Persuasion, Task Interruption and Health Regimen Adherence

Timothy Bickmore, Daniel Mauer, Francisco Crespo, and Thomas Brown

College of Computer and Information Science, Northeastern University
Boston, Massachusetts, USA
{bickmore, daniel, crespof}@ccs.neu.edu, brown.tho@neu.edu
http://www.ccs.neu.edu/research/rag

Abstract. Cueing strategies, such as real-time reminders, are among the most effective methods of persuading individuals to perform healthy behaviors such as taking their medication and exercising. However, these reminders often represent a task interruption for users who are engaged in work activities. This paper presents the results of a study which explores strategies for interrupting users at work to perform a healthy behavior, in which the primary outcome of interest is long-term adherence to a desired health behavior change regimen. We find that the degree of perceived politeness of interruptions is positively correlated with predicted long-term adherence, but negatively correlated with short-term compliance. We also find that, among several interruption coordination strategies previously explored in the literature, empathic interruptions are superior overall in gaining both short-term compliance and long-term adherence.

Keywords: Interruption, relational agent, embodied conversational agent, politeness, health 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 [21]. 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 [12]. 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. Timed reminders are examples of "cueing" or "stimulus control" processes, which involve changing an individual's environment so that it presents a conditional stimulus to perform a desired health behavior (e.g., keeping a gym bag by the door as a physical reminder to work out). It has been shown that such techniques

have been used by most individuals who have successfully changed their health behavior [23].

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 [14] in addition to many other factors such as the emotional state of the user [15] 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 (Fig. 1) [2,3]. In this effort, we are exploring different interruption modalities and strategies that can be used by the advisor to persuade users to perform a healthy behavior while they are working at routine office tasks. We are particularly interested in maximizing *long-term* adherence, summed over weeks or months of regular use. A very insistent or annoying interruption may be effective at gaining compliance in the short term, but individuals may be likely to use the device less frequently (or discontinue use altogether), resulting in an overall loss of adherence. On the other

Fig 1. PDA and Relational Agent Interface

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 device over time. Thus, we hypothesize a curvilinear relationship between the perceived "politeness" of interruption strategies and health behavior adherence, holding all other factors constant.

While we plan to conduct a longitudinal study of the PDA-based advisor, we chose first to conduct short-term studies of interruptions so that we could explore as many design options as possible. Consequently, we used self-reported "desire to continue

use" as a proxy for long-term adherence. While the validity of this measure with respect to actual long-term use needs to be established, there is an analogue in the study of human relationships, in which self-reported relational commitment has been found to be one of the best predictors for relationship longevity [4].

In the remainder of this paper we first present a review of related work in task interruption and health behavior promotion, followed by a description of the PDA platform used in our intervention work. We then describe the experimental setup used in the current effort, followed by a detailed description of the study we conducted. We conclude with a discussion of the study results and their implications and future research directions.

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 interventions. One notable exception is a body of research on reminder systems for individuals with cognitive impairment using pagers [13] and PDAs [17].

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 [11]. 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") [22]. Preliminary evaluation indicated that acceptance among older adult users was high, although results of any efficacy evaluations have not been reported.

2.1 Real-Time Health Behavior Compliance Studies

Goetz, *et al*, evaluated user compliance to the requests of a "nurse" robot to perform exercise [10]. 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).

The Breakaway system used an ambient display to provide real-time feedback to users when they had been seated at work for too long, in an attempt to get them to go for periodic walks [16]. Unfortunately, the system was only evaluated with a single user and only anecdotal results were 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 our focus is primarily on user responsiveness to the interruption.

Gillie and Broadbent [9] showed that even a very short interruption can be significantly disruptive and affect task performance, and Cutrell *et al.* [7] found that even an *ignored* interruption can negatively affect performance.

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 [19,20,24]. 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 [8].

Arroyo, *et al.* found that different interface modalities (e.g. heat, light, sound, odor) carry varying degrees of "disruptiveness" [1]. 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 presented a wearable system that periodically interrupted users and asked them (via text-based prompts) to annotate whether or not this was a good time to interrupt, and to specify their current stress level and activity [18]. 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 Experimental Platform

We have developed a general purpose social agent interface for use on handheld computers (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. 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 and the nonverbal behavior displayed in synchrony. User inputs are constrained to multiple choice selections and time-of-day specifications 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 [6], which automatically adds specifications for agent nonverbal behavior. In addition, each word of each utterance is processed by a viseme generator (based on the freeTTS text-to-speech engine [25]) that provides the appropriate sequence of mouth shapes the agent must form in order to give the appearance of uttering that word.

Interruption behavior can be very flexibly defined, since the scripting language supports 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. User failure to respond to an interruption (or any agent utterance) can also be handled in a flexible manner.

The architecture of the run-time system on the handheld is shown in Fig. 2. The actions of the system are primarily controlled by a finite state machine, which is built at run time according to the XML script. The Agent / Interface module comprises the relational agent itself (graphics, animations, audio, etc.), as well as areas for text output and user input in the form of clickable buttons which effect state transitions. During time periods in which the script does not explicitly specify agent actions, the idle action system takes over control of the agent, randomly performing various idle behaviors (eye blinks, posture shifts, etc.).

The run-time software was developed entirely in Macromedia Flash, and we are using Dell Axim X30 Pocket PC computers for development and experimentation.
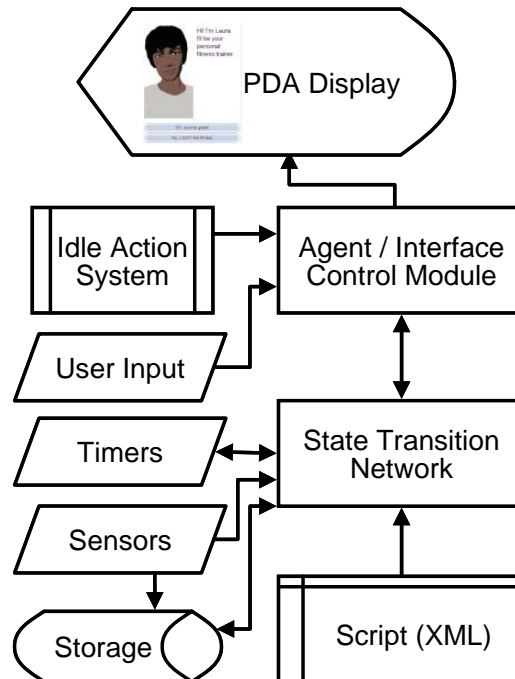

Fig 2. Run-time System Architecture

4 Experimental Method

We conducted a study to evaluate our hypothesis about the relationship between politeness of interruption and compliance, holding as many factors constant as possible. To do this, we evaluated the impact of a range of audio alert tones that were pre-selected to vary in perceived politeness. We used "wrist rests" as the health behavior to be promoted (secondary task), and web searching and typing answers to questions on a desktop computer as the representative primary task.

4.1 Experimental Design

The study is a four treatment, randomized, within-subjects design experiment, in which the sound used to signal the start of the interruption is varied between treatments (AUDIO1-AUDIO4), with the order of presentation completely counterbalanced. Each interruption strategy was evaluated using a different 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.

4.2 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 desktop computer. Following instruction on the primary task, subjects were told how to interact with the agent on the PDA, 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 subjects "Remember, we are keeping track of how many questions 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 primary task together with the advisor telling subjects the importance of wrist rests set up a dilemma for subjects that could often be visibly observed in their behavior following each interruption (e.g., false starts at task resumption, throwing hands in the air, rolling eyes, facial display of exasperation, etc).

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 Apparatus

Study subjects were seated at a desktop (primary task) computer with 17" color monitor, keyboard and mouse (see Fig. 3). 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 (overhead, left and right frontal), audio, and the computer screen video were continuously recorded during the study for subsequent analysis.

There was one program always running on the primary 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 program prevented users from pasting text; attempts to do so would produce a dialog box telling them they had to answer "in their own words". The program also presented a "DONE" button that subjects were instructed to push once they had finished answering the question. If subjects pushed this button before they had typed 40 words, another dialog box informed them how many more words they needed before their answer would be considered complete. Once they successfully finished a question, an audio tone was sounded and they were presented with the next question. At the end of the session (normally 10 minutes) a different tone was sounded and a dialog box appeared informing them that "TIME IS UP" for the primary task. Questions for the primary task program were designed to be answerable in 3-5 minutes; data from the study indicated that subjects actually took on average 3.2 (SD 1.2) minutes, and answered 2.0 (SD 1.13) questions per session, using an average of 59.4 (SD 14.2) words per answer. Sample questions are "Describe some of the features of HD DVDs." and "What is the greenhouse effect?".

Four PDAs were running the agent software described above, with four different 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."
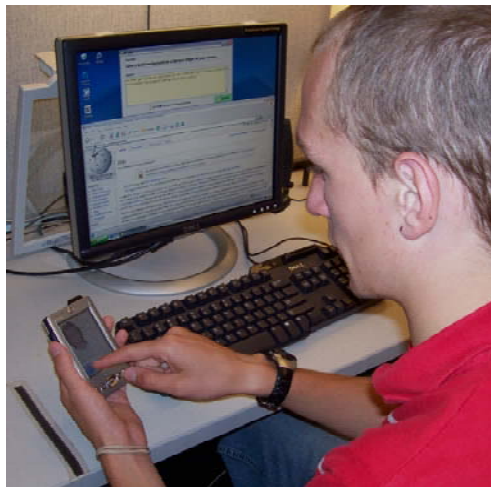

Fig 3. Experimental Setup

The only allowed user response is "OK, I'll rest now". After the user selects this response, the PDA display turns off until the next interruption.

4.4 Selection of Stimulus Alert Sounds

Four alert sounds were selected for use in the study, which varied from very polite (AUDIO1, a subtle "ping") to very impolite (AUDIO4, a loud klaxon). Selection was performed by first identifying 210 candidate sounds on the Internet using the FindSound.com search engine with keywords "alarm", "beep", "phone", "alert" and "chimes", and only selecting relatively brief sounds (< 5sec) of high quality. This list was manually reduced to 128 sounds by removal of clips we felt were duplicates. The resulting sounds were rated by seven members of our research group on a 7-point, "polite" to "annoying" scale, and four sounds that evenly spanned the range of averaged scores were selected for use in the study.

4.5 Measures

Self-report. Following each condition, subjects were asked to rate the wrist-rest agent on the self-report scales shown in Table 1, covering various ratings of the agent as well as 'desire to continue' using the agent.

Table 1. Self-Report Measures for Rating Interruption Methods

Measure	Question	Anchor 1	Anchor 7
POLITE	How polite was the advisor?	not polite	extremely
ANNOYING	How annoying was the advisor?	not annoying	intolerable
CONTINUE	How much would you like to continue working with the advisor?	not at all	very much

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 (using SPSS single measure 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 (TASKS).

4.6 Participants

Twenty-nine 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). The average wrist-rest behavior stage of change (adapted from [23]) was 'preparation', but subjects spanned the entire scale from 'pre-contemplation' to 'maintenance'.

4.7 Results

Comparison among study conditions was performed using repeated measure ANOVAs in SPSS.

Results are shown in Table 2. Most differences across conditions are highly significant (Fig. 4). Subjects did perceive a significant difference in politeness/annoyingness across conditions (manipulation check). LSD post hoc analysis indicated significant differences among all pairs of conditions except for AUDIO1 and AUDIO2.

Desire to continue using the advisor varied directly with politeness ratings, confirming part of our hypothesis. Post hoc analyses indicated significant differences between AUDIO4 and the other conditions, but not among any other pairs of conditions.

Fig. 5 shows short-term compliance behavior (rest time) for the two interruptions in each evaluation session. Rest time following initial exposure to each stimulus sound (REST 1) did appear to vary according to our hypothesis (subjects rested longer for more annoying sounds), but this variation was not significant. By the second exposure to each stimulus sound (REST 2), 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. Post hoc analyses on REST2 indicated significant differences between AUDIO4 and the other conditions, but not among any other pairs of conditions.

There were no significant effects of perceived politeness on primary task performance.

Table 2. Study Results, Mean(SD)

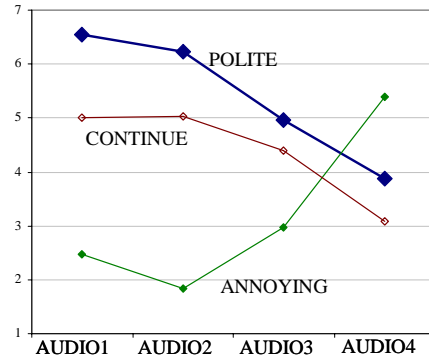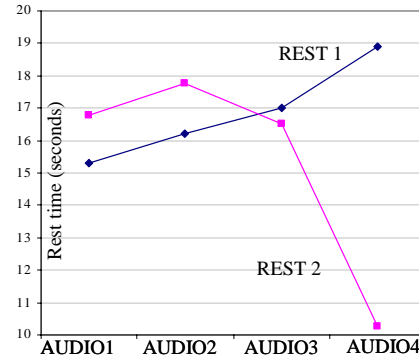| Measure	AUDIO1	AUDIO2	AUDIO3	AUDIO4	Sig.
POLITE	6.54(1.63)	6.23(1.58)	4.96(2.09)	3.89(2.66)	p<.001
ANNOYING	2.48(1.76)	1.84(0.85)	2.96(2.09)	5.40(2.77)	p<.001
CONTINUE	5.00(2.51)	5.04(2.46)	4.39(2.30)	3.08(2.40)	p<.001
REST 1	15.30(3.06)	16.16(3.12)	17.02(3.29)	17.78(3.71)	n.s.
REST 2	16.79(3.08)	17.76(3.41)	16.52(3.57)	10.25(2.62)	p<.05
TASKS	3.69(1.29)	3.45(1.15)	3.38(1.02)	3.38 (1.12)	n.s.

Fig 4. Self-Report Measures

Fig 5. Compliance (Rest Times)

5 Conclusion

Overall, our hypotheses regarding the relationship between politeness and compliance were supported. We were surprised to find that compliance would drop off so quickly with a very annoying interruption, but this was further confirmed by video analysis showing several subjects shutting off the second exposure AUDIO4 alarm with a disgusted expression and continuing to work without stopping. In sum, this study indicates that an appropriate level of politeness must be used when interrupting and persuading users in order to maximize long-term effectiveness.

Future Directions

There are many interesting directions of future research in this area. Our next study involves manipulation of the perceived social distance between the user and advisor. Following Brown & Levinson's theory of politeness [5], we expect this to moderate the relationship between politeness and compliance, such that as familiarity grows, users are more accepting of less polite forms of interruption. Our planned field study of the Just-in-Time advisor agent will also provide feedback on the long-term efficacy of the interruption strategy we finally implement for the system. Finally, 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., [14]) to design systems that have the overall best chance of promoting long-term health behavior regimen adherence.

References

1. 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.
2. Bickmore, T. and Mauer, D.: Modalities for Building Relationships with Handheld Computer Agents. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI) (2006).
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. Brehm, S., Intimate Relationships, McGraw-Hill, New York (1992).
5. Brown, P. and Levinson, S. C., Politeness: Some universals in language usage, Cambridge University Press, Cambridge (1987).
6. Cassell, J., Vilhjálmsson, H., and Bickmore, T.: BEAT: The Behavior Expression Animation Toolkit. SIGGRAPH '01 (2001) 477-486.
7. Cutrell, E., Czerwinski, M., and Horvitz, E.: Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. INTERACT'01 (2001) 263-269.
8. 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.
9. Gillie, T. and Broadbent, D.: What makes interruptions disruptive? A study of length, similarity and complexity. Psychological Research 50, 1 (1989) 243-250.
10. Goetz, J., Kiesler, S., and Powers, A.: 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 (2003)
11. Haigh, K., Kiff, L., and Ho, G.: The Independent LifeStyle AssistantTM (I.L.S.A.): Lessons Learned. Assistive Technology 18, 2006) 87-106.
12. Haynes, R., McDonald, H., and Garg, A.: Helping Patients Follow Prescribed Treatment. JAMA 288, 22 (2006) 2880-83.
13. Hersh, N. and Treadgold, L.: Neuropage: The rehabilitation of memory dysfunction by prosthetic memory and cueing. NeuroRehabilitation 4, 1994) 187–197.
14. Ho, J. and Intille, S. S.: Using Context-Aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices. CHI (2005)
15. 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.
16. Jafarinaimi, N., Forlizzi, J., Hurst, A., and Zimmerman, J.: Breakaway: An Ambient Display Designed to Change Human Behavior. CHI 2005 (2005) 1945-1948.
17. 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.
18. Liu, K. and Picard, R.: Embedded Empathy in Continuous, Interactive Health Assessment. CHI Workshop on HCI Challenges in Health Assessment (2005)
19. 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.
20. 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.
21. 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.
22. 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.

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

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.

25. Walker, W., Lamere, P., and Kwok, P.: freeTTS opensource repository. at http://freetts.sourceforge.net/ (2006).