Constructing and Evaluating Sensor-Based Statistical Models of Human Interruptibility

James Fogarty
Human Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213
jfogarty@cs.cmu.edu

ABSTRACT
A person seeking a colleague's attention is normally able to quickly assess the colleague's interruptibility. In contrast, current computer and communication systems interrupt at inappropriate times or unduly demand attention because they have no model of human interruptibility. In my dissertation, I pursue system support for sensor-based statistical models of human interruptibility. Applications could use such models to negotiate interruptions at appropriate times, thus offering the potential for significant advances in human computer interaction.

Author Keywords
Interruptibility, context-aware computing, sensor-based interfaces, situationally appropriate interaction, managing human attention, machine learning.

INTRODUCTION
In order to provide system support for reliable sensor-based statistical models of human interruptibility, my dissertation pursues four primary contributions. I first contribute a series of studies to examine the feasibility of creating sensor-based statistical models of human interruptibility. I then contribute a system to support the use of sensor-based statistical models of human interruptibility in a variety of applications. I next contribute a notification application built using this system. Finally, I contribute an evaluation of the sensor-based statistical models of human interruptibility built by my system in the context of this notification application.

FEASIBILITY STUDIES
Many different sensors seem like they might relate to interruptibility, but the uncertainty surrounding their actual usefulness make it very likely that implementing and then evaluating them would result in significant time and resources being spent on sensors that are ill-suited or sub-optimal to predict interruptibility. My initial feasibility work instead collected 600 hours of audio and video recordings from the normal environments of four office workers and used a Wizard of Oz technique to simulate the presence of a variety of potentially useful sensors [2, 4]. While these recordings were collected, the office workers provided self-reports of their interruptibility at random time intervals. Statistical models based on the simulated sensor data distinguished "highly non-interruptible" situations from other situations with an accuracy as high as 82.4%, significantly better than a chance accuracy of 68.0% that could be obtained by always predicting that people were not "highly non-interruptible" ($\chi^2(1, 1344) = 31.13$, $p < .001$).

To evaluate this performance, the collected audio and video recordings were shown to human observers who estimated the interruptibility of the people in the recordings [2]. The human observers distinguished "highly non-interruptible" situations from other situations with an accuracy of 76.9%. This study shows that our sensor-based statistical models created from simulated sensors perform significantly better than human observers ($\chi^2(1, 3072) = 5.82$, $p < .05$), indicating that this approach can support models that are sufficiently reliable to support negotiated interruptions.

My most recent feasibility study extended these results by deploying real sensors into the normal environments of ten office workers with more diverse job responsibilities [3]. I deployed sensors to detect motion, whether the phone was off its hook, whether the door was open, closed, or cracked, the audio level in the office, the active computer application, and the level of mouse and keyboard activity. A model of all ten workers, ignoring the differences in their job responsibilities and working environments, had an accuracy of 79.5%, better but not significantly different from human observers ($\Delta z = 1.34$, $p \approx .18$). More accurate models resulted when I examined subsets of the data from subjects with similar job responsibilities and working environments. A model of two first-line manager subjects had an accuracy of 87.7% ($\Delta z = 1.34$, $p < .001$), a model of five researcher subjects had an accuracy of 81.1% ($\Delta z = 0.89$, $p \approx .37$), and a model of three interns working in offices shared with another intern had an accuracy of 80.1% ($\Delta z = 0.17$, $p \approx .86$). These results indicate that statistical models of human interruptibility created from real sensors can perform as well as or better than human observers for a variety of office workers.

SYSTEM SUPPORT FOR INTERRUPTIBILITY MODELS

While my feasibility studies have shown the potential for creating sensor-based statistical models of human interruptibility, significant effort is still required to include such models in applications. To address this, I am building AmIBusy, a system that provides mechanisms for logging sensor data, collecting observations of interruptibility, and automatically analyzing the collected sensor logs and interruptibility observations to create statistical models of human interruptibility. For applications desiring a generic estimate of interruptibility, the interruptibility observations necessary for a statistical model can be provided by a standard mechanism, such as once per day random prompts for a self-report. Applications interested in a more specialized estimate, such as whether a person will be receptive to a notification on a mobile device, can build models by providing observations of the specialized notion of interruptibility, as when a person reads a notification on a mobile device or indicates that it was inappropriate.

AmIBusy will be extensible to support a wide variety of sensors, including sensors running asynchronously, sensors running as a child thread, and complex data sources like those that might be provided by the Context Toolkit [1]. Because raw sensor data is typically inappropriate for creating statistical models of human interruptibility, AmIBusy automatically explores combinations of sensors and temporal sequences of sensor readings. This allows programmers to focus on the core functionality of a sensor, rather than on how it might interact with different sensors.

Finally, AmIBusy will combine data from multiple people with the goal of creating more reliable models of human interruptibility. Different sensors are predictive for people in part because they spend their time doing different things. AmIBusy will automatically cluster observations into situations based on their corresponding sensor readings. It will then build a statistical model of the interruptibility of people in each situation. This approach will allow a useful and meaningful aggregation of data from people with different job responsibilities and working environments.

NOTIFICATION APPLICATION

While my focus is on providing useful mechanisms to support models of human interruptibility in a variety of applications, I will also use AmIBusy to build a notification application. Notification applications are interesting in the context of interruptibility because, even though notification streams can be of high value, individual notifications are often of low value and can be considered disruptive when delivered at inappropriate times. This has generated significant interest in peripheral displays for notifications, but I am also interested in directly addressing the problem of delivering notifications at appropriate times. Using the responses of people to notifications, such as clicking for further information or indicating that the notification was delivered at an inappropriate time, my application will build a sensor-based statistical model of the receptiveness of people to its notifications. I will focus on sensors that can be deployed in software to a typical laptop, such as analyses of computer activity, audio as heard by a laptop's built-in microphone, and estimates of location based on network connectivity. While AmIBusy will support a wide variety of sensors, focusing on software-deployable sensors will allow a larger evaluation than would be feasible if I needed to deploy hardware.

EVALUATION OF MODEL CONSTRUCTION

My evaluation of sensor-based statistical models of human interruptibility created by AmIBusy in the context of my notification application will examine two primary issues. The first will be the ability of models to predict the receptiveness of a person to a notification. To examine this issue, I will deliver some notifications at random times. I expect that these randomly-timed notifications will be better received when the models indicates that a person is interruptible, as measured by a lower likelihood of indicating that a notification was delivered at an inappropriate time and a higher likelihood of clicking for additional information about the notification.

The second issue I will examine is how the specialized notion of interruptibility measured by feedback to my notification application relates to the more generic notion of interruptibility measured in my feasibility studies. I will have a subset of the users of my notification application provide self-reports of interruptibility like those collected in my feasibility work. I will then examine how well models based on feedback to the notification application can predict the self-reports. Because feedback like that given to the notification application can be collected much more subtly than self-reports, models based on this sort of data offer the potential for additional advances.

CONCLUSION

Sensor-based statistical models of human interruptibility offer the potential for significant advances in human computer interaction. My dissertation will pursue this potential by contributing a series of feasibility studies, a system to support applications using sensor-based statistical models of human interruptibility, a notification application created with this system, and an evaluation of the models created in the context of this notification application.

REFERENCES

1. Dey, A.K., Salber, D. and Abowd, G.D. (2001) A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. *Human-Computer Interaction (HCI) Journal*, *16* (2-4). 97-166.
2. Fogarty, J., Hudson, S., Atkeson, C., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J. and Yang, J. (2004) Predicting Human Interruptibility with Sensors. *To Appear, ACM Transactions on Computer-Human Interaction (TOCHI)*.
3. Fogarty, J., Hudson, S. and Lai, J. (2004) Examining the Robustness of Sensor-Based Statistical Models of Human Interruptibility. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2004)*, 207-214.
4. Hudson, S., Fogarty, J., Atkeson, C., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J. and Yang, J. (2003) Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2003)*, 257-264.