Interruptability Prediction Using Motion Detection

Peter Vorburger, Abraham Bernstein, Alen Zurfluh

Department of Informatics, University of Zurich, Zurich, Switzerland
{vorburger, bernstein, azurfluh}@ifi.unizh.ch

Abstract. People are subjected to a multitude of interruptions. This paper introduces an approach to predict a person's presence and interruptability in an office-like environment. To determine these two states we record data as audio, motion detection, and the time of the day representing a person's context. We show that motion detection data outperforms audio data both in presence and interruptability prediction.

Introduction

An ordinary office day is often disrupted by interruptions. Some of them might provide some benefit to an office worker [1, 16] but others are annoying and detrimental to work performance.

This paper investigates a novel way of predicting a person's interruptability in an office based setup. More specifically, we gather information about the subject's context by considering audio, motion detection, and the time of the day. First, we show how well we can predict the subject's presence in the office from our observations where motion detection outperforms the other two data streams. Second, we show that we can predict his/her interruptability with good accuracy. Furthermore, we demonstrate that motion detection is superior to audio and that a combination of all data streams reaches even higher prediction accuracy. Finally, we also show that dividing the motion detection information into different sectors representing different activity regions of the subject improves the prediction power. The presented analysis takes into account the temporal aspect of office work by using machine learning algorithms that can detect temporal patterns.

The paper is organized as follows. First, we succinctly discuss the related work focusing on interruptability prediction and context-awareness, specifically for office-like environments. Then, we introduce the experiment we conducted to support our claims. This involves presenting the methodology and the technical setup. After evaluating the gathered data and analyzing the results we close with a discussion and the prospect of future work.

Related Work

Ever since the landmark study by Minzberg [15], researchers have investigated the activity of knowledge workers in organizational settings to find ways on how to improve their performance. In the past years, researchers have started to pay more attention to the effect of interruptions on individual performance. This issue has become increasingly important as new technologies are likely to increase the number of occasions for interruption [11].

One stream of research is focused on finding the cost of interruption [4, 5]. Extending this line of work Hudson and colleagues [8, 9] use an empirical sampling technique and qualitative interviews to find research managers' attitude towards interruptions.

In a series of studies McFarlane, [12-14] examined four methods for deciding when to interrupt someone during multitasked computing.

Summarizing the (recent) findings on the effect of interruption on people we can conclude that all studies agree that *interruptions can, indeed, have disruptive effects on performance*. The studies, furthermore, conclude that *the extent of the effect is heavily dependent on the current context* of the interrupted person *as well as the nature of the interruption*, whether it contains a desired piece of information or not.

Consequently, it is important to be able to predict the nature of an interruption and the current context of a person's activity in order to reduce the interruption's detrimental effects and improve overall work performance.

In a study called "Coordinate" [7], a prototype service is presented that logs all the meeting information stored in a user's calendar and several additional properties. They predict a person's attendance at the meeting, a person's interruptability, and location. Extending this line of work, Horvitz and colleagues [6] present methods for inferring the cost of interrupting users.

Hudson and colleagues [9] present a so-called "Wizard of Oz" feasibility study to predict people's interruptability. They simulate a sensor-equipped office using a video and audio recording of the office, which are then hand-coded by people determining features such as the number of people currently in the office, who is speaking, what task objects are being manipulated, whether the phone is on or off the hook, and other similar facts about the environment. In a follow-up study [3] they equipped an office with real physical sensors. They placed microphones in the office, logged the beginning and end of each non-silent interval, after applying a speech recognition tool that detected conversations. Additionally, two magnetic switches, one near each side of the top of the door frame, allowed them to sense whether the door was open, cracked, or closed. Two motion sensors were put in each office. Another magnetic switch was used to determine whether a person's phone was physically off its hook. Software on each subject's computer logged, once per second, the number of keyboard, mouse move, and mouse click events in the previous second. It also logged the title, type, and executable name of the active window and each non-active window. The interruptability annotation was done the same way as introduced in the preceding study by audibly prompting the subjects (i.e., self-reporting). All this information together with the interruptability annotation was used to build interruptability predictors. To gain insight into the generalization of their method they measured three different types of subjects: "managers", "researchers", and "interns".

They found that statistical models of interruptability should adapt to the people and they showed that audio is the most predictive dimension for interruptability in their setup.

There also exist several wearable setups to predict a person's interruptability like the SenSay project [18] or [10]. Nomadic Radio [17] is a wearable computing platform. The platform is audio-only and uses speech recognition, message priority, as well as a contextual notification model to define when a message should be posted on the user's heads-up display. As contextual information they use the likelihood of conversation obtained by melscaled filter-bank coefficients and pitch estimates to discriminate a variety of speech and non-speech sounds. The notification type is then based on the likelihood of speech. Unfortunately, no prediction performance results are reported.

Summarizing, we can say that we did not find any study with the same experimental setup (in particular, none of them used any *motion detection* sensors).

Experiment

Requirements to the Collected Data

In order to be able to make the desired predictions we needed to collect sensor data containing sufficient information about the subject's context in its environment, i.e., his/her office. We, therefore, decided to record both audio and video as well as self-reports provided by the subject on interruptability.

For the motion detection recording we used a camera reporting changes in different sectors of the office as dynamics might be a significant indicator of someone's context or context changes. For simplicity, we did not consider face recognition or any other high level image recognition techniques in this work. A microphone recorded the auditory surrounding of the person in the office.

We prompted the subject to report his/her level of interruptability. The self report of his/her level of interruptability (e.g. How disrupting would an incoming phone call?) has been broken down to five classes in a range from "ok, I don't care" to "do not disturb".

Method

According to [2], there are three ways to conduct experience sampling:

1. *Interval contingent*: Sampling occurs at regular intervals.
2. *Event contingent*: Events of interest trigger the sampling procedure.
3. *Signal contingent*: Sampling is performed randomly over a period of time.

Our concept of the interruptability self-report on a regular basis corresponds to a mixture of interval and signal contingent experience sampling. To ensure an upper and lower limit of the number of annotations we generate acoustic signal (or "beep") every 15 minutes. We also used a variance of 10 minutes on the signal to avoid

"training" the users to expect the signal and thus altering their behavior. According to the subject the frequency of "beeps" turned out not to be too disruptive after some days of experimentation but their occurrence was still frequent enough to collect a significant number of self-reports.

The subject was asked to adhere to the following directions during the experiment. When a "beep" occurs the subject performed self-report (assuming that the person is present in the office). Due to this underlying assumption we can also gather information about the subject's presence in the office.

Data Collection Setup

The environment of the experiment is an office with three work places/locations (Figure 1). The office is typically used by one person only. The two remaining seats are used sporadically by other people as well as by the subject. The subject is a researcher [3].

The audio and video data were recorded by an off-the-shelf webcam (Logitech QuickCam Pro 4000). We added a wide angle lens widening the aperture from 45° to 75°, such that the entire office could be overviewed as shown in Figure 1. The audio recording was set to CD quality but mono instead of stereo (i.e., the settings are 16bit, 44.1 kHz, mono). The recorded video had 320*240 pixels (in color) at 25 frames per second. We compressed the video stream using the XviD codec setting I420 to ensure that one day of recording would fit on one DVD, while ensuring good recording quality. The recorded files were saved in "avi"-format for further processing, keeping the audio and video streams synchronized.

For the self-report we used a modified keyboard. All information sources were collected on a single PC on working days from 8.15am to 6.15pm.

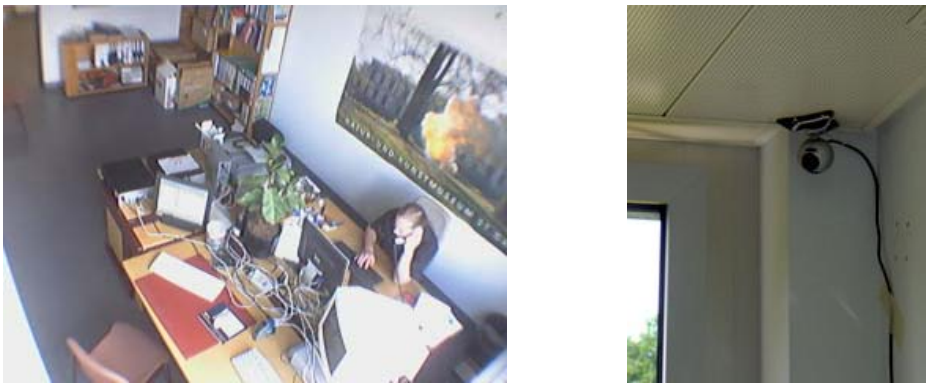

Fig. 1. On the picture on the left, the office is seen through the webcam. The picture on the right, shows the camera mounted in the corner of the office.

Sources of Interference

The experiment was conducted in a real-life environment. As a consequence, much interference influenced the experiment. In this section we provide a list of possible interferences.

Fig. 2. Sources of video inferences. On the left, another person than the subject is in the office. On the right an open window covers the office partly.

The video used as motion detector was sensitive to all kinds of movement in the office. Therefore, quality of collected data suffered from the presence of people other than the subject in the office – especially, when the subject was not present as seen in Figure 2 on the left. Furthermore, disturbances such as objects (like an open window) covering part of camera's view or changing brightness influenced the recording quality.

Background noise interfered with the audio recordings. Sources of such background noise originate from outside the office (e.g., people chatting on the corridor, or the neighing horse on the paddock next to the university) or from inside the office (e.g., computer ventilators).

Error sources in the annotation procedure stem from the subject ignoring "beeps" as well as inadvertent annotation mistakes. Addressing this risk we implemented a control mechanism using a feedback message for impossible annotation sequences.

Finally, as a matter of course this experiment was influenced by the experiment itself. The "beeps" prompting for a report on the subject's interruptability were disruptive for the subject.

Preprocessing

Beside the synchronization of all data streams we had to preprocess the raw audio and video data to get the most appropriate features for our problem. First, we extracted the features from audio (spectral center of gravity, temporal fluctuations of spectral center of gravity, tonality, mean amplitude onsets, common onsets across frequency bands, histogram width, variance, mean level fluctuations strength, zero crossing rate, total

power, and the 10 first cepstral coefficients) as described in [19][1]. This resulted in audio feature vectors of 20 features and one feature vector for every second of the recording.

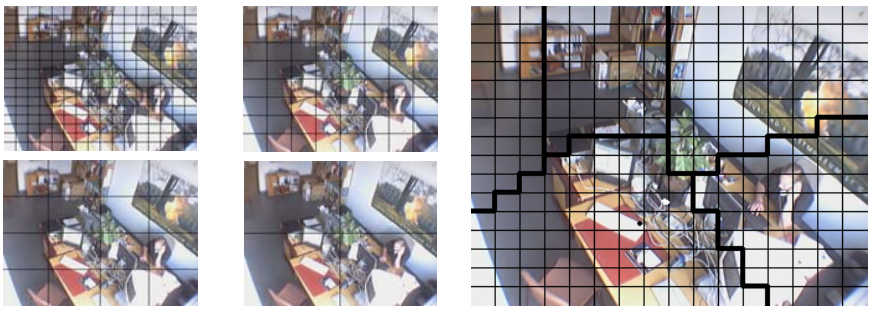

Fig. 3. Illustration of the motion detector features.

As we used video as motion detector we calculated the changes between two frames separated by one second in the video stream. We divided each frame into rectangles of the size of 15x20 pixels resulting in 256 (16x16) distinct fields. To measure the motion in the office we summed the number of changed pixels between the two frames. Hence, for every second of the recording we obtained feature vectors of the size of 256. Based on this large feature set we calculated smaller sets by summarizing the values of adjacent rectangles such that we got feature sets of the size of 64 (8x8), 16 (4x4), 4 (2x2), and 1 as depicted in Figure 3 on the left. Additionally, we created another feature set by dividing the room into five sectors as shown in Figure 3 on the right. The borders of the five sectors correspond to different activity regions. Three regions are located at the three work places; the two other regions are only active when people walk around. Thus, our motion detector is a little more sophisticated than the usually used motion detectors because it distinguishes between different sectors, except where we employed the feature set of size 1.

We constructed a two-dimensional feature vector representing the time of the day by taking the hour and distinguishing between am and pm.

Results

This section presents the results obtained after conducting the experiment. The experiment lasted 41 (working) days. The data set consists of 1349 self-reports.

[1] We would like to thank B. Schiele and N. Kern for sharing their audio preprocessing code with us.

Data Overview (Descriptive Statistics)

Motion Detection

The motion detection data shows patterns as depicted in Figure 4. The usual location of the subject can easily be identified as the bright area. There are other lighter regions near the door and around the second work place. The right picture of Figure 4 additionally shows the sectors of the five features we have chosen. The borders of the particular sections overlap with the motion pattern.

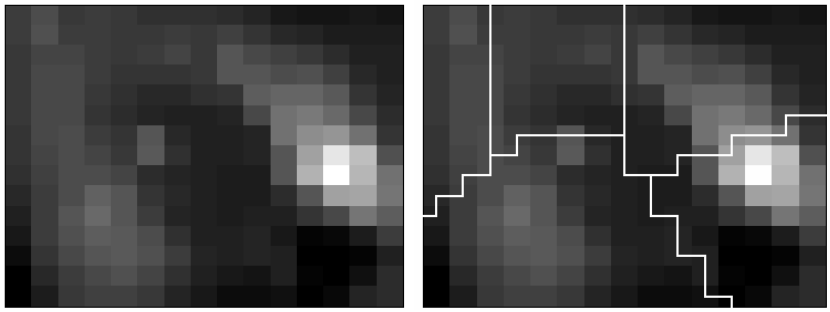

Fig. 4. Histogram of the movements recorded by the video camera. On the right the same picture as on the left but with the sectors representing the employed features.

Presence

Figure 5 shows the overall presence of the subject illustrating that the subject is in his/her office about 45.1% of the time. The histogram on the right graphs the presence depending on the time of the day. The lunch break manifests itself as a dip at noon. The (average) presence decreases at both ends of the day (note that the distinctive decrease at 8am and 6pm are mainly due to the partial recording).

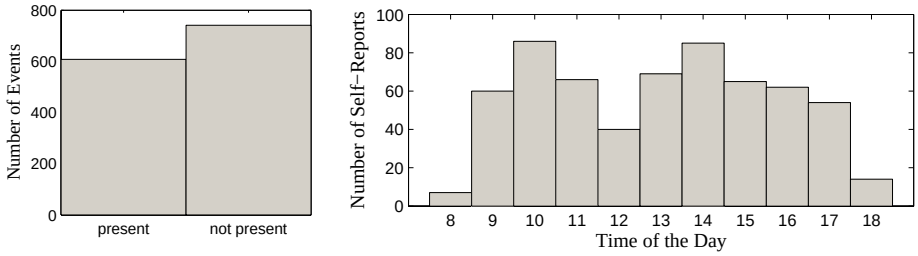

Fig. 5. The histogram on the left shows the overall presence of the subject in the office; the histogram on the right shows the presence vs. the time of the day.

Interruptability

When present, the subject self-reported his/her degree of interruptability on a scale from 1 “easily interruptible” to 5 “not at all interruptible”. Figure 6 shows the

distribution of the interruptability data. Class 2 “quite interruptable” is dominant with a prior probability of 29.3% followed by class 5 with 25.2%.

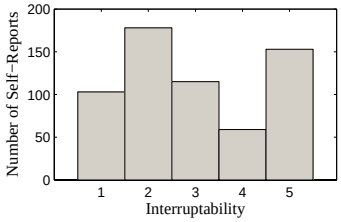

Fig. 6. Interruptability self-reports on a scale of five classes.

Prediction Quality

In this section, we present the prediction results of the subject’s self-reports from the sensor streams. First, we explain the prediction methods followed by the results.

Applied Classification Methods

We used the Weka 3 machine learning software package [20] to predict the subject’s self-reports. For all classification tasks we tested the data with two standard learning algorithms: naïve Bayes and the “J48” decision tree learner. We preprocessed the data by normalizing and discretizing it with the standard Weka algorithms for better predictions. For the predictions, we took data up to 5 minutes prior to the event into account. We incorporated this information by an additional processing of the data by averaging the data (with equal weight) for each self-report. The depth of this averaging defines how much of the information about the past is incorporated. For each original feature the resulting new feature vector then contains the mean and standard deviation.[2] All results reported below are based on a 10 fold cross-validation.

Presence

Both graphs in Figure 7 show the prediction accuracy versus past time. The graph on the left shows the prediction from motion detection evaluating all six possible feature-combinations. The largest feature set (the most finely grained with 256 rectangles per frame) turns out to be the most predictive. The graph on the right, compares the best motion detector prediction with audio. Both audio and motion detection show a distinct maximum at about 20 seconds. Motion detection reaches an accuracy of 96% at 20 seconds using the J48 decision tree classifier outperforming audio that reaches 89.9% using naïve Bayes. By taking only the time of the day into consideration to infer the presence we reach an accuracy of 60.3% which is still better than the prior annotation distribution of 54.9% (see Fig. 8 for the detailed confusion matrices). We combined the three classifications by meta-classifiers on their class prediction

[2] We also tried Markov chains and hidden Markov models. However, they were outperformed by our coarse approach.

probabilities but the results were not better than the prediction from motion detection. Thus, audio and the time of the day do not contribute any new information to achieve better accuracies but might contribute to higher robustness.

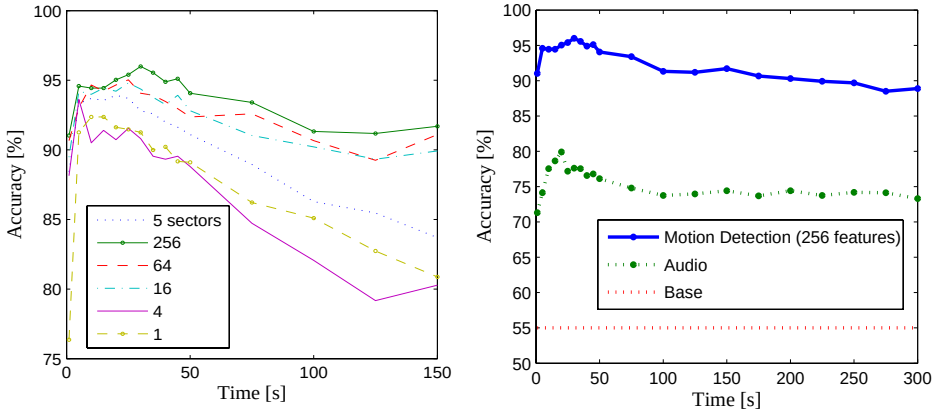

Fig. 7. Accuracies of presence prediction dependent on the past time. On the left, we see the motion detection with all six features for feature selection, on the right, the presence prediction from audio data compared with the prediction of from motion detection.

Presence, audio

		Prediction	
		1	2
Self-Report	1	582	159
	2	98	510
Accuracy: 80.9% Base: 54.9%			

Presence, motion

		Prediction	
		1	2
Self-Report	1	719	22
	2	32	576
Accuracy: 96.0% Base: 54.9%			

Presence, time of day

		Prediction	
		1	2
Self-Report	1	570	171
	2	365	243
Accuracy: 60.3% Base: 54.9%			

Fig. 8. Confusion matrices for presence prediction based on 20 seconds of past data. (Class 1 corresponds to “not present” and class 2 for “present”).

Interruptability

We have two data sets to infer the degree of the subject’s interruptability. This prediction task is a 5-class classification prediction with a base of 29.3%. Figure 9 on the left shows that the 5-sector feature of the motion detector is the most predictive. Figure 9 on the right shows the prediction accuracies of audio and motion detection vs. the time into the past. Both audio and motion detection show very good results with a maximum at 150s. Motion detection is superior (41.6%) to audio (40%) and, furthermore, motion detection seems to be more robust to variations on the time dimension. Predicting the interruptability from the time of the day using naïve Bayes results in an accuracy of 35.9%. Combining audio and motion detection by a naïve Bayes meta-classifier results in a remarkably better prediction result (maximum at 150s: 44.6%) indicating that both sensor inputs provide partly independent information. Combining all three information sources (audio, motion detection, and

time of the day) results in an even better result with a maximum accuracy of 45.4% at 150s. Furthermore, the combination of the three sources results in a much more robust result in terms of time dependency. Figure 10 shows the performance of the different calculations on the basis of confusion matrices.

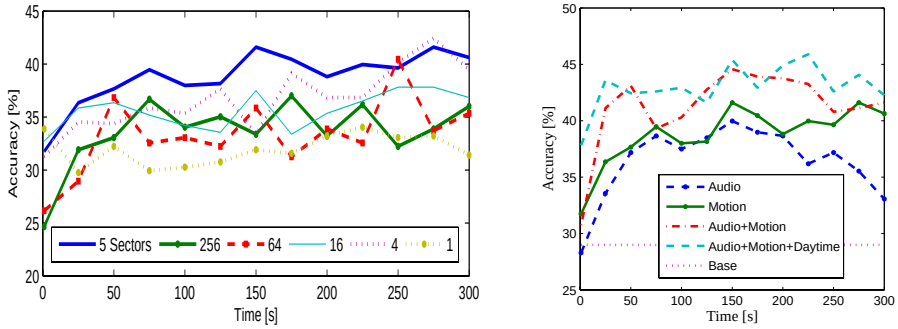

Fig. 9. 5-class interruptability prediction accuracies vs. time to past. On the left, there are the motion detection predictions from all feature setups. The figure on the right shows the audio and motion detection predictions and the combinations.

Interruptability, audio

Self-Report	Model prediction (naïve Bayes) 1	2	3	4	5
1	12	51	11	5	24
2	13	106	17	8	34
3	8	65	14	5	23
4	3	15	2	6	33
5	8	25	8	7	105

Accuracy: 40.0% (at 150s)
Base: 29.3%

Interruptability, motion detection

Self-Report	Model prediction (J48) 1	2	3	4	5
1	27	45	6	0	25
2	22	125	3	3	25
3	15	75	4	1	20
4	9	25	0	1	24
5	12	33	5	7	96

Accuracy: 41.6% (at 150s)
Base: 29.3%

Interruptability, time of the day

Self-Report	Model prediction (naïve Bayes) 1	2	3	4	5
1	1	40	1	0	61
2	5	104	4	0	65
3	4	46	0	0	65
4	1	19	0	0	39
5	4	36	0	0	113

Accuracy: 35.9%
Base: 29.3%

Interruptability, all combined

Self-Report	Model prediction (naïve Bayes) 1	2	3	4	5
1	35	24	18	8	18
2	15	103	38	4	18
3	7	53	34	3	18
4	6	11	9	2	31
5	8	11	22	10	102

Accuracy: 45.4% (at 150s)
Base: 29.3%

Fig. 10. Confusion matrices for the 5-class interruptability detection on the “beep” data set.

Future Work

The major drawback of this study is the experimental setup restricted to only one single subject. To strengthen the external validity of the experiment we intend to conduct this experiment with a broad range of different people.

We plan to replace the camera simulating a motion detector by one or several (for different sectors) daylight and infrared motion detectors for even stronger prediction power.

Finally, we plan to apply our approach to other (non-office based) areas.

Discussion and Conclusions

In this study we successfully introduced motion detection to augment context-awareness in office-like setups.

In order to reach our overall goals we found the following outcomes. We can predict whether a person is present in the office or not, based on motion detection, audio, and daytime data, where motion detection clearly outperforms the others. We showed that we can predict the person's degree of interruptability from these three information sources, where motion detection again turned out to be the most reliable source.

References

1. Cutrell, E.B., Czerwinski, M. and Horvitz, E., Effects of instant messaging interruptions on computing tasks. in *CHI'2000 Conference on Human Factors in Computing Systems*, (The Hague, The Netherlands, 2000), ACM-Press, Pages: 99 - 100.
2. Feldman-Barrett, L. and Barrett, D.J. Computerized experience-sampling: How technology facilitates the study of conscious experience. *Social Science Computer Review, 19*. 175-185.
3. Fogarty, J., Hudson, S.E. and Lai, J., Examining the Robustness of Sensor-Based Statistical Models of Human Interruptibility. in *SIGCHI Conference on Human Factors in Computing Systems (CHI 2004)*, (2004).
4. Gillie, T. and Broadbent, D. What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research, 50*. 243-250.
5. Hess, S.M. and Detweiler, M., Training to Reduce the Disruptive Effects of Interruptions. in *HFES 38th Annual Meeting*, (Nashville, TN, 1994), 1173-1177.
6. Horvitz, E. and Apacible, J., Learning and reasoning about interruption. in *5th international conference on Multimodal interfaces*, (Vancouver, British Columbia, Canada, 2003).
7. Horvitz, E., Koch, P., Kadie, C.M. and Jacobs, A., Coordinate: Probabilistic Forecasting of Presence and Availability. in *Eighteenth Conference on Uncertainty and Artificial Intelligence (UAI '02)*, (Edmonton, Canada, 2002), 224-233.
8. Hudson, J.M., Christensen, J., Kellogg, W.A. and Erickson, T., "I'd be overwhelmed, but it's just one more thing to do:" Availability and interruption in research

management. in *Human Factors in Computing Systems (CHI 2002)*, (Minneapolis, Minnesota, 2002), ACM-Press, 97-104.

9. Hudson, S.E., Fogarty, J., Atkeson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C. and Yang, J., Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. in *SIGCHI Conference on Human Factors in Computing Systems (CHI 2003)*, (Fort Lauderdale, Florida, 2003), ACM-Press.
10. Kern, N. and Schiele, B., Context-Aware Notification for Wearable Computing. in *International Symposium on Wearable Computing (ISWC'03)*, (White Plains, NY, 2003), IEEE Computer Society.
11. Markus, M.L. Finding A Happy Medium: Explaining the Negative Effects of Electronic Communication on Social Life at Work. *ACM Transactions on Information Systems*, *12* (2). pp. 119-149.
12. McFarlane, D.C., Coordinating the Interruption of People in Human-Computer Interaction. in *Human-Computer Interaction - INTERACT'99*, (1999), Published by IOS Press, Inc., IFIP TC 13, 295-303.
13. McFarlane, D.C. Interruption of People in Human-Computer Interaction *School of Engineering and Applied Science*, George Washington University, Washington, DC, 1998.
14. McFarlane, D.C. Interruption of People in Human-Computer Interaction: A General Unifying Definition of Human Interruption and Taxonomy, Naval Research Laboratory, Washington, DC, 1997.
15. Mintzberg, H. *The Nature of Managerial Work*. Harper & Row, New York, 1973.
16. O'Conaill, B. and Frohlich, D., Timespace in the workplace: Dealing with interruptions. in *Human Factors in Computing Systems (CHI'95)*, (Denver, Colorado, 1995), ACM-Press, 262-263.
17. Sawhney, N. and Schmandt, C., Nomadic Radio: Scaleable and Contextual Notification for Wearable Audio Messaging. in *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 99)*, (Pittsburgh, Pennsylvania, 1999), ACM-Press.
18. Siewiorek, D., Smailagic, A., Furukawa, J., Krause, A., Moraveji, N., Reiger, K., Shaffer, J. and Wong, F., SenSay: A Context-Aware Mobile Phone. in *International Symposium on Wearable Computers - Poster Session*, (2003).
19. Syrjälä, J. Context Classification using Audio Data for Wearable Computer *Department of Informatics*, Swiss Federal Institute of Technology (ETH), Zürich, 2003.
20. Witten, I.H. and Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufman Publishers, San Francisco, 2000.