Show Me a Good Time: Using Content to Provide Activity Awareness to Collaborators with ActivitySpotter

Brian Y. Lim
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
byl@cs.cmu.edu

Oliver Brdiczka
Palo Alto Research Center (PARC)
3333 Coyote Hill Road
Palo Alto, CA, USA
brdiczka@parc.com

Victoria Bellotti
Palo Alto Research Center (PARC)
3333 Coyote Hill Road
Palo Alto, CA, USA
victoria.bellotti@parc.com

ABSTRACT
In order to study the effect supporting awareness of a colleague's activity on a collaborator's communication intentions, we developed ActivitySpotter. It is a research tool and awareness display that determines a user's current activity through a semantic analysis of documents s/he accesses and shares this information with collaborators. We ran a user study on 22 participants to investigate how accurately ActivitySpotter represents user activity and whether different representations of activity (presence only, topic keywords, or activity labels) influence awareness differently and lead users to change their contact intention. Our findings suggest that activity content awareness can help users glean more about what their collaborators are doing, especially if they are more socially distant, and can afford screen space to have the display showing. This increase in awareness also positively influences users' intentions to communicate in a socially appropriate manner.

Categories and Subject Descriptors
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors.

Keywords
Awareness, display, collaboration, interruptibility.

1. INTRODUCTION
Knowledge workers have to juggle many tasks at the same time [17] and yet often have to attend to interruptions from colleagues and collaborators. While interruptions can be disruptive, leading to long recovery periods [7] and increased stress [26], they can assist in information sharing and collaboration [28], [36]. Nevertheless, in spite of this benefit, people tend to reject attempted interruptions [36]. One reason for this is the asymmetry of communication goals between the contactor (initiator) and recipient [36]. While a contactor can reduce her cognitive load by "discharging her communication responsibilities," the recipient increases his cognitive load because he has to take on the contactor's communication goals and still recall his original task when he returns to it. Previous research [24] indicates that people would prefer to be interrupted at more appropriate times and about topics or activities that are more relevant to their current activity. We would like to reduce this cognitive load so as to facilitate casual interruptions. One way to do this would be to influence the timing of interruptions to when both the contactor and recipient are doing the same activities (i.e., their topics have *concurrency*). While [26] showed that content-inappropriate interruptions do not affect disruption time, they still cause more stress to the recipient, and we would like to minimize that.

We aim to facilitate this concurrency by showing the potential contactor what her collaborator is doing. To support this requirement and mitigate potential problems with privacy and cognitive load of interpreting low-level activity content information [9] (e.g. document titles, content), we abstract content information through a *semantic analysis*. Abstraction can hide specific levels of detail that users may be concerned about sharing, and the information is more concise and easier to interpret. We then rely on the contactor to exercise sensitivity and optimize the times when he initiates contact given this new awareness. To achieve this, we developed a research tool and activity awareness display, ActivitySpotter (AS), which infers a user's activity via semantic analysis of documents that the user accesses and presents activity content information as low-level detailed topic information, or high-level abstracted labels.

Our contribution is to (i) explore a means to generate high-level activity content information to share through an awareness display, and (ii) provide an understanding of how users appreciate this activity content information of their collaborators and use the information to moderate their contact intention. Our results indicate that AS is reasonably effective at raising users' awareness of their collaborators. This increase in awareness also positively influences their intentions to communicate in a socially appropriate manner.

In the next sections, we discuss activity awareness displays and how AS employs a different type of activity information than has been presented to users previously, we then describe our implementation of AS, the user study that we ran with it and our findings and some recommendations on what information to share so that users can be more socially appropriate contactors.

2. ACTIVITY AWARENESS
Much research has looked into providing people with increased awareness of the world and people around them. We are interested in awareness of others in collaboration (collaboration awareness)

Group'10, November 7–12, 2010, Sanibel, Florida, USA.

rather than information awareness where much work has been done with peripheral displays (e.g., [24], [32]). Gutwin [18] divides collaboration awareness into informal, conversational, structural, and workspace awareness. Informal awareness is knowing "who is around, what they are up to, and whether they are available." Conversational awareness covers verbal and non-verbal cues in face-to-face communication. Structural awareness is understanding of the organizational and social structure between people. Finally, workspace awareness covers the environment and artifacts involved in collaborations. Our study focuses on the effects of increasing *informal* awareness, particularly about what people are up to, to facilitate communication. Next, we discuss previous awareness displays built to support coworkers.

2.1 Awareness Displays

Early work on awareness displays focused on showing presence and activity in a peripheral display. Portholes [[10]] showed frame-grabbed video stills of collaborators at their desks. While this might be very informative about whether one's collaborator is available, it does not say much about what they are working on. Peepholes [[16]] took a more conservative approach using iconic presence indicators to show people's likelihood of being available determined from activity on their computers. Later systems made use of environmental sensors (e.g. door sensors, microphones) to detect and infer presence availability, and present that information to desktop displays (e.g. [[1], [14]]), and door displays (e.g. [[12], [35]]).

All these awareness displays used context information to represent availability. Previous work (e.g. [3], [25]) showed that people also cared about the relevance (content) of the interrupted topic too. We seek to support this by providing content information in awareness displays. One way to do this is to display titles of documents that people are working on. Tee et. al. [33] added screen-sharing capabilities to the Community Bar [27], allowing users to see screens of their collaborators. However, there needs to be a trade-off. While there are privacy measures in place to help obscure the screens when desired, users may prefer to set the obfuscation on all the time and then forget to share their screens at high resolution. In fact, Brush et. al. [4] identified several types of information that users would rather not share, including many types that are collected by the aforementioned displays (e.g. title documents, URLs of webpages, video feed). Hudson et al. [20] identified the dual trade-off of awareness with privacy, and also awareness and obtrusiveness. Document names and screenshots may be too detailed for users to quickly interpret and assess in a peripheral display.

3. ACTIVITY AWARENESS

To investigate the usefulness of activity content for awareness displays, we developed an experimental instrument, ActivitySpotter (AS). As AS is a research instrument, we focused mainly on its ability to collect useful information rather than develop the best possible user experience. In this section, we describe the building of the topic model and activity model that AS uses, how it classifies higher-level content information from documents, and how it displays the information (see Figure 1).

3.1 Topic Modeling from Documents

Previous work (CAAD [29]) proposed grouping software artifacts (documents, folders, web pages, people) of a user's interaction contexts into clusters to represent his task. While CAAS established associations between artifacts and considered

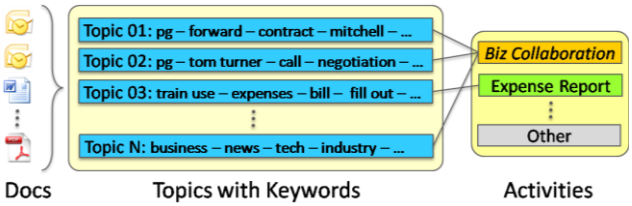

Figure 1. ActivityIndexer builds a topic model through LDA clustering from documents the user has. The user then uses ActivityConfigurator to manually map topics to relevant activities. ActivitySpotter uses the topic model and activity map to determine which topics and activities are related to the document the user accesses.

documents as atomic entities, it does not look at the content *wihtin* the documents. In this paper, we focus on clustering the user's content, in particular, text of documents that he uses. We use semantic analysis to construct a content-based representation of the user's activity.

Using ActivityIndexer, we extract a list of terms (words) from relevant documents (in our prototype, they are just Microsoft Outlook emails sent and received in the past month). Further preprocessing is done to filter the list:

1. Calculate the TF-IDF (term frequency inverse document frequency) [29] values of every term found.
2. Remove terms that have TF-IDF values <.001 or that are stop words (common words such as "the," "is," "about").
3. Group frequently co-occurring pairs of terms together into bigrams (e.g. "expense report," "statistical analysis") and add these to the list of terms.

We then use the terms as input for the topic model estimation using a clustering technique, Latent Dirichlet Allocation (LDA) [3]. LDA has been used for estimating the topics from the term list and occurrence frequencies. In particular, [10] used it to generate email summary keywords. This motivates our using LDA for modeling email content about activities. LDA is better than the simpler TF-IDF method of generating keywords because it models the user's work in terms of higher level topics. This is important to associate documents with the same topics and performs better than simple heuristics derived from keyword matching over TF-IDF. It is a generative model that allows sets of observations to be explained by unobserved groups or clusters that explain why some parts of the observation data are similar. In the context of documents, words can be considered as observations collected in documents. Each word can be interpreted as being generated by one or a small number of topics; the creation of each word is attributable to one of the document's topics. In our approach, we choose 50 topics (pre-tested to provide reasonable accuracy without excessive effort for users to subsequently categorize) to learn from the document corpus of each participant. We specifically used the LDA toolkit, which can be downloaded at http://www.cs.princeton.edu/~blei/lda-c/index.html.

3.2 Activity Mapping

Topic modeling allows topics to be extracted from a user's documents, but these topics are unlabeled and may not relate explicitly to activities that the user does. To relate the extracted topics to the user activities, we provide an activity mapping user interface, ActivityConfigurator, for users to map topics to activities (see Figure 2). In this study, we allow users to define projects that they work on and non-project-specific activities (e.g.

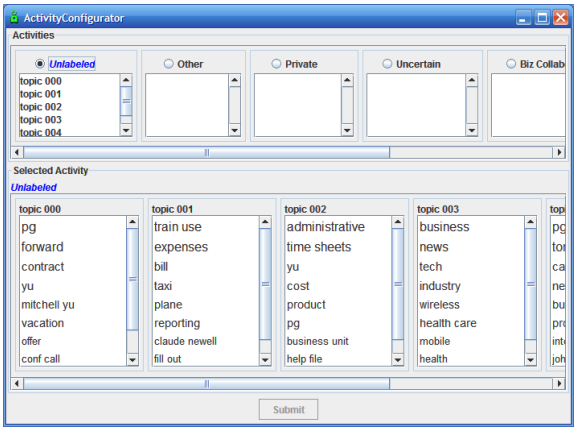

Figure 2. ActivityConfigurator for users to map topics to activities and categories. Users map whole topics (not individual keywords), one activity per topic.

billing, administration) as activities. ActivityConfigurator presents the users with the 50 extracted topics as lists of keywords, ranking the words in order of salience to the topic. More salient keywords are shown at the top of the list in bigger font to help user understand what the topics represent. Users drag whole each topic keyword list (not individual keywords) to an appropriate activity. We provided meta-activity categories: other, private, and uncertain. The user can put topics that do not relate to any pre-defined activity into the "other" category. If she finds that a topic has keywords that make it ambiguous as to which activity it should belong to, she can place the topic in the "uncertain" category. Such topics would not be used in activity determination in the classification stage. The user can also put a topic in the "private" category if she does not want to share any of its keywords.

3.3 Classification of Topics and Activities

With the topic model and activity map done, the user can use ActivitySpotter (AS) to classify his documents according to topics and activities. AS monitors the user's desktop activity and as soon as it detects a window change (*i.e.*, the user switches applications), AS obtains a handle to the current application and extracts the text from the document contained therein. AS can do this for a limited set of applications: Microsoft Outlook and Office applications, Acrobat Reader. For some other applications, it can identify file names from the window title, and extract the text content from the file. AS does not read certain applications (e.g. web browsers) to avoid sharing sensitive information [4]. Using topic inference based on the learned LDA model and the retrieved document text, AS derives a score for each topic. This score is based on the keywords of the topic, their weight within the topic and frequency in the extracted document text. Based on the calculated matching score, we rank the topics, retain the best three matched and calculate their relative confidence values (normalized to 100%).

Activities are classified using the user-defined activity labels, manually associated topics (using activity definition) and the previously selected best three topics (see Figure 1). The three topics vote for their associated activity labels and their scores are added up to generate the resulting activity weight and confidence values (also normalized to 100%).

Figure 3. Three versions of ActivitySpotter: Presence only version showing minimal activity context, Topic Keywords version (expanded) showing presence and topic keywords of the currently accessed document, and Activity Labels version showing presence and activity labels of the currently classified activity content.

3.4 Three Versions of Awareness Displays

To investigate the impact of sharing activity content information, we developed three versions of AS: one version showing only presence, and two content-aware versions (Topic Keywords and Activity Labels) showing presence and activity content in a space-saving (expandable/collapsible), desktop-based, peripheral display (see Figure 3). We did not want to interfere with the user's work too much, and enabled AS to be minimized. The *Presence* version, only shows whether a user is online, offline, or away (had not touched his keyboard or mouse for at least 5 minutes). The *Topic Keywords* version shows three relatively long lists of keywords to represent topics related to the user's activity. Each list shows the ten best keywords for each of the three best topics with regard to the current document the user is working on. The *Activity Labels* version shows a higher level, more abstracted interpretation of the user's activity — the mapped activity labels — but at the cost of greater inaccuracy due to possible mismatch during the user's categorization of activities. Topic keywords and activity labels are prepended with confidence percentages indicating the likelihood AS thinks the respective topic or activity is representing the user's or collaborator's activity.

Figure 1 and Figure 3 illustrate an example where Alice is working on a document, say, `bizdev-pg.doc`. The Presence only version of AS would not indicate that she was working on this document. The Topic Keywords version would show the keywords ("pg," "forward," "contract," etc). The Activity Labels version would display the label "Business Collaboration."

3.5 Limitations and Assumptions

We sought to determine the impact of providing activity content information through an awareness display and how the new awareness (if any) would affect contact desire. As such, to avoid confounds, we did not include contextual information that are known to be useful to help users know when to contact their collaborators (e.g. availability [1], [15], interruptibility [12], breakpoints [22], calendar schedule).

Table 1. Hypotheses regarding the perceived accuracy, resulting awareness of collaborator activity, and cognitive load when using different versions of ActivitySpotter.

	Measure	Hypothesis
H1	Accuracy	Topic Keywords < Activity Labels
H2	Cognitive Load	Topic Keywords > Activity Labels
H3a	Awareness	Presence < {Topic Keywords & Activity Labels}
H3b		Topic Keywords < Activity Labels

Table 2. Hypothesis (H4) of the relationship between awareness and contact desire. Without awareness, people would have a higher contact desire for their collaborator when they are working on the shared activity. However, with awareness, they would moderate their activity according to what they think their collaborator is doing.

***Perceived* Activity Concurrency**	**Collaborator Activity Unaware**	**Collaborator Activity Aware**
Neither	High contact desire	**Least** contact desire
Only Self	High contact desire	Less contact desire
Only Collaborator	Low contact desire	More contact desire
Both	Low contact desire	**Most** contact desire

AS supposes that a knowledge worker's documents are a reasonable representation of his work activity. But, while LDA clustering is reasonably accurate in determining topic clusters, our mapping and filtering process requires end-users to categorize the topics into activities and projects. This explicit labeling of clusters resembles the requirement to label current tasks in the TaskTracer system [12], but is done only once when setting up AS (rather than continually throughout the study). We believe that this is fine for an experimental system running for shorter than a month.

4. HYPOTHESES

We hypothesize that presenting users with activity *content* information would lead to similar improvements in awareness as presenting them with activity *context*. We expect AS to better represent the user's activity than random chance. Given the simplification of abstracting the content information, we hypothesize that the user will be more able to interpret the Activity Labels and rate that version as more *accurate* (or representative of the user's activity) than the Topic Keywords awareness type (H1). On the other hand, in line with [9], we expect the user to struggle more with interpreting Topic Keywords due to increased *cognitive load* in understanding the details (H2). We hypothesize that AS will convey to the user some activity information about her collaborator, and increase her awareness of what her collaborator is doing (H3a). The Activity Labels version will be more effective than the Topic Keywords version because users, feeling the former is more accurate and useful, will use it more to find out about their collaborator's activity (H3b). We hypothesize that having learned her collaborator's activity (and the *concurrency* of her own and the other's activities), the user will appropriately moderate her contact desire: reducing it when she sees that her collaborator is working on an unrelated activity, and increasing it otherwise (H4). Table 1 and Table 2 summarize the hypotheses.

5. METHOD

We used activity information awareness type (Awareness Type) as our manipulation variable with three values embodied in three versions of AS: Presence, Topic Keywords, and Activity Labels.

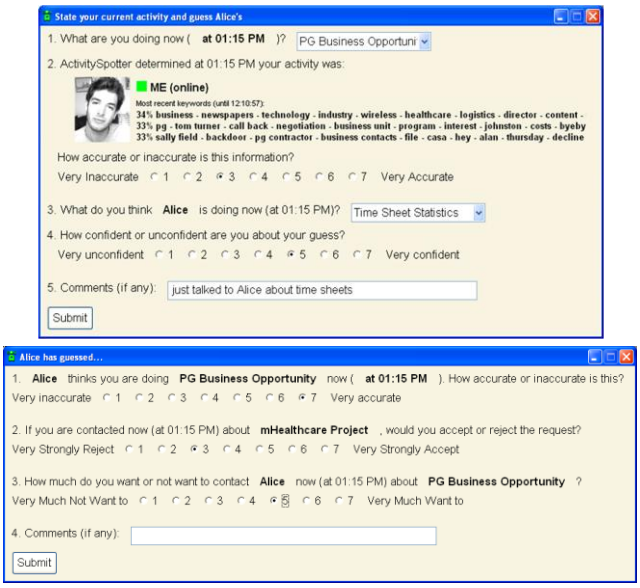

Figure 4. Pair of questionnaires (s1, s2) that participants receive multiple times a day asking them about their and their collaborator's activities and their contact receptivity and contact desire.

Participants used all versions in a within-subject design, for up to five working days per version with a balanced Latin Square arrangement to counterbalance for order and carryover effects. 32 participants were recruited from within one company in pairs as collaborators who shared at least one project (activity) between them.

5.1 Procedure

Participants filled out an initial survey capturing their company role, the projects they were involved in, and activities (or projects) they shared with their paired collaborator, and some that they did not share with the collaborator. The lists of activities were checked by both participants in each pair to verify that activities are correct and understood.

Each participant then installed three pieces of software: ActivityIndexer, ActivityConfigurator, and ActivitySpotter. ActivityIndexer runs first for about 10-30 minutes to index Microsoft Outlook emails from up to one month ago to do the topic modeling. After the indexing was done, participants ran ActivityConfigurator to map the 50 topics to their pre- specified activities (takes about 15-30 min). To ease this process for participants, one of the experimenters was present to help guide the each participant as s/he categorized their topics. Once categorization was finished, participants then started AS. AS would auto-start whenever users restart their computers so that they do not forget to turn it on.

5.2 Measures

Depending on how often both collaborators of a pair were concurrently online at their computer, participants received several pop-up awareness surveys a day. The frequency of pop-ups could be reduced if participants found it too high. Our surveys appeared in a sequence of two for each participant (see Figure 4). The second survey (s2) depended on information provided in the first (s1), and popped up once both participants in the pair answered the first. We asked participants (s1 q2) how accurate

Table 3. Measures from the end-of-phase surveys. Measures were composed from sets of three to four 7-pt Likert scale questions, except for Information Amount which only had two. Survey measures that are cited were adapted from their respective sources (cited). Number of questions in sets (# Q's) and reliability (Cronbach α) of measures indicated. Sets with 4 questions have 2 positive and 2 negative polarity questions.

Measure	Sample Questions (asked in sets of at least 3 questions; all 7-pt Likert: 1=Strongly Disagree to 7=Strongly Agree)	# Q's	α
Awareness with respect to time, Communication & common ground [5]	Over time, I became more and more aware of my partner's plans	3	.883
	My partner and I communicated well with each other	3	
Privacy / sharing	ActivitySpotter reveals too much information about what I am doing	4	.657
Obtrusiveness [5], [32]	ActivitySpotter is distracting	3	.314
Information Amount	I wish Activity Spotter provided more information about my collaborator	4	.880
Usefulness & Satisfaction [18]	ActivitySpotter saves me time when I use it	3	.932
	It is pleasant to use ActivitySpotter	3	
Accuracy Perception	I do not trust the information provided by ActivitySpotter	4	.672

they thought the activity information (Topic Keywords or Activity Labels) was (but not in the Presence only condition). To obtain ground truth for her current activity, we asked the participant what she was doing (s1 q1). We also asked her to guess what she thought her collaborator was doing (s1 q3), and her confidence in her guess (s1 q4). In the second survey, we asked her to rate how accurate her collaborator's guess was (s2 q1), and how receptive she would be if she were contacted about a randomly chosen activity (s2 q2, 50% probability of being her current activity, 50% not). Finally, we asked about her desire to contact her collaborator if that participant were working on a randomly chosen activity (s2 q3, 33% probability each of being her current activity, her collaborator's activity, or neither of theirs). All questions were either multiple-choice or 7-pt Likert scale.

At the end of each phase, participants filled out a survey asking about their perceptions of using that version of AS (see Table 3). Eight participants were selected for 30- to 45-minute exit interviews and asked in detail about their working relationship with their collaborator, how they used AS, their perception of its accuracy and usefulness, their strategies for guessing their collaborator's activities, and their concerns regarding obtrusiveness, distraction, and privacy when using AS.

6. RESULTS

We recruited 15 pairs of participants from the same company. Due to lack of data and participation, we had to drop two pairs. Another pair was dropped because one participant was uncomfortable sharing information with a subordinate. We dropped yet another pair because one had a computer was not powerful enough to run AS. The remaining 11 pairs were distributed across the three AS awareness type condition sequences (orders in our Latin square; 4 pairs in each of two, 3 pairs in the remaining one). While we recruited participants from the same company (similar to the work of [1], [15], [21], [34], [35]), our participants came from a multiple departments, and had

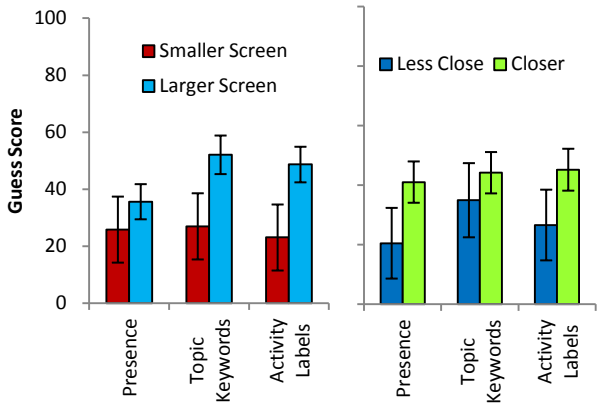

Figure 5. Guess correctness scores of participants using different Awareness Types, Screen Sizes (left) and Closeness (right). There was no significant effect due to Awareness Types in general. Participants who use larger displays had significantly better scores. Participants in closer pairs had marginally better correctness scores. Participants who used larger displays improve their guess more when using Topic Keywords or Activity Labels.

a myriad of job functions. These 22 participants included 14 research scientists, 1 in business development, 1 engineer, 2 in intellectual capital management, 2 principal scientists/fellows, 2 managers. Within-pair relationships were 2 manager-subordinate pairs and 9 co-contributors.

We define an activity as a project or major sub-job (e.g. client project, hiring), and define a shared activity as a project or subproject that both collaborators are involved in. Participants defined 2 to 10 activities/projects (Median=5), declared that they shared 1 to 8 of them (Median=2), and worked together about 1 to 50 hours per week (Median=5), in general. Unsurprisingly, due to people tending to be close to their collaborators, we had 7 pairs defined as close and 4 not so close.

Participants informed us that the surveys popped up occasionally at inconvenient times (e.g. when they were talking to someone at their office) and so they deferred them for a few minutes, hours, or even over a weekend. As the surveys depended on answering questions about activities as soon as they popped up, this led to stale or obsolete data. So we filtered out survey responses where the total time between when survey 1 was first sent, and when the second survey was done was within 5 minutes. We only retained surveys where both members of a pair answered their corresponding surveys. 638 responses were filtered out, leaving 968 responses for analysis.

Due to disparities in schedules between collaborators, some pairs had less time being online at the same time and thus fewer survey responses. 7 participants also took holidays or were away on business trips, so we paused their pair for those days. The number of days that participants were in each activity type condition also varied, from two days (one pair only) to six days. Excluding holidays, participants were involved in the study from 2 to 6 days per condition. Participants filled out between 5 (capped lower bound) and 51 survey pairs (Median=19.5) per condition within the acceptable criteria.

Even though participants only trained AS at the beginning of the study, no significant drop in accuracy was measured across the three phases of the study (p=n.s.).

6.1 ActivitySpotter Accuracy

All versions of AS recorded presence information, topic keywords and activity labels associated with the user's current documents regardless of whether the information was shown to participants. We compared the activity labels determined by AS with the self-reported activities of the participants. This was independent of the activity type condition because what was displayed did not influence the classification. Up to three activities were listed by AS each time, and if the list contained the participant's activity, then there was a match, and AS was considered to have made a correct classification. Because AS was trained differently for each participant, we computed the means of correctness per participant (M=51.1% correct, Std Err=3.9%). Note that if each participant were to randomly guess among their numbers of activities, the mean accuracy would be M=15.0%, Std Err=0.9%.

In all tests mentioned henceforth, we nest participants in pairs and nest these pairs in Awareness Type. Both participants and pairs are set as random effects.

We conducted an ANOVA with Awareness Type as a main effect (R^2=.265). In survey 2, participants rated the accuracy of the Activity Labels version of AS as marginally higher than the Topic Keyword version (M=3.18 vs. M=2.91, p=.09).

6.2 Closeness and Screen Size

From interviews and observation, we realize two factors that may influence how well participants may guess their collaborator's activity: closeness and screen size. **Closeness** is whether the participant is close to his collaborator. Closeness was derived from these criteria:

- *Physically proximate*: in adjacent or nearby offices.
- *Socially close*: communicate frequently, about every day via face to face, IM, phone, or email conversations.

We hypothesize that participants of closer pairs would be more aware of each other and would be more able to guess what their collaborators are doing. **Screen size** is whether the participant uses a large desktop monitor (>19"), or has multiple monitors (>2 monitors, or desktop + laptop). We hypothesize that users with larger screen sizes would be more willing to keep AS displayed (rather than minimized), and would use it (or remember to use it during the surveys) more often.

6.3 Correctness of guessing collaborator activity

We were interested to see if participants were better able to guess what their collaborator was doing when they received information from AS. Guess correctness was computed as a composite score of whether what the participant guessed *matched* what his collaborator self reported (0 or 1), the *confidence* reported by the participant (7-pt Likert scale: 1 to 7) and the collaborator's *rating* of the participant's guess (7-pt Likert scale):

$$Correctness = \left[Matched(guess, actual) \times \left(1 + \frac{Confidence - 1}{6}\right) + \frac{Rating - 1}{6} \right] \times \frac{100}{3}$$

This is normalized to a 100-point score, matched and rating scaled to 1, and confidence is scaled to 2 . We do not factor in confidence if there is no match. The mean score for participants is 41.2.

We conducted a two-way ANOVA with Awareness Type, Closeness and Screen Size as main effects and Awareness Type × Closeness and Awareness Type × Screen Size as interaction effects (R^2=.806). Participants who had a larger or more displays had higher guess scores than those with smaller or fewer displays (M=45.5 vs. M=25.3; p=.045). Participants of closer pairs had marginally higher scores (M=43.5 vs. M=27.4, p=.1). Participants who used larger displays also had marginally higher guess scores when they received activity content information (contrast of Presence awareness < {Topic Keywords & Activity Labels}, p=.045). Furthermore, among these participants, this effect was marginally significant for socially distant pairs (contrast, p=.068), but not for those who were more close (p=n.s.).

6.4 Behavioral Measures

We conducted a one-way ANOVA with Awareness Type as the main effect and the time it took participants to do the first survey that asked them to guess their buddy's activity (R^2=.074). We found there were marginally significant differences (p=.06). A pairwise Student's t test found that the participants completed the Presence awareness type fastest (M=35.0sec, Std Err=3.5, p<.05). A one-tailed (because we hypothesized H2) contrast found that participants took marginally longer when using Topic Keywords type than Activity Labels (M=43.7sec, Std Err=3.7 vs. M=38.8sec, Std Err=3.4; p=.09).

6.5 Contact Desire and Contact Receptivity

The concurrency of activity of participants and their collaborators (neither, self, collaborator, both) is calculated from matching the randomly chosen activity in the survey question about contact desire (s2 q3) to the activity that each participant in the collaboration pair reported that he or she was doing. For example, if the randomly chosen activity is one of the defined shared activities, but only the participant's collaborator was doing it at the time, then the concurrency is "collaborator."

To simplify the analyses and increase statistical power, we combined Awareness Type to two values Presence and Content-Aware (Topic Keywords and Activity Labels), and combined Concurrency into two values: when Both collaborators were working on a shared activity at the same time, and Neither+Either for other cases. To test contact desire, we conducted a two-way ANOVA with Awareness $Type_2$ and Concurrency as main effects and Awareness $Type_2$ × Concurrency as an interaction effect (R^2=.293). Participants had a higher contact desire when they and their collaborator were both working on a shared activity at the same time compared to other times (M=2.98 vs. M=2.63, p=.004). Participants moderated their context desire when using the content-aware versions of AS, *i.e.*, they lowered their contact desire when both of they and their collaborators were not working on a shared project (M=3.00 vs. M=2.52, p<.001). On the other hand, there was no significant difference between concurrency types when they used the Presence only version of AS (p=n.s.).

To test contact receptivity, we conducted a two-way ANOVA with Awareness Type and Current Activity as main effects and Awareness Type × Current Activity as an interaction effect (R^2=.468). Current Activity refers to whether the activity the participant is being asked about is what he is currently doing. There was no effect due to Awareness Type (p=n.s.). However, participants had higher contact receptivity if they were to be contacted about what they are currently doing (M=4.79 vs. M=3.91, p<.001) than if they were to be contacted about something else.

Table 4. Factors considered for multi-level analysis of various end-of-phase survey measures. Grayed out indicate not examined as a factor. R^2 denotes how well the considered factors model the data.

Measure	Awareness Type	Closeness	Screen Size	R^2
Awareness & Comm	✓	✓	✓	.885
Privacy / sharing	✓	✓		.279
Information Amount	✓			.402
Usefulness / Satisfaction	✓	✓	✓	.832
Accuracy Perception	✓	✓		.733

Table 5. Results of the end-of-phase surveys. Tests are conducted as multi-level one-way ANOVA with up to three factors. Values are on balanced 7-pt Likert scales. Underlined indicates significantly lower than other groups, * indicates significant difference.

	Awareness Type				
Measure	**$M_{Presence}$**	**M_{Topic}**	**$M_{Activity}$**	**F**	**p > F**
Awareness & Comm.	2.60	2.69	3.08	-	n.s.
Privacy / sharing	4.37	3.86	4.10	7.13	.004*
Information Amount	3.00	3.67	3.54	4.58	.02*
Usefulness / Satisfaction	2.34	2.26	2.49	-	n.s.
Accuracy Perception	3.49	2.98	3.23	5.32	.01*

	Closeness			
Measure	**M_{Close}**	**$M_{Less\ Close}$**	**F**	**p > F**
Awareness & Comm.	3.26	2.23	7.51	.01*
Privacy / sharing	3.85	4.37	3.23	.09
Information Amount				
Usefulness / Satisfaction	2.58	2.14	-	n.s.
Accuracy Perception	3.18	3.29	-	n.s.

	Screen Size			
Measure	**M_{Larger}**	**$M_{Smaller}$**	**F**	**p > F**
Awareness & Comm.	3.23	2.25	7.26	.02*
Privacy / sharing				
Information Amount				
Usefulness / Satisfaction	2.82	1.90	5.05	.03*
Accuracy Perception				

6.6 End-of-phase Surveys

Table 3 shows reliability of scale measures used in the end-of-phase surveys. Table 4 shows the factors that were analyzed for each measure. Questions measuring obtrusiveness are not well correlated with one another and not analyzed further. Results are summarized in Table 5. **Awareness with respect to time, Communication & common ground:** Participants who were in closer pairs and had larger or more displays were more aware of collaborator activity. **Privacy / Sharing:** A Tukey HSD test found that the Topic Keywords Awareness Type provoked the most privacy concerns ($p<.05$). **Information Amount** varied across Awareness Types, and participants felt that the Presence awareness type was the least informative (contrast Presence vs. others: $p=.006$). **Usefulness / Satisfaction**: Participants were less satisfied with AS when they used smaller screens ($p=.03$). **Accuracy Perception:** A Tukey HSD test found that participants perceived the Topic Keyword activity information to be the least accurate and trustworthy ($p<.05$).

7. DISCUSSION

In this section, we discuss our quantitative results in light of qualitative information learned from participant interviews. Table 6 summarizes the results in terms of our hypotheses.

Table 6. Summary of hypotheses and whether our results satisfied them.

	Measure	Hypothesis	Verified?
H1	Accuracy	Topic Keywords < Activity Labels	Marginal
H2	Cognitive Load	Activity Labels < Topic Keywords	Yes
H3a	Awareness (Guess correctness)	Presence < {Topic Keywords & Activity Labels}	Yes, & with moderators
H3b		Topic Keywords < Activity Labels	No
H4	Awareness → Contact Desire	Contact desire moderated by awareness of activity concurrency	Yes

H1: *Activity Labels are perceived to be marginally more accurate than Topic Keywords.* Participants reported (in the survey) that both content-aware versions of AS were not very accurate (Median<4), but they felt that the Topic Keywords were less accurate, and some mentioned errors they saw in the keywords. Some participants thought that the keywords barely changed:

"A lot of the time it seems that the same keywords would be there." — P.2A (Participant A from pair 2)

Also, Participant 9A thought that the words that appeared as keywords also tended to be unrelated to what he thought his collaborator was doing.

H2: *Participants experience more cognitive load when interpreting Topic Keywords than Activity Labels.* Participants took marginally longer (by 4.9 sec) to do the first survey when using the Topic Keywords version than the Activity Labels version of AS. This suggests that they took longer to interpret the keywords so that they could judge whether AS was accurately representing their activity and that they took longer to interpret the keywords of their collaborator's activity if they referred to them. Participants indicated that they appreciated the Activity Labels display version of AS where they could "quickly assess" what their collaborator was doing:

"The keywords [and presence] were not that helpful. … Later on it got more useful when it got "fixed" [activity labels version]" — P.5A

"I found those [Activity] Labels to be better because it is not just a bunch of words, that I don't know where they're pulled from." — P.12A

Along with [9], this corroborates that simplifying awareness information can reduce cognitive load on users and increase their rate of assimilating that information.

H3a: *Providing activity content information (Activity Labels or Topic Keywords) can help participants be more aware of what their collaborator is doing, especially if they have a **big display**, and are **not already close** to their collaborator.* Our results agree with [29] that closer pairs of participants already have a higher awareness of each other and less close pairs may benefit from AS more than closer pairs. Indeed, some participants pointed out that they already have frequent contact multiple times a day so AS did not really help them. Rather, they could rely on "out-of-band" information gleaned from earlier face-to-face interactions, emails, IM chats, meetings, and so on to learn about approaching events and deadlines that would influence their collaborator's likely activity:

"They were purely about my intuition about what deadlines we had, whether I just talked with [buddy]." — P.12B

"I often had out-of-band information of what [buddy] was doing. …" — P.3A

"I get updates pretty frequently from him, just by me walking to his office and him walking to my office." — P.8B

Since AS did not recognize certain applications and files (e.g. programming interfaces), it was less useful to two participants who did a lot of programming work during the study period. Participants 12A and 12B spent most of their time programming and said they stopped paying attention to AS after having realized it was not displaying their activity:

"Email that I selected usually had nothing to do with what I was doing most of the day, because I was working on Visual Studio most of the day." — P.12A

"The email that I'm reading is typically not related to what I'm actually doing. So whatever email came in, I would glance at it." — P.12B

However, some participants did benefit in their awareness of their collaborator's activity from referring to AS. Our results indicate that it was mainly users who had larger screen sizes (>19") that were willing to keep AS displayed to improve awareness of their collaborator. AS was clearly not useful enough to trade screen space for on a small screen:

"With [buddy] I didn't have much need for fast synchronous communication, so often it was not terribly useful to know what [buddy] was doing. Actually, this is one of the reasons I was minimizing it most of the time." — P.3A

"I never expanded [AS, to see the keyword list] because it would then take up too much space on my screen." — P.9A

"Another reason why I probably didn't use the keywords more was because I always had to [expand] the windows more, to make it have enough sense of the list." — P.14A

Clearly, activity awareness displays need to be compelling (and accurate enough) for users to be willing to sacrifice screen real-estate.

H3b: *Topic Keywords are not worse than Activity Labels at raising participants' awareness of their collaborators.* Our results did not indicate that participants had more awareness of their collaborators due to being able to interpret Activity Labels more easily than Topic Keywords. In fact, in interviews, participants indicated that they did not have difficulty understanding what the keywords meant:

"I found that the topic keywords were interpretable." — P.12A

However, even though participants retrospectively felt that the information qualities of Topic Keywords are similar to Activity Labels, they were able to more quickly interpret Activity Labels (see H3a). Participants felt that they did not need the details provided by the keywords. Given the inaccuracy of the Topic Keywords at some times, some participants did not bother to expand the display to read the words:

"I didn't think the keywords were too accurate, so there was no incentive for me to expand. So if the first one was inaccurate, I didn't think that the remainder would give me any more information." — P.9A

This is further evidence that low-level, more detailed, Topic Keywords may bring disadvantages when describing activity content to users.

H4: *ActivitySpotter influenced participants to moderate their desire to contact their collaborators.* In agreement with [24], our results show that, when made more aware of what their collaborators are doing through AS, participants were socially appropriate and had a higher contact desire if there was concurrency in their own and their collaborator's activities (i.e., both collaborators working on a shared activity vs. not). This socially appropriate intention would be welcome because participants also indicated that they prefer being contacted about what they are currently doing. In spite of the lack of difference in disruption cost that may arise due to lack of concurrency [26], facilitating contact about aligned activities may be more acceptable and useful.

7.1 Relevance of Activity Content

A few participants reported in interviews that they used AS mainly for presence information and did not much care for the content information in the Topic Keywords and Activity Labels versions. They felt that interruptibility and availability information was sufficient and content did not matter:

"I use [AS] as an awareness indicator. So when I see what he is doing, I know he is there, so I walk up to his office." — P.12A

Participant 5A said that he would only note "how" his collaborator is working, rather than "what" on to decide whether to contact him: *"[I would rather know] what he is working on — no [does not influence whether he would interrupt], how he is working [e.g. on the telephone] — yes."*

On the other hand, Participant 9A said that he used the content information to keep informed of what his collaborator was doing. Although he was not directly involved in some projects, he would then talk casually with his collaborator about those activities when he ran into him in the hallway.

"The active projects that I manage, he's not actively working on them at the right stage. He does some of the related work to that. So the synchronization wasn't exactly right for me to be very interested in what he was doing." — P.9A

Participant 14A suggested presenting activity information in terms defined by the person viewing the information (rather than the person it is about) since it is meant for his consumption.: *"[Buddy] uses different labels that I would for his activities. He partitions his activities in more categories than I would."* Participant 9A indicated that he would be interested in activities being displayed from his collaborator if they were "*relevant* to his core area of interest, *new* or *unexpected*, and *sustained* (not transient)." This would serve to expand his situational awareness of what people are doing around him. These remarks point to possible requirements for any future activity awareness tools.

Several participants felt that AS would be more useful for remote collaboration, since they are less intimately aware of what their off-site collaborators are doing.

"Well, I'm working with some people [at a remote location] on this project, and I have very little insight into how they spend their time on their project. We exchange email all the time, because we are worried about interrupting each other at inopportune times. This might be more useful for our collaboration. For example, I just received an email today, and I asked 'why didn't you just call me with that question?'" — P.14A

7.2 Temporal Granularity of Activity Content

Even though some participants did indicate that knowing activity content was potentially useful, some of them did not wish to share the information at such a fine level of temporal granularity as is presented by AS:

"Monthly average is ok, but I would not like them to see on a day-to-day or hour-by-hour basis." — P.5A

"The granularity of activity detail I needed was just day-by-day, and I didn't need finer detail." — P.12A

Participant 5A also suggested showing historical information in a glanceable (*vs.* merely readable) visualization: *"It would be helpful to maybe, rather than see percentages, to see a bar chart or a schedule across the day, like a slider."*

7.3 Privacy and Sharing Concerns

Privacy was an obvious concern that we thought users of AS might be worried about. In fact, one participant dropped out right after installing the software because he realized he was not comfortable sharing his activity information with a subordinate. In general, though, the end-of-phase survey responses indicate that privacy was not a serious concern (mostly neutral). However, privacy was of greatest concern with the Topic Keywords. In an interview, Participant 8B mentioned that he feared that the keywords might pick up and share a "misspoken" word and compromise his professionalism: *"I guess there is always the risk that the keywords may show words that you do not want to be displayed. … It's not that I do that, but let's say you email somebody and you call him an idiot. It would be a word that is displayed."* Though this fear is legitimate, AS would actually not display previously unidentified words during training. Participant 3A insightfully pointed out that privacy is contextual and can change with respect to time and activity: *"there are keywords that in the context that I reviewed them at the beginning, but at another time, they may be more sensitive."* Participant 9A mentioned that he forgot what he had set as private and not private since categorization, and that his concern had grown. A refresher on his settings would have been helpful. Some participants (e.g. 9B and 5A) were concerned about sharing activity information with someone they were accountable to, as AS could provide a means to track whether one is spending one's time well.

8. FUTURE WORK

In contrast with [1], [15], [26], the results from our study indicate that there may be some value in mediating contact behavior in order to make it more content appropriate. The strategy used by ActivitySpotter of attending to contact intention of the contactor / interrupter is promising in that it can moderate a user's contact intention. With information and control on the side of the contactor, AS inherently assumes that the information is useful enough for her to employ, and that she will be socially appropriate and moderate her contact intention, given her extra awareness. But an alternative approach would be to place control more on the side of the receiver / contacted collaborator. For example, a user could instruct an AS-like tool to allow an IM message to pass if it determines that both the contactor and recipient are currently working on the same activity, and reduce the likelihood of allowing the message to pass otherwise. However some writers ([15], [36]) have reported problems with too much automation in the loop, which would have to be overcome perhaps by merely delaying off topic contact attempts so they can be inspected later.

A number of participants indicated preference for just presence or availability information over content. However, others do believe that there is value in activity content information to improve one's awareness of one's collaborator, especially if this could be more accurate than what AS provides. In this study, we focused only on activity content information rather than activity context information (such as presence, availability [1], [15], interruptibility [14], breakpoints [22], calendar schedules, location). In future work, a combination of activity *context* and *content* information could provide significant improvements to help collaborators manage their awareness and contact intents.

Our results also suggest that AS may be more useful for collaborators who are remotely separated [29]. AS simply seems to provide the kind of information that can be gathered through chance encounters or overheard conversations amongst physically proximate collaborators, albeit on a more continuous basis. Since our study took place in just one building, future work could productively examine benefits for even more distant collaborators.

Our experiment set participants up in pairs and investigated how they made use of activity information about each other, but a more externally valid arrangement might be to investigate the use of AS with a social network of collaborators with multiple connections per person. Participant 9A had indicated an interest in knowing about a group of people that he manages.

9. CONCLUSION

We have proposed a means to identify and abstract a user's activity content (topic) from the documents she accesses. We developed ActivitySpotter as an awareness display and experimental instrument for users to share their activity information with their collaborators. We hypothesized that this display would be able to raise awareness of users, but we did not find significant improvements except for users who used large or multiple displays or were more distant (socially and physically) from their collaborators. We also hypothesized that, when made more aware of what their collaborators are doing, users would want to contact their collaborators more when their activities align or are concurrent, rather than when neither of them is working on the same shared activity. They would do this to minimize the disruption on the recipient by interrupting about an activity irrelevant to what the latter is working on. We found that users did indeed adjust their contact desire due to the increase in awareness afforded by ActivitySpotter. In its current state as a research tool, ActivitySpotter had some issues of usability and accuracy, but we have learned various ways users use and benefit from such activity awareness. Furthermore, our results suggest a value in sharing activity content information. Though participants did not particularly use ActivitySpotter during the short study period, some expressed appreciation for the increased situational awareness. Perhaps a more compelling awareness display should cohesively provide information about activity context and content with richer visualization, improved glanceability.

10. REFERENCES

[1] Begole, J. B. and Tang, J. C. 2007. Incorporating human and machine interpretation of unavailability and rhythm awareness into the design of collaborative applications. *Hum.-Comput. Interact.* 22, 1, 7-45.

[2] Bellotti, V., Thornton, J. D., Chin, A., Schiano, D. J., Good, N. 2007. TV-ACTA: embedding an activity-centered interface for task management in email. *CEAS 2007*. Mountain View CA.

[3] Blei, D.M., Ng, A.Y., Jordan, M.I. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research 3*: pp. 993–1022.

[4] Brush, A. B., Meyers, B. R., Scott, J., and Venolia, G. 2009. Exploring awareness needs and information display

preferences between coworkers. *CHI '09*. ACM, New York, NY, 2091-2094.

[5] Cadiz, J. J., Venolia, G., Jancke, G., and Gupta, A. 2002. Designing and deploying an information awareness interface. *CSCW '02*. ACM, New York, NY, 314-323.

[6] Convertino, G., Neale, D. C., Hobby, L., Carroll, J. M., and Rosson, M. B. 2004. A laboratory method for studying activity awareness. *NordiCHI 2004*, Tampere, Finland, ACM Press, p.313-322.

[7] Czerwinski, M., Cutrell, E., and Horvitz, E. 2000. Instant messaging: Effects of relevance and timing. *HCI'00*, British Computer Society, 71-76.

[8] Czerwinski, M., Horvitz, E., and Wilhite, S. 2004. A diary study of task switching and interruptions. *CHI '04*. ACM, New York, NY, 175-182.

[9] Dabbish, L. and Kraut, R. E. 2004. Controlling interruptions: awareness displays and social motivation for coordination. *CSCW '04*. ACM, New York, NY, 182-191.

[10] Dredze, M., Wallach, H. M., Puller, D., and Pereira, F. 2008. Generating summary keywords for emails using topics. *IUI '08*. ACM, New York, NY, 199-206.

[11] Dourish, P. and Bly, S. 1992. Portholes: supporting awareness in a distributed work group. CHI '92. ACM, New York, NY, 541-547.

[12] Dragunov, A. N., Dietterich, T. G., Johnsrude, K., McLaughlin, M., Li, L., and Herlocker, J. L. 2005. TaskTracer: a desktop environment to support multi-tasking knowledge workers. *IUI '05*. ACM, New York, NY, 75-82.

[13] Fitton, D. Cheverst, K., Rouncefield, M., and Dix, A. 2007. Exploring adoption in the Hermes door display deployment. *UbiComp in the Office Workshop '07*, Springer.

[14] Fogarty, J., Hudson, S. E., Atkeson, C. G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J. C., and Yang, J. 2005. Predicting human interruptibility with sensors. *ACM Trans. Comput.-Hum. Interact.* 12, 1, 119-146.

[15] Fogarty, J., Lai, J., and Christensen, J. 2004. Presence versus Availability: The Design and Evaluation of a Context-Aware Communication Client. *International Journal of Human-Computer Studies (IJHCS)*, Vol. 61, No. 3, pp. 299-317.

[16] Greenberg, S. 1996. Peepholes: low cost awareness of one's community. In *Conference Companion on Human Factors in Computing Systems: Common Ground* (Vancouver, British Columbia, Canada, April 13 - 18, 1996. M. J. Tauber, Ed. *CHI '96*. ACM, New York, NY, 206-207.

[17] González, V. M. and Mark, G. 2004. "Constant, constant, multi-tasking craziness": managing multiple working spheres. *CHI '04*. ACM, New York, NY, 113-120.

[18] Gutwin, C. 1997. Ph.D. Dissertation, Department of Computer Science, University of Calgary.

[19] Hofmann, T. 1999. Probabilistic Latent Semantic Indexing. In *Proc. SIGIR-99*.

[20] Hudson, S. E. and Smith, I. 1996. Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. *CSCW '96*. ACM, New York, NY, 248-257.

[21] Huang, E. M. and Mynatt, E. D. 2003. Semi-public displays for small, co-located groups. *CHI '03*. ACM, New York, NY, 49-56.

[22] Iqbal, S. T. and Bailey, B. P. 2008. Effects of intelligent notification management on users and their tasks. *CHI '08*. ACM, New York, NY, 93-102.

[23] Lund, A. 2001. Measuring usability with the USE questionnaire. *Usability Interface*. Retrieved on 9 July 2009 from http://www.stcsig.org/usability/newsletter/home-0110.html

[24] MacIntyre, B., Mynatt, E. D., Voida, S., Hansen, K. M., Tullio, J., and Corso, G. M. 2001. Support for multitasking and background awareness using interactive peripheral displays. *UIST '01*. ACM, New York, NY, 41-50.

[25] Mark, G., Gonzalez, V. M., and Harris, J. 2005. No task left behind?: examining the nature of fragmented work. *CHI '05*. ACM, New York, NY, 321-330.

[26] Mark, G., Gudith, D., and Klocke, U. 2008. The cost of interrupted work: more speed and stress. *CHI '*08. ACM, New York, NY, 107-110.

[27] McEwan, G., and Greenberg, S. 2005. Supporting Social Worlds with the Community Bar. *Group'05*.

[28] Nardi, B. A. 1996. A Small Matter of Programming – Perspectives on End User Computer. MIT Press. 1993.

[29] Olson, G. M. and Olson, J. S. 2000. Distance matters. *Hum.-Comput. Interact.* 15, 2, 139-178.

[30] Rattenbury, T. and Canny, J. 2007. CAAD: an automatic task support system. *CHI '07*. ACM, New York, NY, 687-696.

[31] Salton, G., and McGill, M, eds. (1983. *Introduction to Modern Information Retrieval.* McGraw-Hill.

[32] Stasko, J., McColgin, D., Miller, T., Plaue, T., and Pousman, Z. 2005. Evaluating the InfoCanvas Peripheral Awareness System: A Longitudinal, In Situ Study. Graphics, Visualization, and Usability Center, Georgia Institute of Technology, Atlanta, GA, Technical Report GIT-GVU-05-08.

[33] Tang, J. C., Yankelovich, N., Begole, J., Van Kleek, M., Li, F., and Bhalodia, J. 2001. ConNexus to awarenex: extending awareness to mobile users. *CHI '01*. ACM, New York, NY, 221-228.

[34] Tee, K., Greenberg, S., and Gutwin, C. 2006. Providing artifact awareness to a distributed group through screen sharing. *CSCW '06*. ACM, New York, NY, 99-108.

[35] Tullio, J., Dey, A. K., Chalecki, J., and Fogarty, J. 2007. How it works: a field study of non-technical users interacting with an intelligent system. *CHI '07*. ACM, New York, NY, 31-40.

[36] Whittaker, S., Frohlich, D., and Daly-Jones, O. 1994. Informal workplace communication: what is it like and how might we support it?. *CHI '94*. ACM, New York, NY, 131-137.

[37] Wiberg, M. and Whittaker, S. 2005. Managing availability: Supporting lightweight negotiations to handle interruptions. *TOCHI 2005*, 356-387