

# Work Rhythms: Analyzing Visualizations of Awareness Histories of Distributed Groups

James “Bo” Begole, John C. Tang, Randall B. Smith, and Nicole Yankelovich

Sun Microsystems Laboratories  
2600 Casey Avenue, MTV29-230  
Mountain View, CA 94043 USA  
+1 (650) 336-1109  
<firstname>.<lastname>@sun.com

## ABSTRACT

We examined records of minute-by-minute computer activity coupled with information about the location of the activity, online calendar appointments, and e-mail activity. We present a number of visualizations of the data that exhibit meaningful patterns in users’ activities. We demonstrate how the patterns vary between individuals and within individuals according to time of day, location, and day of the week. Some patterns augment the schedule information found in people’s online calendars. We discuss applications for group coordination (especially across time zones) plus opportunities for future research. In light of the popularity of instant messaging, this research identifies some of the benefits and privacy risks associated with the uses of online presence and awareness information.

## Keywords

Work rhythms, awareness, presence, instant messaging (IM), group calendaring, CSCW, sociology of time.

## AWARENESS OF WORK RHYTHMS

Much of our everyday work (and life) has underlying rhythms that give it structure. Often, people have regular times when they arrive at their workplace, break for lunch, and go home. Zerubavel’s [15] work on the sociology of time identified more fine-grained features of temporal regularity that can occur *within* a work day. He demonstrated how the awareness of the surrounding activities in some work environments, such as hospitals, give people a fairly accurate sense of time without looking at a clock.

Many factors in today’s work environment make it difficult to share such a sense of time among colleagues. Flexible work hours, telecommuting from home, and work teams distributed across geographic sites create situations where coworkers may be both physically separated and temporally shifted across time zones. These factors make it harder for work groups to share an awareness of the temporal rhythms of their colleagues that can help coordinate their activities.

Research projects have explored the use of video [7], audio, snapshots [2], and physical proxies [4] to provide shared awareness for distributed groups. Commercial products

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CSCW’02, November 16-20, 2002, New Orleans, Louisiana, USA.  
Copyright 2002 by Sun Microsystems, Inc. 1-58113-560-2/02/0011

have also helped restore awareness cues for distributed groups. Palen [11] studied how shared access to online calendars affects group coordination of activities. Nardi *et al.* [10] showed how the use of instant messaging (IM) at work has helped restore some sense of awareness among distributed workers and made it easier to establish contact with each other. In addition to these real-time awareness systems, Mynatt *et al.* [9] visualized the *history* of activity from sensors placed in an elderly occupant’s house to allow remote caretakers to monitor noticeable changes over time that might require follow up.

Our work on the Awarenex research prototype [13] focuses on restoring awareness cues to distributed work groups. We use this awareness and IM system in our own work group which is split between sites separated by 2500 miles and a three-hour time difference. In using Awarenex over time, we noticed that it not only provides an immediate sense of who is reachable for contact, but also helps us recognize typical long-term patterns of when our remote colleagues arrive and depart for the day, break for lunch, etc. That insight led us to explore what could be learned from analyzing the history of people’s online activity patterns. Our initial aim was to explore how patterns in people’s work activity would help identify convenient times to make contact (i.e., help reduce phone tag). The coordination patterns that emerged from our analysis suggested an extension of Zerubavel’s work to explore how groups distributed over distance and time zones can share a social sense of time.

This paper describes the analysis of records of computer and e-mail activity. Such exploration raises obvious concerns about privacy and electronic surveillance. We want to contribute to the ongoing public debate of those issues [1] by exploring what information can be inferred from online awareness data. With the popularity of IM systems and ubiquity of electronic sensors, it is important to proactively understand the uses of such information, both good and ill, so that the tradeoffs can be weighed appropriately. Because of the importance of these concerns, we discuss the privacy implications more fully in a later section.

## COLLECTING AND VISUALIZING AWARENESS DATA

For this research, we collected logs of computer interaction (e.g., keyboard, mouse activity), the location where that activity occurred, online calendar appointments, and e-mail activity for 20 users for up to ten months. The computer interaction logs came from our Awarenex prototype, which indicates if a user is actively using their computer input

devices. If the input devices are not used for more than a minute, Awarenex will report how long they have been inactive, updating every minute. This information often gives a useful sense of whether the user is “present”, suggesting whether it is a good time to attempt making contact. Awarenex also indicates the location where the user’s activity occurred (e.g., office, home, lab).

Since e-mail plays such a major role in group communication, we wanted to look for patterns in e-mail activity. To study e-mail activity, we inserted a proxy between the users’ e-mail clients and servers. The proxy detects when a user begins reading a message by noticing the retrieval of the body of a message from the user’s e-mail server. Although many e-mail clients automatically retrieve e-mail headers periodically, the retrieval of a message body is always the result of a user action to view the message for the e-mail clients used in this study (CDE Mailer and Netscape Messenger). The proxy also detects when a user sends an e-mail message by noticing the posting of a new message to the user’s e-mail server. Note that this data does not capture the time spent reading or composing e-mail, only the pulses of fetching and sending messages.

### Groups that Participated in the Study

We studied the activity logs of three different groups:

- Ourselves—the research group that designed Awarenex. Our group has two East Coast and three West Coast members (separated by a three-hour time zone difference). Our group’s records began in May 2001.
- Affiliated researchers—others in our research organization. All four members are located on the East Coast. Their activity records began in May 2001.
- Distributed support team—an Awarenex trial group with four East Coast and five West Coast members, one of whom uses two California sites. This group is unaffiliated with our research, providing independent end user feedback. Their activity records began in July 2001.

In addition to the Awarenex activity logs, we collected e-mail logs on five users since October 2001. All participants provided informed consent to record and study their activity for the purpose of this research.

### Limitations of the Activity Data

Some limitations of the activity data provide important context for this study. First, interaction with the computer is reported down to the minute—the system reports if the user has been active at any time during a minute. The data will not show any activity patterns that occur within a minute.

Another limitation is that activity data is only recorded when users are logged in to Awarenex. Most users regularly logged in while in their office or laboratories, but some users did not always log in when working on the computer from other locations (e.g., home).

More importantly, our activity data does not exactly indicate when people are available for communication in at least two ways. First, even when users are not interacting with their computer, they may still be nearby and thus reachable for communication. For example, if a user is reading a document (either within the office or on her computer screen without touching her mouse for more than a minute), Awarenex will show that she is inactive. Yet she is

reachable through her computer and could have responded to an incoming IM. In this way, the data underestimate the times that a user may actually be reachable for contact.

Second, a user’s interaction with their computer does not mean that they are necessarily *receptive* to communication or interruption. In fact, if they are busily working on their computer, they may be less receptive to interruption. Nevertheless, users are *reachable* when interacting with their computers, which is a prerequisite condition for contacting them. Patterns suggested by the activity information need to be taken in context with other cues to help determine if it is socially appropriate to attempt initiating contact.

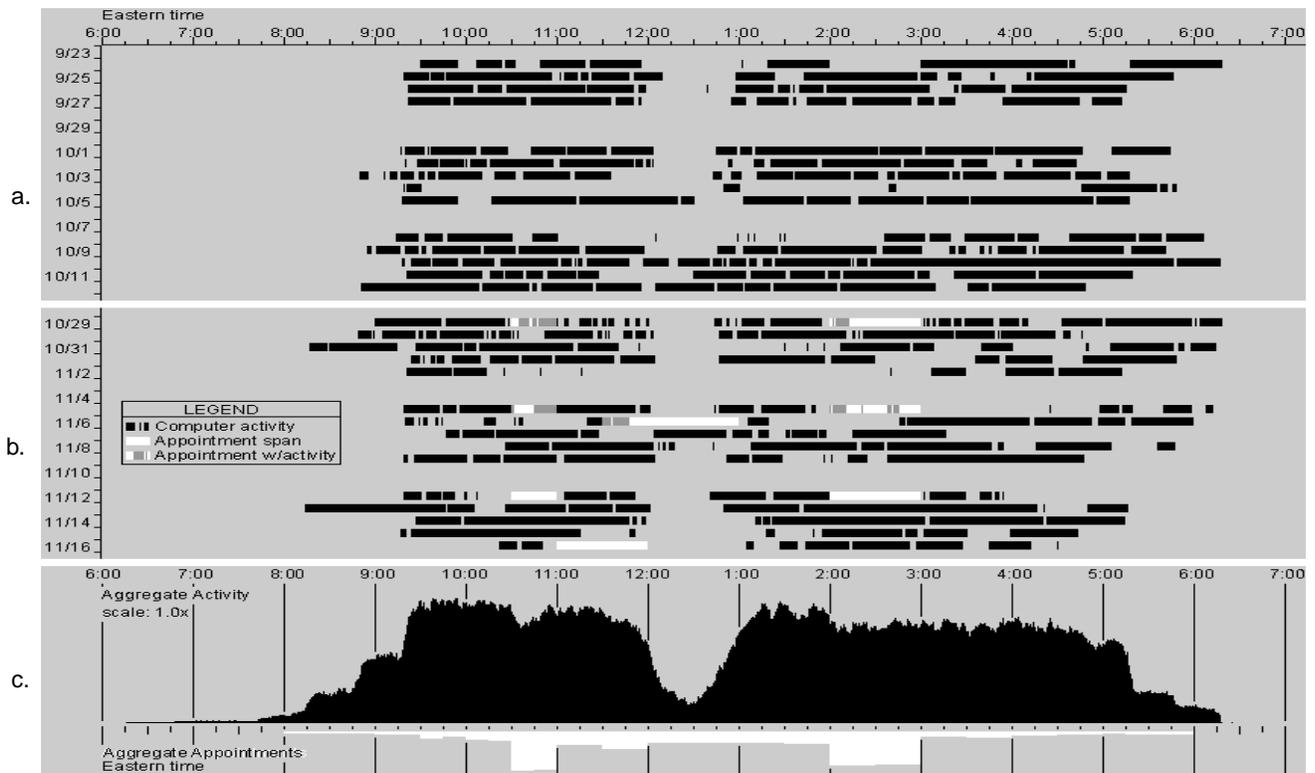
Real-time awareness systems, such as IM, also rely on activity data that does not exactly reflect users’ availability. This limitation does not appear to diminish the utility of IM in helping initiate contact, in part because IM interactions are designed to be so lightweight. It is not too expensive to IM someone even if they are not there, or to dismiss an IM that comes at an inopportune time. Similarly, the design of any applications using rhythmic inferences from activity data needs to accommodate this limitation of the data.

While the activity data logs do not exactly give us the availability status information about the user that we would like, these data can all be automatically collected without burdening the user. We wanted to explore what meaningful patterns could be identified in the activity data that is readily available. Our research explores the constraints of what these data allow us to infer and suggests what additional data might be useful to provide more context.

### Visualizing Computer Activity Data

We created a tool to visualize the computer activity data for our analyses. Figure 1 shows a subset of the activity data for an individual. Figure 1a displays each minute that the person interacted with his computer during the day for three weeks. Time within a day is represented horizontally along a timeline (shown at the top). Activity during a specific minute of a day is shown with a short vertical tick. Durations of continuous (minute-to-minute) computer activity show up as horizontal bands of black, whereas durations of inactivity emerge as spans of the grey background. The activity plots for a number of days are stacked and aligned vertically in an arrangement known as an *actogram*, a visualization technique used in circadian rhythm research [8].

The actogram affords visually detecting patterns that emerge according to time of day, such as typical onset of activity at the beginning of the day, the lunch break, and when activity stops for the day. Many periods of inactivity can be accounted for by scheduled appointments. Figure 1b continues the actogram for the same person over another three weeks and adds online calendar information. Appointment periods are overlaid as translucent white bands such that computer activity that occurs during the appointment shows through as dark grey. For example, the plot for 10/29 shows two appointments occurring at 10:30-11:00 am and 2:00-3:00 pm. The user was active during much of the first appointment and at the beginning of the second. Activity during an appointment can occur either because the appointment was skipped or it allowed the user to continue to interact with his computer (e.g., a phone call appointment). In either case, the user is reachable despite



**Figure 1.** Visualization of one user's activity showing (a) the actogram of three weeks of computer activity (active periods shown in black), (b) another three weeks of activity with online appointments overlaid in white (computer activity during an appointment appears as dark grey), and (c) aggregate of activity and appointments over the entire ten month data set. Overall patterns such as typical arrival, lunch and departure can be seen in the actogram alone (sections a and b) whereas the aggregate view exhibits other patterns, such as the stepped arrival and departure pattern.

the existence of an appointment in his calendar. Depending on the situation (e.g., appointment details) it may be appropriate to still honor the reserved time, but the activity data provides more context than the user's calendar.

Some patterns are easily seen in the actogram, but more emerge in the aggregate histogram, as seen in Figure 1c. Using the same horizontal timeline, the height of the graph is the sum of the days over the entire ten month data set that the user was active at each minute. As in the actogram, overall patterns of typical start and end of day and lunch are apparent, but the aggregate exposes additional features. Notice that the start and end of the day in Figure 1c have a stepped pattern with an interval of approximately 30 minutes between steps occurring at around 15 and 45 minutes past the hour. This pattern has useful implications if you are trying to reach this person because, for example, if he is not active at 8:20, you can infer that he is not likely to become active before 8:45. Similarly, at the end of the day, if he is active at 5:20, he is likely to continue to be reachable until at least 5:45. This pattern can be verified by examining individual days in the actogram, and we later learned that it was the result of using a public bus to get to and from work.

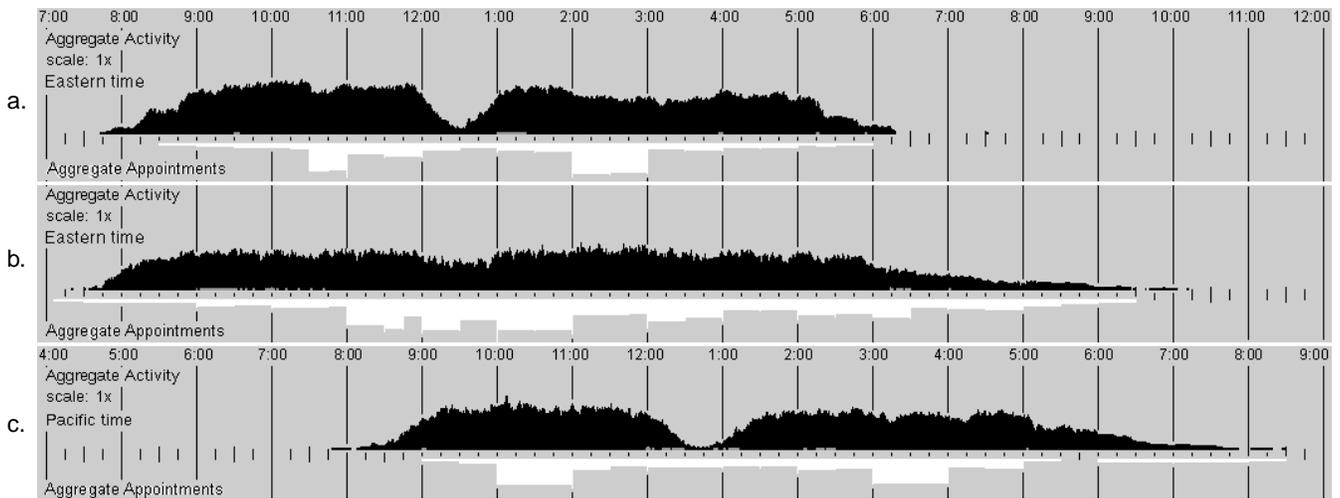
The aggregate in Figure 1c also shows a dip in activity starting at approximately 10:30 am. As in the actogram, we first looked at scheduled appointments to explain such periods of decreased computer activity in the aggregate. Appointments are also aggregated and plotted along the negative vertical axis below the activity aggregate in Figure 1c. Notice that a stack of appointments in the appointment

aggregate between 10:30 and 11:00 corresponds to the activity aggregate dip. Thus, some dips in aggregate computer activity are accounted for by recurring appointments.

Notice that the drop in activity at 10:30 is more pronounced than the gradual return to the basal activity level by 11:00. Looking at instances of the recurring appointment in the actogram of Figure 1b, we see that while the activity generally ceases at the beginning of the appointment period, activity tends to resume earlier than its scheduled end time. This illustrates how activity data can augment calendar information to more accurately indicate when someone can be reached. The appointment would indicate this person is not reachable between 10:30 and 11:00 but the activity record shows that he may be reachable earlier than 11:00.

### Comparing Aggregate Patterns Among Individuals

Comparing the visualizations of aggregate computer activity and scheduled appointments among a group of users presents several patterns that can be useful in group coordination, as shown in Figure 2. The aggregate shows if a person has very regular times for starting and ending the workday and breaking for lunch. The user in Figure 2a shows well-defined regularity for these features, while the user in Figure 2b has a much wider variance of starting and stopping work and taking lunch (if lunch is taken at all). The user in Figure 2c shows the effect of being time-shifted by three hours. This user shows fairly regular daily patterns in starting activity and taking lunch, but a fairly wide variability in ending activity. Many people show a more regular pattern for starting the workday than for ending it.



**Figure 2.** Comparison of aggregate over all days of computer activity and scheduled appointments between two East Coast (a,b) users and a West Coast (c) user, showing the actual overlap of activity between individuals in different time zones.

When comparing users across time zones, the visualization provides a more accurate sense of the periods when work activity overlaps. For example, the users in Figure 2 work in time zones that differ by three hours. While one might naively expect the group to have a five hour work overlap, the actual overlap depends on the individual work rhythms. Users 2a and 2c overlap for only three to four hours because both end their days regularly and take lunch at the typical times in their respective time zones. In contrast, the overlap between users 2b and 2c is about five hours or longer because user 2b does not take lunch regularly and often works late at the end of the day.

Adding aggregate appointments to the comparison affords finding optimal times when a group can meet. Valleys in the activity aggregate and peaks in the appointment aggregate indicate times when someone would be less likely to be available for an appointment. For example, the people in Figure 2 potentially have overlapping availability between 4:30-5:15 Eastern (1:30-2:15 Pacific).

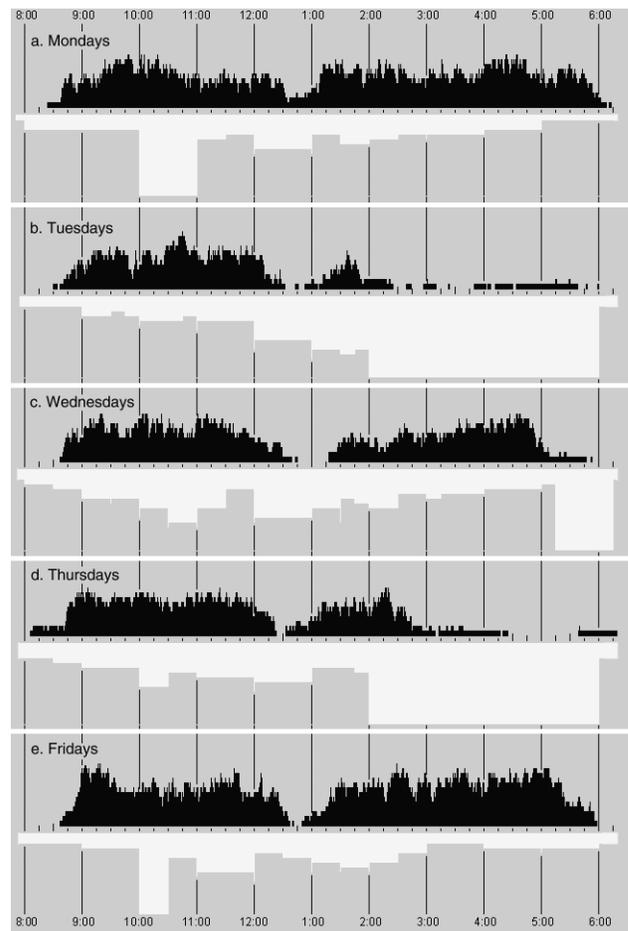
### Day of Week Patterns

Figures 1 and 2 have shown activity patterns across all days but people's work rhythms may differ according to the day of the week (i.e., Sunday, Monday). Figure 3 shows the aggregate activity for a person for each weekday. The variation from day to day suggests that recurring appointments or other regular constraints (e.g., child-care pickup, work from home) can influence when rhythmic features occur. Figure 3a shows that on Mondays he typically arrives around 8:40, takes lunch from 12:30 to 1:10, and leaves around 5:50. These times differ on other days of the week.

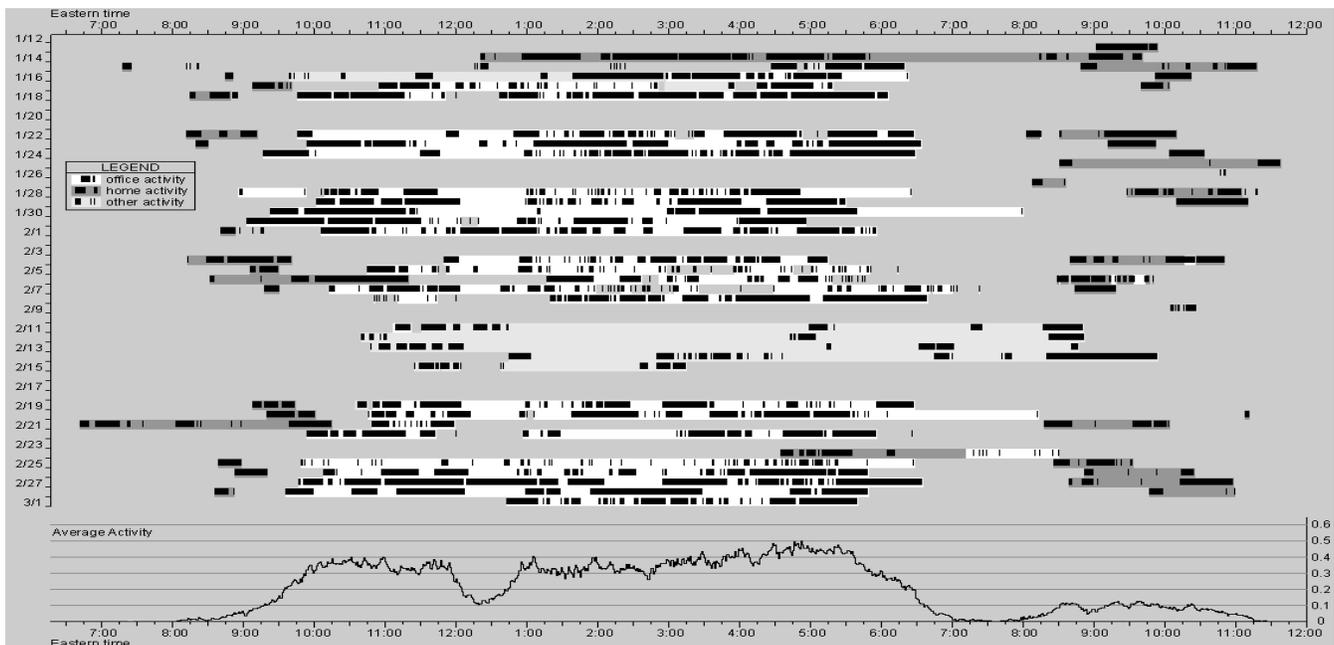
In particular, he has very little activity on Tuesday and Thursday afternoons, which is accounted for by a recurring appointment that starts at 2:00. Notice that on Tuesdays his activity actually stops around 1:45 whereas on Thursdays his activity continues until approximately 2:45. Here again is an example of how the historical activity patterns more accurately indicate a person's reachability than his online calendar information alone.

Closer examination reveals other interesting interactions between computer activity and scheduled appointments.

Looking at Figure 3c, which is the user's graph for Wednesdays, the data show that he has a recurring appointment at 5:15. The computer activity aggregate shows that computer activity drops off sharply at about 4:50. This twenty minute window before the scheduled start of the appointment reflects the preparation and travel time involved in getting to the appointment by 5:15.



**Figure 3.** Aggregate data for each weekday for a user, demonstrating different day-of-week rhythm patterns.



**Figure 4.** Actogram showing location information. The background behind and between activity is shaded according to locale. Patterns differ according to the locale in which activity occurs. Bottom graph shows mean activity over entire data set.

This pattern of an unscheduled buffer time, which can occur before and after an appointment, is most noticeable for recurring appointments in the aggregate. It can also be seen in individual appointments that involve travel or other preparation. If the appointment information indicates that travel is involved, the data can suggest what amount of buffer time should be expected around that appointment.

Some recurring appointments effectively become deadlines (e.g., a recurring appointment that ends the day). Recurring deadlines sometimes correlate with increased computer interaction in trying to accomplish tasks before this deadline. This pattern is seen in Figure 3c, where activity increases in anticipation of the recurring 5:15 appointment. This pattern may indicate a time region when the user is less receptive to interruptions.

#### Location-dependent Patterns

In addition to rhythms that vary depending on day of week, rhythms may differ depending on where the activity occurs. Location information is available based on the locale (e.g., office, home, lab) registered for the Awarenex client being used. Figure 4 shows the activity along with locale information for an individual over a seven week period. Activity is grouped according to locale by shading the background of spans of activity that occur in the same locale.

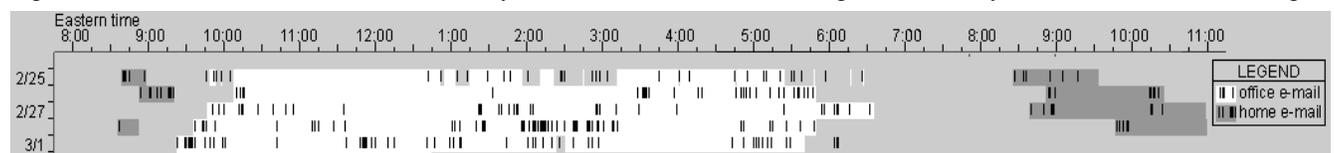
Figure 4 shows activity in four distinct locales: office (white), home (dark grey), and two separate satellite offices (light grey). Note that the patterns of daily onset of activity differ among the locales. For this person, activity often begins at home between 8:15 and 9:00. Usually, after some

period of activity at home, activity ceases for a minimum of 35 minutes and resumes in her office. On those days where her day begins in the office (e.g., 1/28–1/31), the onset of activity occurs between 9:00 and 9:30. These differences in the onset of activity depending on locale may be used to determine appropriate times to reach her. When she logs in from home, she is often reachable earlier than when she comes into the office. When she has ended her activity at home in the morning, she may be expected to be reachable in the office no earlier than 35 minutes later.

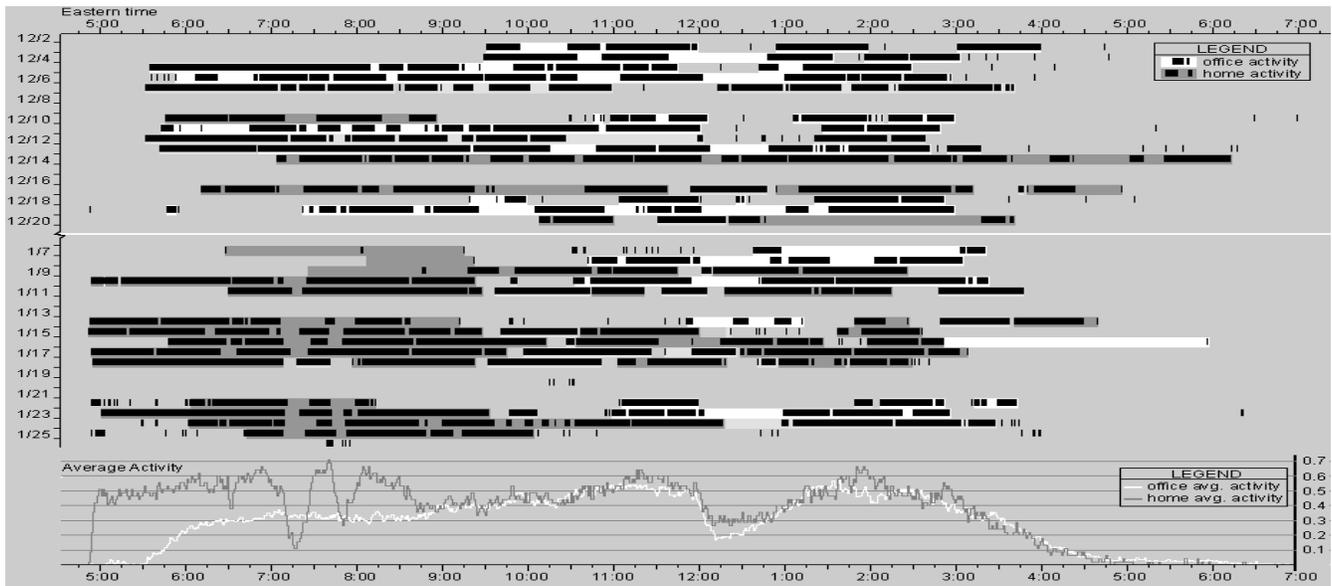
Figure 4 also shows that this person worked in a different office during a visit to the West Coast (in a time zone three hours later) during the week of 2/11–2/15. The onset of her activity during that week is later than usual and there are wider gaps between activity periods when she works in this locale. Again, colleagues who want to reach her may find it useful to know that her reachability differs in this locale.

The bottom of Figure 4 shows the average activity over time, i.e., the mean number of times the person was active for each minute. This alternative to the visualization of aggregate activity over time seen in previous figures gives the approximate probability that the user will be active for each minute based on the history of past activity. Note that this person's probability of activity does not exceed 50%, giving some indication of the amount of her work activity that does not involve interacting with a computer.

We also recorded e-mail activity and Figure 5 shows that when the same user is active at home (both in the evenings and mornings), she always retrieves e-mail messages.



**Figure 5.** E-mail activity (shown in black) and location information for one week of same user shown in Figure 4.



**Figure 6.** Activity and location information for an individual. This person switched from office to work-from-home after 1/7. His activity rhythm has changed significantly (especially in the mornings) as can be seen by the difference in the mean activity graph at the bottom (white is mean activity up to 1/7 and dark grey is mean activity after 1/7).

Remote colleagues in a later time zone may find this information useful because they can expect that she will likely receive e-mail sent after she has left the office, and possibly have a chance to respond before the remote colleagues end their work day.

Figure 6 shows another example of activity patterns that differ according to location. This person began working from home full time beginning on 1/7. Notice that his typical onset of activity at home (around 4:45) begins approximately 45 minutes earlier than his previous typical onset in the office (around 5:30). These differences are reflected both in the actogram and in the mean activity graphs at the bottom. The mean activity before 1/7 when his activity was primarily in the office is shown in white and the mean activity after 1/7 is in dark grey. If his colleagues know when he plans to work at home, they can infer that he is likely to be reachable earlier than when he plans to work in the office.

Additionally, when working at home, there are two regular breaks in activity around 7:15 and 7:50 which are again useful to know for colleagues wishing to contact this person around that time. While the mean activity during the mornings is noticeably different before and after 1/7, the mean activity in the afternoons is rather similar. Since this person now works at home, awareness of these rhythmic patterns can help restore some of the missing cues his colleagues can use to coordinate contacting him. Note that the probability of computer activity for this person (who is a programmer) is higher at home than in the office, especially in the mornings, and is generally higher than that of the person shown in Figure 4 (who is a manager).

#### Activity Patterns Within Individual Days

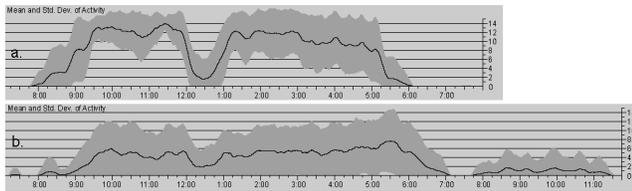
The patterns discussed so far are derived from the activity of many days. The data within individual days also display some meaningful patterns that are not found in the aggregate. Looking at the actogram of individual days in Figure 4, the data exhibit a pattern where she tends to have more extended durations of computer interaction at the beginning

and end of the day, whereas the midday shows more interrupted patterns of activity. This person is a research manager and her pattern of sustained versus interrupted activity shows a suggestive correlation with the results of Hudson, *et al.* [6] who found that research managers reported being more available for interruptions during the middle of the day. Another user's data (not shown) exhibit more extended computer interactions in the afternoon compared to his morning. These patterns have no regular rhythm and do not show up in the aggregate. Yet, these regions of uninterrupted activity may indicate portions of days when users are less receptive to interruption

#### Variability of Patterns Within and Between Individuals

The aggregate activity levels we have described so far represent the total aggregate (Figures 1c, 2 and 3) and mean (Figures 4 and 6) activity over time but these do not indicate the variability of the activity levels within or between individuals. To calculate a meaningful measure of variance, we need to convert the activity data from binary (active versus inactive for a given minute) to non-binary. We performed this conversion by specifying a window of a particular number of minutes, counting the number of minutes within that window that the person was active for each day, summing the counts for the same time window over all days, and calculating the mean and standard deviation for that window over all days. Sliding the window minute-by-minute through the day generated mean and standard deviation values for the entire timeline.

Figure 7 shows the mean and standard deviation for two individuals using an arbitrary window size of 15 minutes calculated for a particular day of the week (Tuesday). Looking at the data for the person shown in Figure 7a, we see that he has generally narrower standard deviation in the morning than in the afternoon. Several people show a pattern where their activity in the morning has a narrower standard deviation band than in the afternoon (suggesting that their mornings are more predictable than their afternoons). Comparing between two individuals, the person shown in



**Figure 7.** Mean (black) and standard deviation range (dark grey) for two individuals showing variance between individuals and within an individual throughout the day.

Figure 7a has a generally higher level of computer interaction than the person in 7b and an overall narrower standard deviation. This indicates that the certainty of predictions about activity and reachability also varies among people.

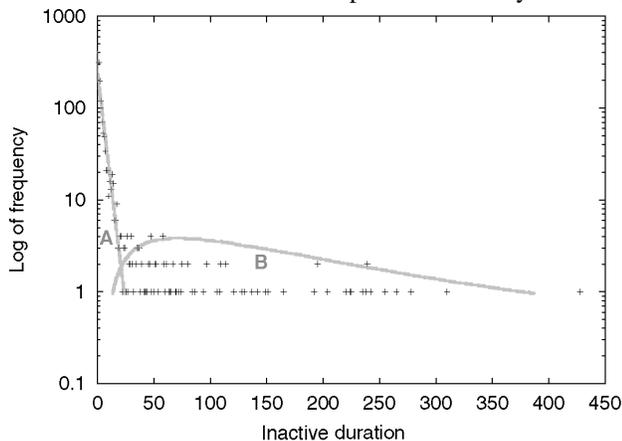
The different activity levels among individuals is a reflection of different styles and patterns of computer interaction influenced by many factors including experience level (e.g., novice versus expert), technology affinity (casual versus “power” user), and role in organization (e.g., manager, programmer, administrative assistant, student).

### Predicting the End of Inactive Periods

By complement, the computer activity logs also provide the periods during which users are not active at their computer. We analyzed these periods of inactivity to determine if we could predict when users might resume activity and therefore become reachable. When inactivity is caused by being away from the computer, predicting when the person will return can suggest good times to try to make contact in the future. For this analysis, we excluded the inactivity between the end of one day and the start of the next.

Figure 8 shows a representative frequency distribution of the duration of inactivity periods for a user. The distribution shows that there are many inactive periods of short durations and few periods of any specific long duration. For example, Figure 8 shows that there are 312 inactive periods of one minute duration but the frequency decreases rapidly until around ten minutes duration, after which the number of occurrences of an inactive period of any specific duration remains relatively small.

This shape of the distribution naturally reflects that there are more causes for short interruptions of activity than long



**Figure 8.** Histogram of durations of inactivity with frequencies plotted according to a logarithmic scale. The data suggest two components: short, random pauses (A) and longer breaks away from the computer (B).

ones. We theorize that the distribution may be the combination of two processes that contribute to the overall shape of the curve, which are shown in overlay in Figure 8. Distribution A contains the many short inactive durations which result from normal pauses during a task, performing tasks not related to the computer, short distractions, etc. The user is typically effectively reachable since they are still near their computer and in any case shortly will be reachable since these durations are brief. We observed that the data that we are characterizing by Distribution A lie along a straight line on the logarithm plot, which is consistent with events generated through a random, Poisson process.

Another component of the curve characterizes lunch breaks, scheduled meetings, visits to a colleague’s office, and other activities that take the user away from her computer, shown as Distribution B. For this data, many of the breaks have a duration of around an hour, but the distribution of this component is right skewed resulting from the variability of longer durations before returning from such interruptions. Specific work settings may have other recurring temporal features that define other components that contribute to the overall curve.

### DESIGN IMPLICATIONS

Studying these patterns and rhythms suggests a number of design implications for applications of this research. As mentioned earlier, our primary interest is in helping workers who are distributed across distance and time zones to coordinate among each other. We are intrigued by the opportunities to share the cues contained in the rhythmic patterns with distributed workers so that those people could interpret this information to help them negotiate good times to make contact. Another application of this work would be to program software agents that could make inferences based on the data and present them to users.

### Group Coordination Applications

Clearly, for those with well-defined rhythms for starting and ending the workday and eating lunch, those patterns can suggest promising times to attempt to contact them. While not every person (nor every work role) exhibits well-defined rhythms, knowing whether they exist for an individual is a useful bit of information toward helping coordinate contact. For those who do exhibit rhythms, there are more opportunities for applying the analysis of their activity data to create a shared sense of time among them.

### Suggesting Good Times to Make Contact

Beyond the overall rhythm for arriving at work, leaving and returning from lunch, and leaving at the end of the workday, the analysis can provide more fine-grained details. For example, we described how the pattern in Figure 1c of a user’s stepped arrival and departure pattern has predictive power about his availability at the beginning and end of the day. Our experience with the visualization tool for analyzing the data suggests the possibility of designing visualizations of activity that could be presented to end users so they could infer good times to make contact based on rhythmic patterns. Especially when initiating contacting someone with whom you are unfamiliar, it would be helpful to have a sense of their patterns of when they can be reached.

Also, for those who have a regular lunch pattern, the histogram of inactive durations around lunch exhibits a normal

distribution from which a mean and standard deviation can be calculated. If you were to check a person's status during lunchtime, this statistical description can help predict when she will return from lunch. For example, if her mean lunch duration is 52 minutes, and she has been inactive for 27 minutes when you check, you might expect her to return in about 25 minutes.

#### *Predicting Return from Inactivity*

Predicting the return from lunch is a special case of the analysis of inactive durations. In general, if you find someone to be inactive when you want to contact them (with no calendar appointment to account for it), it would be useful to know when they might return. The inactive duration analysis can help predict a range within which a user will likely become active again. For example, if someone has been inactive for six minutes, the system might report a 90% chance that the user will become active within another 27 minutes, based on the history of inactivity duration data.

A related question that the data could help answer is whether it would be worth waiting a certain period of time for the person to return. For example, the system could answer the question "what is the probability that the person will be back within the next 5 minutes?" This is useful when a caller would be willing to wait a short period if the other person is likely to return within that period.

#### *Augmenting Online Calendar Accuracy*

Considering the computer activity patterns in combination with scheduled appointments is particularly useful. There are several reasons why the computer activity logs show some patterns that would not be found in a person's calendar. Tullio et al. [14] observed that people's calendars are only a plan for their use of time, and may substantially differ from what actually occurs. In some ways, the logs more closely represent their actual activity as evidenced by their use of a computer. For example, computer activity recorded during a scheduled appointment may indicate that the person did not actually attend that appointment, or that the appointment started later or ended earlier than scheduled.

Furthermore, rhythms in the computer activity data indicate some very regular patterns that may not be entered into the calendar. Breaking for lunch is one example of an event that happens regularly in many people's workday, but is not usually scheduled in their calendar (unless it involves other people). Similarly, starting and ending the workday are typically not scheduled in a calendar, but can be clearly seen in the computer activity data for many. Beyond daily arrival, departure and lunch patterns, some people have other regular periods of inactivity at certain times during the day (e.g., the two breaks between 7:00 and 8:00 in Figure 6). Other regular breaks may not occur at a specific time, but occur between transitions of activity in different locales (e.g., the gaps between home and office activity in Figure 4).

Also, some of the patterns seen in the computer activity data augment what is shown in the calendar. The buffers of no activity before and after appointments (both recurring appointments and individual ones) indicate travel, preparation/recovery, and other time surrounding a scheduled appointment. Again, these would not be accounted for by the appointment itself, but are observable in the computer activity data in relation to the appointments.

In these ways, the computer activity data can describe people's behavior in ways beyond what a person's calendar can. Taking the computer activity data and schedule data together can be a more powerful predictor of when people could be reached for communication.

#### *Identifying E-mail Reading Patterns*

We also found that the record of an individual's e-mail reading pattern is a useful indication of when a sender can expect a message to be read. This information may be useful in determining whether sending e-mail would be an effective way to contact someone, based on the urgency of the need to make contact and the reading patterns of the recipient. Further analysis could identify typical patterns for how quickly a user responds to e-mail.

#### *Restoring Cues to Negotiate Initiating Contact*

The direction of applications that we intend to pursue builds on the notion of social translucence (Erickson & Kellogg [3]) of providing awareness cues that allow people to socially negotiate appropriate action. We would like to prototype applications that transmit activity rhythm cues among electronic correspondents (especially those who are physically remote from each other) to allow them to more effectively and naturally negotiate coordinating contact among them. Most communication technologies place most of the burden of managing communications on the receiver (e.g., e-mail filtering, phone screening, immediately responding to IMs). By contrast, physically co-located colleagues have rich cues available to them (e.g., closing the office door, averting eye contact, awareness of office activity patterns) to help mutually negotiate when and whether to start an interaction. Transmitting activity rhythm cues to remote colleagues can inform their decisions about when and how to initiate contact. These cues could help restore a sense of negotiation between sender and receiver in ways that allow them to more gracefully establish contact at convenient times for both parties.

#### **Other Applications**

Beyond helping groups coordinate, analyzing the logs suggested some other applications of time-related patterns. For example, reviewing your own work activity history can be helpful in setting preferences for managing when and how people can contact you. An analysis of historical logs of when and where you are active as well as calendar appointments could provide useful defaults for user settings of when to route calls to your home office rather than your corporate office, for example. It may also help you identify regularly occurring time patterns of when you prefer to have contact and when you would rather not be disturbed. Using these rhythmic patterns to help configure your own settings is an application of this work that avoids any privacy exposure of that information to others.

A different twist on using this data is to monitor the use of shared computer and networking equipment. With renewed interest in centralized computing resources (e.g., Application Service Providers, centralized compute servers), having an accurate and personalized model of how a collection of users actually use their computers would be helpful. This information would help in configuring how many users can share a computing resource as well as identifying optimization strategies for improving performance. For example, a

better understanding of users' inactive behavior could identify a threshold inactive duration (tailored for that person) after which the system could swap out the user's computing resources into a cache to free up those resources for others.

### PRIVACY IMPLICATIONS

Collecting, analyzing, and presenting computer activity data raise obvious concerns about revealing sensitive information about people's activities. Even though co-located workers currently have an awareness of much of this activity information, Grudin [5] observes that collecting and distributing such data electronically heightens concerns of being susceptible to undesirable uses. For example, users have expressed concern that managers could use this data to monitor employees' computer activity and make inferences about their productivity. While addressing all of the privacy issues is not within the scope of this paper, we hope that our research helps inform the larger discussion about appropriate uses and safeguards needed to reap the benefit in applications without creating undue privacy exposure.

Our *research* has required greater exposure of people's computer activity logs to scrutiny than any of the *applications* would need. This research has involved centralized collection and required detailed analyses of peoples' activity patterns to understand how those patterns correlate with actual activity and reachability. We have considered how applications and systems resulting from this research can be designed to ameliorate some of the privacy concerns.

First, while the data and visualizations presented here contain fine-grained details, applications could present inferences drawn from the data or considerably abstracted views which would be much less revealing than the data itself. For example, a simple application of this data would be to query the system for good times to reach someone. The system would respond with suggested times to make contact without exposing all of the information used in the inference. If users decide they are comfortable exposing more information, the system might provide an abstracted visualization of their rhythm patterns. For example, an image that only shows an overview of a user's workday (when he starts, takes lunch, and ends) with fuzzy boundaries that indicate uncertainty about exact times would be useful without revealing sensitive details.

Another approach is to regulate who gets entrusted with what kind of information. As is currently the case with access to online calendars, some people (e.g., assistants, close associates) might be allowed to see more details than others (e.g., strangers, your manager). Applications may need to allow for several different levels of disclosure and an interface for setting who has access to what information. Those applications that only reflect the rhythmic patterns back to you (e.g., as defaults in user settings) avoid the privacy concerns of sharing this information with others.

Thus, applications can be designed that are sensitive to users' privacy and protective measures can help ensure that detailed data are not accessible by unauthorized users. Yet, the existence of the data and the determination of who is "authorized" to access it still raises concerns. For example, privacy policies can be structured such that upper level managers or human resources personnel are considered "authorized" to access all data. Ultimately, all such data are

subject to the authority of a court subpoena, after which the system's intended access controls will no longer apply as the data become subject to the rules of legal proceedings.

In the end, as with many applications (e.g., e-mail repositories, shared calendars, IM, web cookies), users must assess whether the benefits gained outweigh the privacy risks. Along the lines of viewing privacy as a property right [12], we think of the cost/benefit analysis in terms of a "privacy economy"—people make choices about how much privacy disclosure to "pay" based on the benefits they perceive they are receiving. Our work explores what benefits might be gained and exposes some of the associated privacy costs. The design and development of applications based on this research needs to be sensitive to the potential privacy risks, and the users should also be able to opt out of these applications or negotiate an acceptable place on the spectrum of privacy protection, based on their sense of personal benefit and privacy risks.

### FUTURE ANALYSIS AND APPLICATION WORK

Our research has opened up more opportunities than we have been able to explore to date. The computer activity logs are ripe for more data mining and other analysis. We plan to develop programmatic techniques to detect features that we have discovered in our analyses described in this paper. We will integrate these feature-detection techniques into applications to explore their usefulness. These steps would also enable us to validate how well historical patterns predict future activity.

Besides looking for more patterns in the activity data, there are other kinds of analyses to consider. An analysis of the duration of computer activity episodes, akin to what we presented about inactive durations, might also reveal patterns or structure to that activity. More comparisons of patterns among people could help identify group patterns.

One area of interest for further exploration is the difference between being *reachable* for communication and being *available* for it. Availability depends not only on physical presence but also on mental receptivity to communication. Computer activity data clearly indicate when a user is sitting in front of a computer and thus can be reached through a variety of media (e-mail, instant message, nearby phone). Yet, being currently engaged in a task with the computer may also indicate a time when the user is not receptive to being interrupted.

Conducting a study similar to the Hudson *et al.* [6] work, which probed users' availability preference at random times throughout the day, among the people that we have activity data for could help identify correlations between being reachable and available. Still, there is no way to account for people's mental receptivity to interruption (even with those who are co-located). Applying what we have learned about people's activity rhythms provides useful cues about being reachable, but we expect that people will always have to rely on a social context to make decisions about when and how to initiate contact.

We are quite interested in developing applications that make use of this research. We expect to develop prototypes that help support the coordination of distributed teams, perhaps by integrating more features into the Awarenex [13]

prototype. We plan to observe these prototypes in use to examine how effective they are and understand how people react to the privacy issues that arise.

#### TOWARD A SHARED SENSE OF TIME

This study analyzed computer activity information which included data on the time and location of keyboard/mouse activity, and e-mail reading/sending. We explored the data using a number of visualizations. We noted a number of temporal, rhythmic patterns in the activity record. Different patterns emerge at different levels of analysis: overall patterns appear in the day-by-day activity graphs (actogram), other features emerge in the aggregate across all days. Some rhythm patterns depend on the day of the week and the location in which the activity occurred. Studying the rhythm visualizations provided concrete representations for some intuitive notions of being time-shifted, but also yielded intriguing and rich details about our own group's activity, some of which were surprising to us even though we had been working together for over two years.

The observations reported here contribute to the general understanding of peoples' rhythmic interaction with computing devices. Furthermore, the patterns suggest a number of applications to help remote colleagues coordinate communication. For example, the distribution of inactive periods may be used to suggest when someone will likely return from being inactive. In addition to work rhythm patterns, the data give a history of activity that is often more accurate than online calendar schedules. With respect to privacy, we feel this work contributes to the ongoing public debate by exploring what information can be inferred from the online presence data that current technologies are making increasingly available. A proactive understanding of the uses of such information is necessary to consider the tradeoffs of privacy and awareness. By studying work rhythms among distributed teams, we hope to find ways to create a shared sense of time among remote colleagues.

#### ACKNOWLEDGEMENTS

We want to thank Helen Cunningham, who suggested many ways of exploring and analyzing the data that helped shape this research. We thank our intern, Mark Bilezikjian, who helped process the log data and reviewed earlier versions of this paper, and Frank Richichi, who helped collect the e-mail data. We also thank Brian Raymor for helping us confront the privacy issues raised by this work. We especially want to thank the Awarenex users that entrusted us with access to their activity logs for analysis for this work.

#### REFERENCES

1. Agre, Philip and Marc Rotenberg (Eds.), *Technology and Privacy: The New Landscape*, Cambridge, MA: MIT Press, 1997.
2. Dourish, Paul and Sara Bly, "Portholes: Supporting Awareness in a Distributed Work Group", *Proceedings of the Conference on Computer Human Interaction (CHI) '92*, Monterey, CA, May 1992, pp. 541-547.
3. Erickson, Tom and Wendy Kellogg, "Social Translucence: An Approach to Designing Systems that Mesh with Social Processes." *Transactions on Computer-Human Interaction*, Vol. 7, No. 1, 2000, pp 59-83.
4. Greenberg, Saul and Hideaki Kuzuoka, "Using Digital but Physical Surrogates to Mediate Awareness, Communication and Privacy in Media Spaces", *Personal Technologies*, Vol. 4, No. 1, January 2000, pp. 182-198.
5. Grudin, Jonathan, "Desituating Action: Digital representation of Context", *Human-Computer Interaction*, Vol. 16, Nos. 2-4, 2001, pp. 269-286.
6. Hudson, James, Jim Christensen, Wendy Kellogg, and Thomas Erickson, "'I'd Be Overwhelmed, But It's Just One More Thing to Do: Availability and Interruption in Research Management", *Proceedings of the Conference on Computer Human Interaction (CHI) 2002*, Minneapolis, MN, April 2002, pp. 97-104.
7. Mantei, Marilyn, Ronald Baecker, Abigail Sellen, William Buxton, and Thomas Milligan, "Experiences in the use of a media space", *Proceedings of the Conference on Computer Human Interaction (CHI) 1991*, New Orleans, LA, 1991, pp. 203-208
8. Moore-Ede, Martin, Frank Sulzman, and Charles Fuller. *The Clocks That Time Us: Physiology of the Circadian Timing System*, Cambridge, MA: Harvard Univ. Press, 1982.
9. Mynatt, Elizabeth, Jim Rowan, Sarah Craighill, and Annie Jacobs, "Digital family portraits: Supporting peace of mind for extended family members", *Proceedings of the Conference on Computer Human Interaction (CHI) 2001*, Seattle, April 2001, pp. 333-340.
10. Nardi, Bonnie, Steve Whittaker, and Erin Bradner, "Interaction and Outeraction: Instant Messaging in Action", *Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW) 2000*, Philadelphia, December 2000, pp. 70-88.
11. Palen, Leysia, "Social, individual and technological issues for groupware calendar systems", *Proceedings of the Conference on Computer Human Interaction (CHI), 1999*, Pittsburgh, May 1999, pp. 17-24.
12. Post, David G. "Privacy, Property, Cyberspace", *American Lawyer*, November 1997.
13. Tang, John C., Nicole Yankelovich, James Begole, Max Van Kleek, Francis Li and Janak Bhalodia, "ConNexus to Awarenex: Extending awareness to mobile users", *Proceedings of the Conference on Computer Human Interaction (CHI) 2001*, Seattle, April 2001, pp. 221-228.
14. Tullio, Joe, Jeremy Goecks, Elizabeth Mynatt, and David Nguyen, "Augmenting Shared Personal Calendars", *Proceedings of the Symposium on User Interface Software and Technology (UIST) 2002*, Paris, in press.
15. Zerubavel, Eviatar, *Hidden Rhythms: Schedules and Calendars in Social Life*, Chicago: The University of Chicago Press, 1981.