. Hersh, N. and Treadgold, L.: Neuropage: The rehabilitation of memory dysfunction by prosthetic memory and cueing. *NeuroRehabilitation* 4, (1994) 187-197.
15. Ho, J. and Intille, S. S.: Using Context-Aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices. *CHI* (2005)
16. Hudson, J. M., Christensen, J., Kellogg, W. A., and Erickson, T.: "I'd Be Overwhelmed, but It's Just One More Thing to Do": Availability and Interruption in Research Management. . (2002) 97-104.
17. Johnson, W., LaBore, C., and Chiu, Y.: A Pedagogical Agent for Psychosocial Intervention on a Handheld Computer. *AAAI Fall Symposium on Dialogue Systems for Health Communication* (2004)
18. Kim, H., Burke, D., Dowds, M., Robinson Boone, K., and Park, G.: Electronic Memory Aids for Outpatient Brain Injury: Follow-up Findings. *Brain Injury* 14, 2 (2000) 187-96.
19. Liu, K. and Picard, R.: Embedded Empathy in Continuous, Interactive Health Assessment. *CHI Workshop on HCI Challenges in Health Assessment* (2005)
20. McFarlane, D. C.: Comparison of Four Primary Methods for Coordinating the Interruption of People in Human-Computer Interaction. *Human-Computer Interaction* 17, 1 (2002) 63-139.
21. McFarlane, D. C. and Latorella, K. A.: Coordinating the Interruption of People in Human-Computer Interaction. . *IFIP TC.13 International Conference on Human-Computer Interaction* (1999) 295-303.
22. Mokdad, A. H., Marks, J. S., Stroup, D., and Gerberding, J.: Actual causes of death in the United States, 2000. *JAMA* 291, (2004) 1238-45.
23. Pollack, M. E., Brown, L., Colbry, D., McCarthy, C. E., Orosz, C., Peintner, B., Ramakrishnan, S., and Tsamardinos, I.: Autominder: An Intelligent Cognitive Orthotic System for People with Memory Impairment. *Robotics and Autonomous Systems* 44, (2003) 273-282.
24. Robertson, T. J., Prabhakararao, S., Burnett, M., Cook, C., Ruthruff, J. R., Beckwith, L., and Phalgune, A.: Impact of interruption style on end-user debugging *CHI* (2004) 287 - 294.