Embedded Empathy in Continuous, Interactive Health Assessment

Karen K. Liu
Microsoft Corporation
1 Microsoft Way
Redmond, WA USA
+1-425-705-2079
kkliu@microsoft.com

Rosalind W. Picard
MIT Media Laboratory
20 Ames St. E15-020a
Cambridge, MA 02142 USA
+1-617-253-0611
picard@media.mit.edu

ABSTRACT

As an increasing number of new technologies are turning a strong focus on health assessment applications, new engineering and design challenges emerge. Challenges such as inference, modeling, data mining, and feedback for long-term usage arise. This paper argues that embedding empathy into the design of these interactive systems can potentially be vital in the acceptance and success of these types of technologies. This paper discusses three pieces of work that illustrate that designing systems that are intentionally empathetic can play a significant role in creating a better user experience in human-computer interactions.

1. INTRODUCTION

There are a variety of emerging health applications that monitor a person's physiological signals over time (Tran 2000; Miller 2001; Cole 2002) -- from a growing interest to monitoring elders in cognitive decline (Morris, Lundell et al. 2003), to systems that monitor and understand stress. Many of these systems may remain implicit monitoring systems, relying solely on sensed data, such as a diabetic (Kinsella 1999) or cardiac monitoring system that relays information to a central server or alerts friends or family in warning situations. However, a subset of these types of health applications will assess affect and medical states and provide feedback to users, thus, requiring interaction with the user. The development of these types of embedded health assessment technologies for long-term interactions present many significant design and engineering challenges. Since the success of such devices is dependant on acceptance from users over time, it is vital that these systems adaptively respond with dialogue and relational strategies that specifically address any disruption the system incurs upon the user. This paper argues that by embedding empathy into these systems we can design for enjoyable, long-term interactions in health assessment systems.

The rest of this paper describes work on embedding empathy in systems (section 2) and conclusions (section 3).

2. EMBEDDED EMPATHY

The following three pieces of work give evidence to show that empathetic systems can play key roles in contributing to a better user experience. We will briefly highlight the first two, and then focus on the third, which is the newest.

2.1 Responding to User Frustration

Klein appears to have been the first to deliberately design a computer system to respond empathetically to a frustrated user (Klein, 1999). His technique involved a simple scripted dialogue, and was evaluated with 70 subjects in a computer task designed to be frustrating, followed by one of three responsive conditions. All three conditions were friendly and chatty, and about the same length, while only one condition responded empathetically to a person's expression of frustration (the two controls either 1. ignored it or 2. asked about it but didn't respond.) A strong behavioral result was obtained: subjects in the empathetic condition engaged significantly longer in a follow-on task than did subjects in the two controls. The follow-on task was such that time spent on it was likely to be related to reduced level of frustration; hence, the behavioral results suggested that the computer's empathy significantly helped people handle the frustration the computer had induced earlier (Klein, Moon et al. 2002).

2.2 Relational Agents

Relational agents are computational artifacts designed to build long-term, social-emotional relationships with their users (Bickmore 2003; Bickmore and Picard 2004). Bickmore's Laura agent in the FitTrack system used relational strategies – sensitivity to your affective state, remembering your previous interactions, and immediacy behaviors – to build and maintain a relationship with users who were undergoing a month-long program to increase their exercise levels. Laura was the first such relational agent designed to maintain a long-term relationship with a user and (compared to a nearly identical but non-relational agent) was rated by users significantly higher on likeability, trust, respect, feelings that it cared for them, and willingness to continue interacting with it. These ratings held both after one week and after one month of interacting approximately every other day with the agent. While several factors (in addition to empathy) contributed to the impact of the relational agent, it is reasonable to expect that the agent's empathy played a key role in several of the significant findings, such as the finding that users felt significantly more "cared for" by the relational agent.

2.3 A Mobile System for Health Annotation

An interactive, health application has been developed for data collection, annotation, and feedback that explores using empathetic dialogue to facilitate a more satisfying and less stressful user experience.

2.3.1 System Design

The system has been developed on a mobile platform and uses affect-sensitive strategies to engage users and allow for real-time annotation of stress, activity and timing information through text and audio input. The platform supports continuous, wireless, and non-intrusive collection of heart signal data, accelerometer, and pedometer information, as well as automatic labeling of location information from context beacons. Figure 2-1 shows the overall architecture for the system.

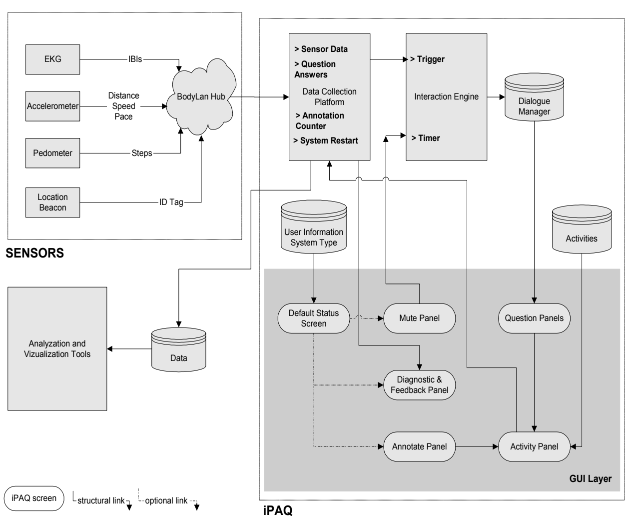

Figure 2-1 System Architecture

The system consists of a sensor layer that communicates to a *Data Collection Platform*. The *Interaction Engine*, which schedules interactions to interrupt the user for annotations either through a timer or through triggers from the sensor information is based off of the Context Aware Experience Sampling (CAES) Engine (Intille, Rondoni et al. 2003) developed at MIT. The CAES system aims to improve upon the Experience Sampling Method (ESM) (Csikszentmihalyhi and Larson 1987) -- an *in situ* sampling method -- by using sensing technology to automatically detect events that trigger sampling. The Interaction Engine uses a *Dialogue Manager* to choose from a set of different interaction scripts and possible responses to user input. A GUI layer receives input from scheduled and triggered interactions, as well as the possible question/response scripts, and interacts with the user through different GUI screens. Figure 2-2 shows example GUI screens used in this system.

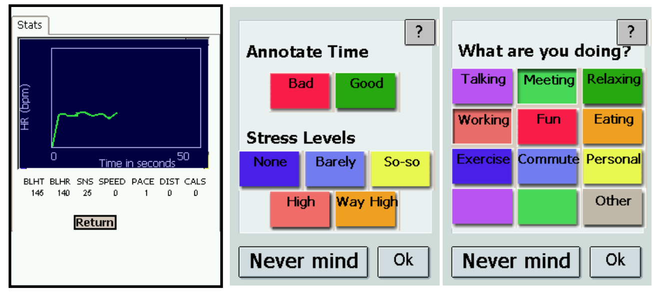

Figure 2-2 GUI Feedback and Annotation Screens

Figure 2-3 shows a sample empathetic interaction and a sample non-empathetic interaction. Both systems employed minimal relational techniques (e.g. social greeting, friendly); however, only one system responded directly to a user's stress levels with empathetic responses.[1]

```
S: Morning, Jane!
S: Do you have a minute?
U: Yes.
S: You know the drill -- feeling stressed?
U: Its there - but not the worst.
S: Wish it was better.  Hope things start
looking up.
S: Thanks so much for all your input.
```

```
S: Morning, Jane!
S: Do you have a minute?
U: Yes.
S: You know the drill -- feeling stressed?
U: Its there - but not the worst.
S: Thanks so much for all your input.
```

Figure 2-3 Sample Empathetic Interaction (top) Sample Non-Empathetic Interaction (bottom)

The sensor hardware, shown in Figure 2-3, consists of a set of custom-developed physiological sensors from Fitsense Technology (Fitsense 2004) – a Pulser for electrocardiogram (ECG) information, a Footpod for accelerometer information, a Pacer for pedometer information, and Location Beacons for environment context information. A BodyLAN Hub (BLH) connects

[1] There was one other difference between the empathetic and non-empathetic systems: the interruption timing algorithms. For the empathetic system, interruptions often triggered off of large changes in the heart signal and off of context changes. While this was initially designed to be of benefit, it wound up interrupting people a lot more which one would predict would lead to users not wanting to use this system for the final session, but this was not what was found.

the iPAQ to the different sensors on a BodyLAN wireless network.

Figure 2-4 Sensor System

2.3.2 Experiment

The system has been evaluated with ten subjects who used either the empathetic or non-empathetic system for four days, then used the opposite system for another four days, and finally, were asked to choose which system to continue interacting with for the last four days.

The hypotheses of this experiment included:

- Subjects will find the empathetic system to be less disruptive and frustrating to use and will have a better user experience while using the empathetic system.
- Subjects will choose to continue working with the empathetic system over the non-empathetic system.
- Users will report fewer perceived interruptions if the system is less stressful to use.

Measures that were used to evaluate the system included self-report ratings as well as a new measure we propose, the *relative subjective count (RSC)*. Relative subjective count is inspired by the relative subjective duration (Czerwinski, Horvitz et al. 2001) – a metric that takes the user's estimated time to complete a task divided by total time to complete the task as an implicit probe for measuring user frustration or satisfaction. In this study, the user's estimated number of interruptions (collected during each end-of-day log) divided by the actual number of interruptions was used as a measure for probing user frustration. If the subject greatly overestimated the number of interruptions, they were considered to be more frustrated than if they underestimated the number of interruptions. Since subjects were led to believe that there were three sessions in the study, the strongest behavior measure for evaluating user preference would be that the user chooses to work with the empathetic system for the final session.

2.3.3 Results

The empathetic system was rated as significantly less stressful on users t(14), $p = .26$[2], and seven out of ten of the subjects chose to continue working with the empathetic system after having used both systems.

Users were also asked to answer a series of questions based on their experience with the stress awareness study overall. The hypothesis here is that the subjects who ended with the responsive system would *remember* having a better experience overall even if in actuality they didn't (Redelmeier and Kahneman 1996). Subjects who ended with the empathetic system rated that their desire to continue with the stress awareness system was significantly higher than subjects who ended with the non-empathetic system, t(8), $p < .0008$.

To measure daily perceived frustration from interruptions, each user was asked to fill out their interrupt estimate in the end-of-day logs each night. There was a significant difference in low RSC, t(45), $p < .00035$, while using the empathetic system versus the non-empathetic system. The RSC (number of times people thought they were interrupted vs. how many times they actually were) was significantly lower in the empathetic condition. We hypothesize that RSC correlates with user frustration. The strongest behavior measure was that seven out of ten subjects chose to continue working with the empathetic system for the final session. The final system preference for all users is shown in Figure 3-3.

The strong preference for the empathetic system, lead us to believe that specifically embedding empathy into a system that directly addresses a user's affective state can facilitate a less frustrating and more enjoyable experience over time, even in tasks that are highly disruptive.

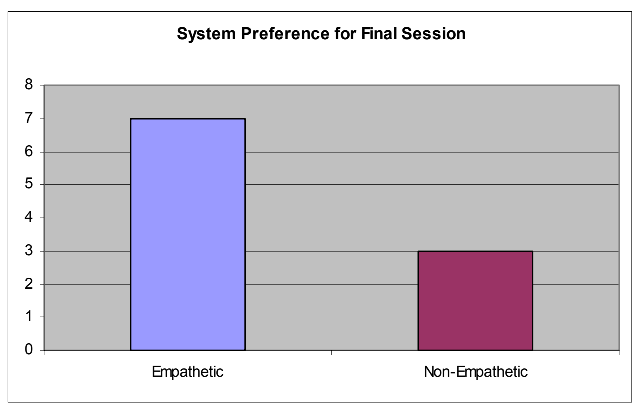

Figure 2-5 System Preference for Final Session

[2] The responses of two subjects were not included because they experienced an unusually high number of interruptions in the empathetic case as a result of a bug in the system.

4. CONCLUSION

While there is still significant work to be done in investigating the effects of embedding empathy into interactive systems, these three examples support the idea that embedding empathy into health technology systems that interact with people in ongoing ways can potentially be vital in the acceptance and success of these types of technologies.

ACKNOWLEDGMENTS

Thank you to Stephen Intille, Eric Horvitz, and Tim Bickmore for their insight and feedback for this work. We also thank Tom Blackadar, Joe Wronski, and Bala Biyer at Fitsense Technology for their help with the hardware. This research was supported in part by the Media Lab's Things That Think consortium.

REFERENCES

Bickmore, T. (2003). Relational Agents: Effecting Change through Human-Computer Relationships. Media Arts and Sciences. Cambridge, Massachusetts Institute of Technology.

Bickmore, T. and R. Picard (2004). "Establishing and Maintaining Long-Term Human-Computer Relationships." ACM Transactions on Computer Human Interaction (ToCHI).

Cole, A. M., B.S., Tran, Binh Q., Ph.D. (2002). Home Care Technologies for an Aging Population. Proceedings of the State of the Science Conference on Telerehabilitation.

Csikszentmihalyhi, M. and R. Larson (1987). "Validity and Reliability of the Experience-Sampling Method." Journal of Nervous and Mental Disease **175**: 526-536.

Czerwinski, M., E. Horvitz, et al. (2001). Subjective Duration Assessment: An Implicit Probe for Software Usability. Proceedings of IHM-HCI 2001, Lille, France.

Fitsense (2004). www.fitsense.com.

Intille, S., J. Rondoni, et al. (2003). A Context-Aware Experience Sampling Tool. Proceedings of the Conference on Human Factors and Computing Systems.

Kinsella, A. (1999). Improved Care for Diabetic Populations: The Need for Telehealthcare and Alternatives to Conventional Care Services. Background Paper for Home Care Technologies for the 21st Century.

Klein, J. (1999). Computer Response to User Frustration. Media Arts and Sciences. Cambridge, MA, Massachusetts Institute of Technology.

Klein, J., Y. Moon, et al. (2002). "This Computer Responds to User Frustration: Theory, Design, Results, and Implications." Interacting with Computers **14**: 119-140.

Miller, C. (2001). Automation as Caregiver: The Role of Advanced Technologies in Elder Care. Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society.

Morris, M., J. Lundell, et al. (2003). New Perspectives on Ubiquitous Computing from Ethnographic Study of Elders with Cognitive Decline. UbiComp, Seattle, WA.

Redelmeier, D. A. and D. Kahneman (1996). "Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures." Pain **66**(1): 3-8.

Tran, B. Q. (2000). Wireless Physiologic Monitoring: Applications and Implementation for Home Healthcare. Presented at the World Congress on Medical Physics and Biomedical Engineering, Chicago, Illinois.

ABOUT THE AUTHORS

Karen K. Liu is a Program Manager on the Visual C# team at Microsoft where she owns the C# interaction in the Visual Studio editor – including intellisense, smart tags, colorization, and code snippets. Prior to joining Microsoft in August 2004, she was a research assistant in the Affective Computing Research Group at the Massachusetts Institute of Technology (MIT) Media Laboratory. Her research interests include designing intelligent human-computer interactions in systems that can sense and respond to perceptual information with a focus on human and machine learning in health and educational applications. Liu holds a Masters degree in Media Technology from MIT and a Bachelor's degree in Computer Science from Wellesley College.

Rosalind W. Picard is founder and director of the Affective Computing Research Group at the MIT Media Laboratory and is co-director of the Things That Think Consortium, the largest industrial sponsor organization at the lab. She holds a Bachelor's in Electrical Engineering with highest honors from the Georgia Institute of Technology, and Masters and Doctorate degrees, both in Electrical Engineering and Computer Science, from the Massachusetts Institute of Technology. She is the author of over a hundred peer-reviewed scientific articles in multidimensional signal modeling, computer vision, pattern recognition, machine learning, and human-computer interaction, and is well known internationally for her pioneering book, Affective Computing, (MIT Press, 1997), laying groundwork for giving machines skills of emotional intelligence.