Who, What, and When:
Supporting Interpersonal Communication over Instant Messaging

Daniel Avrahami
Human-Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA
nx6@cs.cmu.edu

ABSTRACT
Instant Messaging, or IM, has been growing in popularity for personal and work-related communication. The limited awareness provided by current IM systems combined with the asynchronous nature of IM result in messages often arriving at inconvenient or disruptive times. The goal of my research is to alleviate some of the shortcomings of IM through the understanding and modeling of IM interaction in the context in which it takes place. This work uses three complementary steps: Predictions of responsiveness to IM (*when*), observations and predictions of interpersonal relationships (*who*), and use of properties of human dialogue to support balancing of responsiveness and performance (*what*).

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces; H1.2 [Models and Principles]: User/Machine Systems.

General terms: Design, Human Factors, Measurement

Keywords: Responsiveness, Availability, Awareness, Instant Messaging, IM, Predictive Models.

INTRODUCTION
Inter-personal communication through Instant Messaging, or IM, is gaining increasing popularity in the work place and elsewhere. A recent report estimated that 12 billion instant messages are sent each day. Of those, nearly one billion messages are exchanged by 28 million business users [17]. Despite its popularity, IM suffers from a number of shortcomings. Since IM is asynchronous, messages often arrive when a user is engaged in other tasks [16, 20]. The limited awareness of receivers' state, combined with the ease of initiating communication, thus often result in messages arriving at inconvenient or disruptive moments.

The goal of my research is to alleviate some of the shortcomings of IM through better understanding of factors affecting IM interaction in its context, and make use of this understanding to create predictive statistical models that support IM communication. In order to achieve these goals, I use three complementary steps:

1. Create accurate models that successfully predict responsiveness to incoming IM and investigate the factors affecting responsiveness (*when*).
2. Investigate the effect of interpersonal relationships on IM interaction, and use this knowledge to create statistical models that predict relationships (*who*)
3. Make use of basic properties of human dialogue to provide support for balancing of responsiveness and performance (*what*).

Background
IM programs, or clients, facilitate one-on-one communication between a user and their list of contacts, commonly referred to as buddies, by allowing them to easily send and receive short textual messages ("instant messages"). In its early days, IM gained its widest use supporting social communication, primarily among teenagers. Teens used IM primarily for socializing and planning social events, but also for coordinating schoolwork [10]. More recently, organizations are recognizing the value of IM and its benefits as a light-weight communication medium for both co-located and distributed teams, with uses ranging from quick questions and clarifications, coordination and scheduling, to discussions of complex work (see, for example, [16, 19]). As noted by Nardi et al., users may feel overwhelmed by incoming messages and, as a result, will sometimes choose to turn their IM client off, ignoring all IM communication altogether.

Indeed, incoming instant messages join an ever growing number of interruptions a person is exposed to. Those include interruptions external to the computer, such as people stopping by to ask a question, as well as interruptions from various computer applications (such as email alerts and calendar notifications). A large number of studies have been performed showing the negative effect of interruptions on people's performance. Gillie and Broadbent, for example, showed that even a very short interruption can be disruptive [9], while Cutrell et al. showed that even an ignored interruption can have a

UIST '06, October 15-18, 2006, Montreux, Switzerland

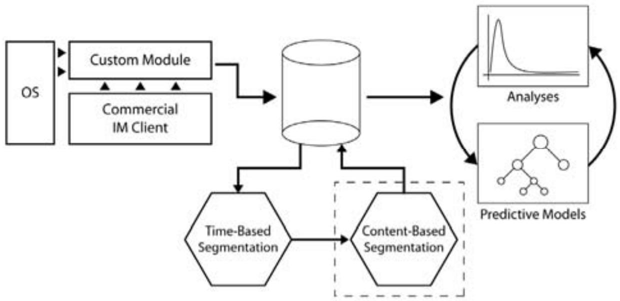

Figure 1: Process illustration: Data collection, processing, analysis and modeling.

negative effect [6]. As the use of IM is growing, and in particular in the work place, the inability to detect a buddy's state can often result in communication breakdowns with negative effects on both communication partners. If, however, we were able to accurately predict whether a user was likely to respond to a message within a certain period of time, then some of these breakdowns could be prevented.

In this document I describe completed and planned research. I start by describing a custom data collection mechanism that I implemented and a number of important privacy preserving measures used. I then describe the construction of predictive models of responsiveness to incoming IM. Next, I describe findings from analysis on the effect of interpersonal relationships on IM communication, and the use of these findings for the creation of predictive models. I finish by describing a tool that allows users to balance performance and responsiveness to IM.

DATA COLLECTION MECHANISM

My work is founded on the collection of field-data representing naturally occurring IM interaction in its context (Figure 1). These data are used to examine communication patterns and for the construction of predictive statistical models in support of communication. I have created a data collection mechanism, implemented as a custom plug-in module for Trillian Pro, a commercial IM client. This module is used to record all IM and desktop events. Participants are able to use this client for all their IM communication with their existing buddies, which allows for collecting high volumes of real IM and desktop interaction over extended time periods. In an ongoing data collection effort, I have so far collected approximately 5,200 hours of data, with nearly 90,000 instant messages. These messages were exchanged between 16 participants (both graduate students and industrial researchers) and over 400 buddies.

Privacy of Data

I have taken a number of measures to preserve, as much as possible, the privacy of participants and their buddies. The text of messages is not recorded unless I receive specific permission from the participants. Otherwise, messages are masked in the following fashion: Each alpha character is substituted with the character 'A' and every digit is substituted with the character 'D'. Punctuation is left intact. For example, the message "my PIN is 1234 :-)" is recorded as "AA AAA AA DDDD :-)".

When a participant opens a message window to a buddy for the first time (and that buddy is online), an alert is sent to the buddy notifying them of the participation in the study. Buddies of participants who had provided the additional permission to record the text of messages are notified with a different alert message that instructs them of a simple mechanism for temporarily masking messages.

Finally, for determining that two events were associated with the same buddy, we create a unique ID for each buddy (using an MD5 cryptographic hash) and store the ID of the buddy instead of the buddy-name itself.

WHEN: PREDICTING RESPONSIVENESS TO IM

As mentioned above, predicting *responsiveness* -- whether a user is likely to respond to an incoming message within a certain time period -- may be used to reduce disruptions and unsuccessful communication. Using the collected field-data described above, I have created a set of statistical models (using AdaBoosting on Decision-Tree models) that can predict, with accuracy as high as 90.1%, *responsiveness* to incoming instant messages [3]. Specifically, these models predict whether a user will respond to a buddy's attempt to start a new session -- I define an *IM session* to be a set of instant messages that are exchanged within a certain time delay between one another -- within 0.5, 1, 2, 5, and 10 minutes (see Table 1). A comparison showed that all these models perform significantly better than the prior probability baseline ($G^2(1,3161) \geq 916$, $p<.001$) (Prior probability represents the accuracy of a model that picks the most frequent answer at all times.)

Such models of responsiveness can be used in a number of different ways to support communication. Models can be used, for example, to automatically provide different "traditional" online-status indicators to different buddies (results from [1] suggest that, given information about the receiver, senders would be able, and willing, to time their messages to accommodate for the receiver's state). Alternatively, these models can be used to increase the salience of incoming messages that may deserve immediate attention if responsiveness is predicted to be low (such as in [2]). Additionally, models can be used by systems that show a list of potentially responsive experts to users who are looking for help or support, while hiding others.

Predict response within	30sec	1min	2min	5min	10min
Full Set	79.8	83.8	87.0	89.4	90.1
Buddy-independent	79.8	83.7	87.0	89.4	89.3
Baseline	54.7	55.9	63.8	72.0	75.4

Table 1. Accuracy (in %) of models compared to baseline by feature sets (Full vs. Buddy-independent) and prediction class (30secs, 1, 2, 5, and 10 minutes)

Buddy-Independent Models

Next, to understand the role that buddy-identity plays in predicted responsiveness, I created a second set of models that use only information about the participant, excluding any information about the buddy who initiated the session (*buddy-independent*). To my surprise, the buddy-independent models performed nearly as well as the first set of models (as high as 89.3%), and the differences were not found to be significant (see Table 1). These models are interesting also from an application's point of view, since buddy-independent models will predict, at any given moment, a single level of responsiveness, while my earlier set of models could predict different levels of responsiveness to different buddies.

Content-driven Session Segmentation

In the work presented above, I used time delay, rather than content, to determine whether two messages belonged to the same IM session. Not using content, however, may result in a number of problems. For example, a late reply may not represent the start of a new session, rather a part of the previous one. In order to achieve reliable identification of session boundaries, I plan to employ a two-step *superset-pruning* approach (see Figure 2). First, the delay between messages is used to segment the data into an overly large set of sessions. By choosing a short delay threshold most session boundaries will be identified, while at the same time, other gaps will be incorrectly identified as boundaries. In the second step I will use content analysis techniques, including assignment of dialogue-act labels, and identifying the use of pronouns, to examine the relationship between the end of each session and the beginning of the session that follows it. This will allow me to determine whether the boundary between two sessions was correctly identified or whether the two are in fact part of the same session, and should be merged. I will evaluate this process by comparing its results to manual coding of segment boundaries.

WHO: RELATIONSHIPS AND IM COMMUNICATION

A receiver's availability and responsiveness to communication may often depend on the identity or role of the communication partner. As more and more people use IM for social as well as work-related communication, I wanted to investigate the effect of relationship on basic characteristics of IM communication. Previous research (e.g., [7]) showed that interpersonal relationship type has significant effects on different aspects of communication, including the quality, purpose and perceived value of the communication.

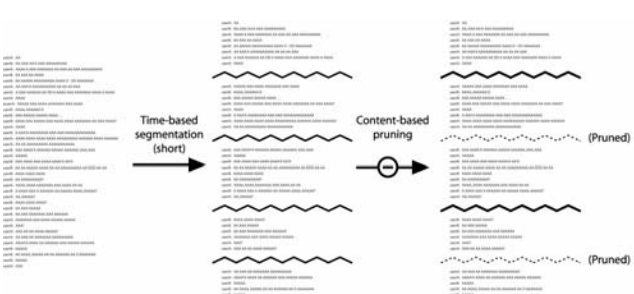

Figure 2: A two-step content-driven method for identifying IM session boundaries.

An analysis of my data showed the significant effect of relationship, independent of message content, on a number of basic communication characteristics (e.g., session duration, message length, exchange rate, etc.) [4]. I found, for example, that buddies in a social relationship had significantly longer sessions than buddies in a work relationship, exchanged more messages, and took more turns. Surprisingly, however, message exchange rate was significantly lower for buddies in a social relationship compared to mixed relationship (both work and social) (M=4.6 vs. M=6.2 messages per minute; p=.003) and marginally significant compared to work relationship (M=4.6 vs. M=6.0 messages per minute; p=.078). A possible explanation for this finding is that users focus less of their attention on conversations with social buddies but give more attention to conversations with buddies with whom they work. I also found that the length of messages exchanged between buddies in a work relationship were significantly longer, on average, than messages exchanged between buddies in a social relationship (M=38 vs. M=30 characters-per-message; p=.002). It is indeed possible that conversation between buddies in a work relationship is less casual and users construct their ideas more carefully before sending them.

Having investigated the effect of relationship on IM communication, I examined the ability to use these findings for the creation of statistical models that predict the relationship between users, based only on those basic communication characteristics, without the use of message content. Predictive models of the relationship between IM users could be used in a number of ways for augmenting IM clients, but could also be used to propagate predictions of relationships to other communication mediums. One of the models generated was able to predict, with accuracy of nearly 80%, whether a user and a buddy are in a work or social relationship. Significantly better (p<.001) than the prior probability of 52%. These models were generated as a two-step prediction process, based on nominal logistic regression (for more details see [4]).

WHAT: USING PROPERTIES OF HUMAN-DIALOGUE

In an attempt to alleviate the problem of IM disrupting work on an important task, or a user being forced to ignore incoming messages in order to maintain workflow, I have created a tool called QnA that augments a commercial IM client [2]. By monitoring the content of incoming and outgoing messages, QnA helps users identify messages that potentially require a quick response, and messages that they are expecting. More specifically, QnA provides users with lightweight notifications of incoming questions and incoming responses to their own questions. If, however, QnA determines that the user has already attended to the message, then its notifications are suspended. This allows users to smoothly transition between work-states while seeming responsive. To determine whether a message contains a question, it is compared against a large set of string matching rules. (The message is also matched against a second set of rules created to avoid identifying 'false-positives'). Determining that the user has attended to a message was achieved by including a delay period before a notification, during which, user interaction is observed.

RELATED WORK

In addition to research mentioned throughout this document, my research joins a growing body of work investigating the use of machine learning techniques to infer and predict human situation and action. Related research includes work on predicting importance and actions on email [11], likelihood of meeting-attendance [13, 18], appropriate times for interruptions [12, 14, 15], and presence [5]. Prior work, such as [12, 14], collected naturally occurring behavior as data, using participants' self reports as the labels of ground truth. Other work, such as [15], and [8] used the behavior of subjects participating in a lab experiment to train their models. In contrast with the work mentioned above (but similar to [5] and [13]), the work presented in this document involves the creation of predictive models trained using *naturally occurring human behavior*. An added benefit of using naturally occurring behavior as the source for learning is that a model deployed as part of a system would be able to continuously observe user behavior to train and improve its performance without requiring any intervention from the user.

CONCLUSIONS

I have presented planned and completed work on analysis and generation of predictive modeling in support of interpersonal communication over IM. This work's contribution to the HCI field spans both theoretical and applied aspects. From a theoretical point of view, this work provides insights into the factors that influence interpersonal communication patterns and responsiveness. At the applied level, this work provides predictive statistical models that can be used in many useful applications. Finally, this work promotes the creation of tools that use predictive models generated from naturally occurring interaction.

ACKNOWLEDGMENTS

I would like to thank my advisor Scott Hudson. This work was supported in part by the National Science Foundation under grants IIS 0121560 and IIS 0325351, and this material is based in part upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. NBCHD030010.

REFERENCES

1. Avrahami, D., Gergle, D., Hudson, S.E., and Kiesler, S. Improving the match between callers and receivers: A study on the effect of contextual information on cell phone interruptions. *Behaviour & Information Technology* (in press).
2. Avrahami, D., and Hudson, S.E. QnA: Augmenting an Instant Messaging Client to Balance User Responsiveness and Performance. In *Proceedings of CSCW 2004*. 515-518.
3. Avrahami, D., and Hudson, S.E. Responsiveness in Instant Messaging: Predictive Models Supporting Inter-Personal Communication. In *Proceedings of CHI 2006*. 731-740.
4. Avrahami, D., and Hudson, S.E. Communication Characteristics of Instant Messaging: Effects and Predictions of Interpersonal Relationship. In *Proceedings of CSCW 2006*. to appear.
5. Begole, J.B., Tang, J.C., and Hill, R. Rhythm Modeling, Visualizations, and Applications. In *Proceedings of UIST 2003*. 11-20.
6. Cutrell, E., Czerwinski, M., and Horvitz, E. Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. In *Proceedings of INTERACT 2001*. 263-269.
7. Duck, S., Turr, D., Hurst, M., and Strejc, H. Some evident truths about conversations in everyday relationships: All communications are not created equal. *Human Communication Research, 18*, 2 (1991), 228-267.
8. Fogarty, J., Ko, A.J., Aung, H.H., Golden, E., Tang, K.P., and Hudson, S.E. Examining Task Engagement in Sensor-Based Statistical Models of Human Interruptibility. In *Proceedings of CHI 2005*. 331-340.
9. Gillie, T., and Broadbent, D. What Makes Interruptions Disruptive? A Study of Length, Similarity, and Complexity. *Psychological Research, 50* (1989), 243-250.
10. Grinter, R., and Palen, L. Instant Messaging in Teen Live. In *Proceedings of CSCW 2002*. 21-30.
11. Horvitz, E., Jacobs, A., Hovel, D. Attention-Sensitive Alerting. In *Proceedings of UAI 1999*. 305-313.
12. Horvitz, E., Koch, P., and Apacible, J. BusyBody: Creating and Fielding Personalized Models of the Cost of Interruption. In *Proceedings of CSCW 2004*. 507-510.
13. Horvitz, E., Koch, P., Kadie, C.M., and Jacobs, A. Coordinate: Probabilistic Forecasting of Presence and Availability. In *Proceedings of UAI 2002*. 224-233.
14. Hudson, S.E., Fogarty, J., Atkeson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C., and Yang, J. Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. In *Proceedings of CHI 2003*. 257-264.
15. Iqbal, S.T., and Bailey, B.P. Leveraging characteristics of task structure to predict the cost of interruption. In *Proceedings of CHI 2006*. 741-750.
16. Isaacs, E., Walendowski, A., Whittaker, S., Schiano, D.J., and Kamm, C. The Character, Functions, and Styles of Instant Messaging in the Workplace. In *Proceedings of CSCW 2002*. 11-20.
17. Mahowald, R.P., *Worldwide Enterprise Instant Messaging Applications 2005-2009 Forecast and 2004 Vendor Shares: Clearing the Decks for Substantial Growth*. Technical Report, IDC Market Analysis, 2005.
18. Mynatt, E., and Tullio, J. Inferring Calendar Event Attendance. In *Proceedings of IUI 2001*. 121-128.
19. Nardi, B.A., Whittaker, S., and Bradner, E. Interaction and Outeraction: Instant Messaging in Action. In *Proceedings of CSCW 2000*. 79-88.
20. Voida, A., Newstetter, W.C., and Mynatt, E.D. When Conventions Collide: The Tensions of Instant Messaging Attributed. In *Proceedings of CHI 2002*. 187-194.