

A Model for Human Interruptability: Experimental Evaluation and Automatic Estimation from Wearable Sensors

Nicky Kern^{*†}, Stavros Antifakos^{*}, Bernt Schiele^{*†}, Adrian Schwaninger^{‡§}

^{*}Perceptual Computing and Computer Vision, ETH Zurich, Switzerland

[†]Multimodal Interactive Systems, Darmstadt University of Technology, Germany

[‡]Max Planck Institute for Biological Cybernetics, Tübingen, Germany

[§]Department of Psychology, University Zurich, Switzerland

Abstract

For the estimation of user interruptability in wearable and mobile settings, we propose in [8] to distinguish between the users' personal and social interruptability. In this paper, we verify this thesis with a user study on 24 subjects. Results show that there is a significant difference between social and personal interruptability. Further, we present a novel approach to estimate the social and personal interruptability of a user from wearable sensors. It is scalable for a large number of sensors, contexts, and situations and allows for online adaptation during run-time. We have developed a wearable platform, that allows to record and process the data from a microphone, 12 body-worn 3D acceleration sensors, and a location estimation. We have evaluated the approach on three different data sets, with a maximal length of two days.

1 Introduction

Today's computer and Ubicomp systems are often annoying since they interrupt users at the most inappropriate moments. Interruptions may come from a wide variety of sources, such as email notification, incoming cell-phone calls or simply a system message that the network connection has failed. A major problem in our opinion is that those systems do not model and recognize the user's current context and situation in order to estimate his interruptability. This is of special importance in today's and future mobile systems, where such interruptions may occur in any situation and from an increasing number of sources. In this paper we explore the example application of interruptability estimation for a wearable notification system.

There is always two sides to a notification: a cost and a utility. While the utility mainly depends on message type, sender, content, and importance, the cost mainly depends on

the interruptability of the user. In this paper we concentrate on estimating the user's interruptability, since the estimation of the message utility is a research topic on its own.

Previous related work mainly concentrates itself on interruptability estimation in office settings [5, 2, 4] or highly specialized settings, such as Military Command Control [11] or Space Shuttle Monitoring [3]. They often model the interruptability as a uni-dimensional variable, eg. 'interruptible' vs. 'non-interruptible'. Considering mobile and wearable settings, where an interruption may occur at any time in any situation, we previously proposed [8] to distinguish between the personal and social interruptability of the user. We define the *personal interruptability* to be the interruptability of the user. The *social interruptability* indicates the interruptability of the user's environment. This differentiation is introduced for context-aware notification on a wearable computer, where notifications such as a loud beep not only interrupt the user but also the social environment of that user. We depict the interruptability using the two-dimensional *interruptability space* (see Figure 1).

This paper has three principal contributions: firstly, we conducted a user study to lend support to the necessity of distinguishing social and personal interruptability (see Section 2). Secondly, we propose a novel, scalable approach for estimating the user's interruptability (see Section 3), that overcomes several restrictions our previous algorithm had. For one, we employ an unsupervised learning mechanism to extract low-level context from acceleration, audio and a location sensors. This unsupervised learning can be done incrementally – and therefore online – making the system scalable and adaptive to novel situations. We compare the novel approach against the previous one and evaluate it on two sets of data of 3.5h and 2 days length respectively (see Section 5). As the third contribution, we have developed a hardware platform that allows to record, process, and annotate data from 12 3D acceleration sensors, a microphone and a location sensor in online and in realtime (Section 4).

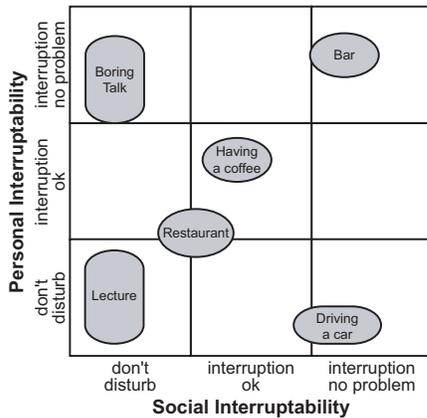


Figure 1. Social and Personal Interruptibility for Various Situations

2 Human Interruptibility: An experimental Study

The validity of the model presented in the introduction was investigated in an experiment using a set of videos of 47 prototypical situations. These situations were selected to provide a diverse sample of everyday situations. For each 5-second situation movie clip, participants rated how much it would disturb them to receive a phone call if the phone was set to ring or set to vibrate (assessment of personal interruptibility). In addition, participants had to indicate how much they believe it would bother the people in the environment, if their phone rang audibly (social interruptibility). Note that the modality (ring vs. vibration) is very important for testing the degree of independence of these two dimensions. If the phone rings audibly, it is reasonable to assume that personal and social interruptibility would be relatively well correlated, because the ring tone does of course affect the social environment as well. Testing the degree of independence of these two dimensions would be more valid if a modality is used for assessing personal interruptibility that does not alert the social environment automatically. This is fulfilled when the phone is set to vibrate. Therefore, the interesting correlation is between ratings of personal interruptibility when the phone is set to vibrate and social interruptibility when the phone is set to ring.

Method and Procedure. 24 students from either the computer science department of ETH Zurich or the psychology department of the University of Zurich participated in the study. Half of the participants were female. The mean age of participants was 24.6 years (stdev. 4.1 years). 23 of the 24 participants reported to own a mobile phone. The experiment was conducted using a personal computer running Windows XP with the screen resolution set at 1280x1024 on a TFT screen. A Java program was written to display



Figure 2. A screenshot of the program used for the Experiment. Below the video there are 3 sliders to assess the interruptibility, and a “next” button to start the next video.

the videos and to accept the users answers (see Figure 2). The Program covered the full computer screen, except the windows task bar. The videos were recorded using a digital video camera. All scenes always depicted the same actor during routine activities. From the video recording we extracted a total of 94 different 5-second videos. These consisted of 2 different instances of 47 different prototypical situations. We chose the situations to best possibly sample the interruptibility space to our own belief. The situations varied between: looking at shop windows, sitting in a nice restaurant, listening to an interesting lecture, and studying in the universities library, amongst others. For each video the participants had to provide a rating for each of the three following questions using graphic rating scales (on a scale from *little* – value 0 to *much* – value 1):

- How much does it disturb **you**, if you receive a phone call in this situation? (Mobile phone rings audibly)
- How much does it disturb **you**, if you receive a phone call in this situation? (Mobile phone vibrates, you notice it)
- How much does it disturb **your environment**, if you receive a phone call in this situation? (Mobile phone rings audibly)

The first two questions assess the participants personal interruptibility, depending on the means of notification. The third question assesses the participants social interruptibility for one specific modality. The independent variable *situation* was randomized. 47 different situations and 2 repetitions per situation were used. This results in 94 trials per participant, which took about half an hour to complete.

Results. For each participant we calculated the correlation between ratings of personal interruptibility when the

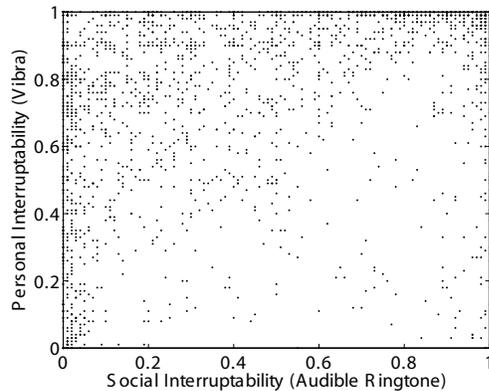


Figure 3. Modalities personal-vibra vs. social-ring. Each point represents the annotation of one person for one situation. Axes run from 0 – not interruptable to 1 – highly interruptable

phone was set to vibrate and ratings of social interruptability when the phone was set to ring (individual correlation values). On average, a correlation of $r = .40$ was found (stdev = $.28$), which is consistent with the assumption of two relatively independent dimensions for personal vs. social interruptability. As depicted in the scatter plot in Figure 3, the different situations used in the experiment did sample the interruptability space relatively well.

These results were different, when participants had to rate personal interruptability assuming that the phone was set to ring. In this case, personal and social interruptability were much higher correlated (mean $r = .79$, stdev = $.19$). Apparently, many participants assume that in this case the environment is disturbed to a similar degree as they feel being disturbed themselves. The scatter plot depicted in Figure 4 shows the associated changes in the interruptability space. As indicated by the standard deviations of individual correlation coefficients, substantial differences between people were observed in both cases. It would therefore be interesting for future studies to investigate which personal-ity variables or even other variables such as gender can predict the degree of independence of personal and social interruptability. Such results could also provide important information for the estimation of settings of individually adapted notification systems. For example, preliminary analysis of the present study suggest that the correlation between personal and social interruptability might be generally higher for men than for women ($r = .84$ vs. $r = .73$ for same modality and $r = .44$ vs. $r = .37$ for different modality).

Discussion. The results presented above show that there is a significant difference between personal and social interruptability. Especially when regarding modalities that make it possible to choose whether the whole environment or just oneself is interrupted. Interestingly, the study shows that

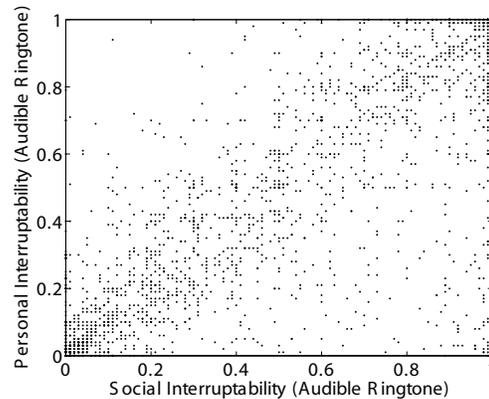


Figure 4. Modalities personal-ring vs. social-ring. Each point represents the annotation of one person for one situation.

there are differences amongst people in the way they assess personal and social interruptability. This fact speaks in favor of having a personal interruptability estimation system that adapts to the user's preferences.

3 Interruptability Estimation from Sensors

In [8] we have introduced an algorithm for estimating the users' personal and social interruptability from data obtained from wearable sensors. It combines low-level context information using tendencies to obtain an estimate of the interruptability. The low-level context is based on one acceleration, an audio, and a location sensor. Every sensor is classified into about 5 classes, e.g. 'walking' for the acceleration sensor. The tendencies indicate the likelihood of the interruptability for every position in the space for every sensor. The tendencies are weighted using the recognition score of the low-level contexts and added together. The maximum within the interruptability space then indicates the users' current interruptability.

This approach has three principal limitations: firstly the low-level contexts are recognized using classifiers that are trained in a supervised manner. This requires labor-intensive acquisition of labeled training data, and thus limits the extensibility towards new contexts. Furthermore, it is not possible to recognize contexts that require more than a single sensor modality. Secondly, the tendencies are hand-crafted – a labor-intensive and tedious work that has to be repeated every time a new context or sensor is added. Thirdly, the tendencies are modeled as uni-modal Gaussians, an assumption which is not necessarily true.

We present here two novelties to this approach: by modeling the tendencies in a non-parametric way we avoid the uni-modality restriction and also allow for direct learning of the tendencies. By clustering the sensor data, we also ob-

tain the low-level contexts automatically. Combining both measures, the novel approach learns low-level contexts fully unsupervised, and thus scales for a large number of contexts and situations, and also allows for online learning and adaption of the system during runtime. The new approach is compared to our previous algorithm using the same data set and error measure.

3.1 Learning Multi-Modal Tendencies

We represent the tendencies in a non-parametric way in order to overcome the unimodality limitation of the Gaussian tendencies. We therefore divide the interruptability space in bins $b(\hat{x}, \hat{y})$ at a given resolution, where (\hat{x}, \hat{y}) refers to the (two-dimensional) bin number. A non-parametric tendency $\hat{T}(\hat{x}, \hat{y}, s)$ maps a constant value to every bin. With a sufficiently fine resolution of the grid, we can represent arbitrarily complex tendencies.

Using such tendencies, we can calculate the sensor estimate for every bin, over all sensors (with recognition score $l(s, t)$):

$$\widehat{SE}(\hat{x}, \hat{y}, t) = \sum_s \hat{T}(\hat{x}, \hat{y}, s) \cdot l(s, t)$$

The actual interruptability can then be found by selecting the bin with the highest value from the sensor estimate.

Learning. Given a set of low-level context readings and the corresponding ground truth, the tendencies can be estimated using standard optimization techniques. The ground truth is given as a single point (x, y) in the interruptability space. In order to allow for small variations, we blur the ground truth to the bin (\hat{x}, \hat{y}) and its adjacent bin using a Gaussian function with variance σ . This gives us for every time step t a ground truth grid $g(\hat{x}, \hat{y}, t)$.

Given the ground truth grid and the sensor readings, we obtain a linear equation for every bin and time step: $\sum_s \hat{T}(\hat{x}, \hat{y}, s) \cdot l(s, t) = g(\hat{x}, \hat{y}, t)$. Using the equations of all time steps, this represents an over-determined system of linear equations that can be solved using least-squares.

3.2 Finding Low-Level Context Automatically

In our previous approach, low-level contexts were extracted using classifiers that were trained in a supervised manner from the data of a single sensor each. To overcome the necessity of labeled training data, we extract them automatically by clustering the sensor data.

We concatenate the feature vectors of the different sensors to form a long d -dimensional vector. We then perform a k -means clustering on the concatenated feature vectors, which returns k cluster centers in the d -dimensional space. In order to keep as much information as possible, the vector of k distances to the k cluster centers is used as new feature

Variant	Social Inter.	Personal Inter.
Hand-crafted tendencies Low-level contexts from supervised training	86.0 %	96.2%
Learned Tendencies (6x6 bins) 5-fold cross-validated Low-level contexts from supervised training	92.72 %	97.75 %
Learned Tendencies (6x6 bins) 5-fold cross-validated Low-level contexts automatically found (20 clusters)	91.0 %	97.4 %

Table 1. Comparison of Recognition Scores for 3 Algorithmic Variants

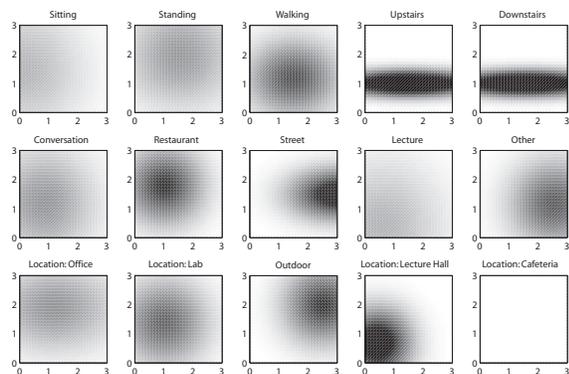
vector. Again, this model allows incremental extension of the model during run-time of the system, simply by reclustering the data to form new low-level contexts.

3.3 Evaluation of the Novel Approach

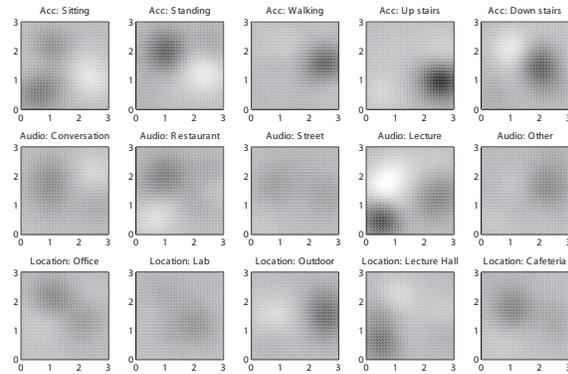
We have compared both the original and novel approaches against each other using the original data set from [8]. This data set contains a 37 min. recording from a single biaxial accelerometer, a microphone, and a location estimate from the closest WLAN access point. It contains a walk on the street, (part of) a lecture, a visit to a restaurant and several conversations. The performance numbers given in table 1 indicate the percentage of the time where the result is in the right bin or one of its adjacent bins (using a 6x6 grid), a comparable error measure to the previous one.

As a first observation, we see that the assumption of unimodality does not hold. Figure 5(b) shows the tendencies that were automatically learned from the low-level contexts used previously. Even more, the tendencies are not all non-negative, i.e. there are parts, that ‘punish’ certain areas. Thus, the recognition scores are with 92.72% and 97.75% for social and personal interruptability higher with only 86% and 96.2% respectively (see Table 1).

Table 1 summarizes the recognition scores for the original system and our novel approach on the original data set. Comparing the automatically found low-level contexts (third row) and the manually trained ones (second row), we find that the former perform only marginally worse (91%/97.4% instead of 92.72%/97.75% for social and personal interruptability, respectively). However, they still perform better than the original results. Given the fact, that these contexts are found in a completely unsupervised manner and the approach thus scales to a large number of contexts, this small performance penalty is acceptable.



(a) Hand-crafted Gaussian Tendencies (white denotes 0, black 0.2).



(b) Automatically Learned Tendencies on a 30x30 Grid (white denotes -0.22, black 0.4.).

Figure 5. Hand-crafted and automatically learned tendencies

4 Platform

As the basis for our further research, we have developed a hardware and software platform that allows both for recording of a large number of data streams, for their on-line processing, and online annotation (see Figure 6(a)). It allows for entire days of operation: not only does it supply sufficient battery power, but it is also easy enough to put on and comfortable and unobtrusive to wear. Using a wrist-worn display it permits to annotate data online without introducing a bias into the data.

Main Unit. The main unit of our platform is a CharmIT Pro wearable computer, a Linux system with a hot-pluggable battery-driven power supply. It provides many serial and USB ports for attaching sensors, but little built-in IO, such as keyboard or display, that would not be useful in a backpack. We use a MicroOptical HMD and a Twiddler for easy access (without having to take the computer out of its bag).

We use the CharmITPro's power supply as the central power supply for our entire platform. Our experiments show that the systems runs continuously for 6 hours on a single set of batteries. Since they are hot-pluggable, the system can well be used for recordings over long periods of time.

The entire system has been packed in a backpack, modified to accommodate all components and for easy wearing and putting-on. The batteries are stored on the outside of the backpack, and cable channels allow to place cables and sensors. Holes in the outer fabric allow for ventilation.

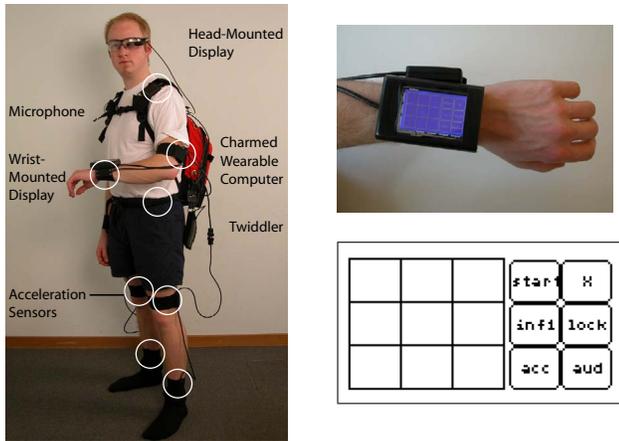
Directly attached to the main unit are a microphone with a USB audio device and a wireless LAN adapter. The microphone is attached to the chest strap of the backpack and records both the voice of the user and environmental sounds. The wireless LAN adapter both provides network access and a rough location estimate.

Acceleration Sensors. The platform supports 12 3D acceleration sensors, based on the work of Kern et al. [9]. They are attached to both shoulders, elbows, wrists, hips, knees, and ankles. Figure 6(a) shows where the nodes are placed. The 12 sensors are grouped into two groups of six sensors each. Each group of acceleration sensors is attached to one Smart-It [1]. The shoulder and hip sensors are kept in the cable channels that are sown to the backpack. The sensors for arm and wrist, and knee and ankle respectively, are wired through a single pluggable cable to the Smart-It. This allows to wear the sensors underneath clothing and easily connect them to the backpack worn above the clothing.

Wrist-Worn Display for Data Annotation. We use a small LCD panel with built-in touch panel, worn on the wrist of the user, as annotation device. It provides quick and easy interaction-at-a-glance, comparable to a wrist watch. This is essential for efficient and sensible data annotation, especially for long recordings.

A simple and quick solution to annotating data is, especially for long recordings, crucial for two reasons: firstly, a clumsy annotation device can introduce considerable bias in the data (Kern et al. [9] report that annotating data using a Compaq iPAQ introduced considerable bias in their data). Secondly, for long recordings it becomes impossible to get continuous annotation of the data, simply because users cannot concentrate long enough. It is therefore necessary to resort to experience sampling techniques [6, 4, 5, 2]. These require simple and quick interaction with the user in order to be the least obtrusive possible.

The display we use provides a resolution of 64x128 pixels and a 5x3 field touch panel. It is attached to the main unit via a standard serial line. Further, it is programmed using simple escape sequences. This allows to write standard C/C++ programs to control the interaction over the display. It is mounted inside a small box that can be strapped to the user's wrist. Figure 6(b) shows the display attached to the user's wrist.



(a) The entire platform as worn by the user.

(b) Wrist-Mounted Display for Data Annotation. Display above, screen-shot of application below.

Figure 6. The Hardware Platform

Our annotation application allows the user to annotate his current interruptability and also to control the recording. Figure 6(b) shows a screen shot of the application. The interruptability space on the left hand side can be used to annotate the user’s interruptability. The buttons on the right hand side allow to start and stop the recording (using the ‘start’ button) and to clear all annotations (using the ‘X’ button).

Software. The software for our system has to cope with multiple streams of data, which can be used in very different compositions for different usage modes. Each sensor produces a stream of data. These streams have to be handled differently, depending on the current usage mode of the system. For mere recording, it is sufficient to write them into a file. For online recognition, features have to be extracted and classified. For debugging purposes it is desirable, that these streams can not only come directly from the sensors, but also be read from files.

To make the system easily adaptable to such different usage modes, we implemented a streaming framework that can easily be reconfigured to adapt to new requirements. Each processing step, such as extracting features from a window of audio data, is encapsulated in a *building block*. These are connected among each other at run time, using a concept similar to UNIX pipes. Central units ensure proper start/stop of the data streams in-between the different building blocks.

5 Experiments

We have validated our approach experimentally on two data sets. The first consists of two recordings of continuously annotated data with a combined length of 3.5 hours. The second consists of two days of 8 hours in which we have 54 samples of the user’s interruptability.

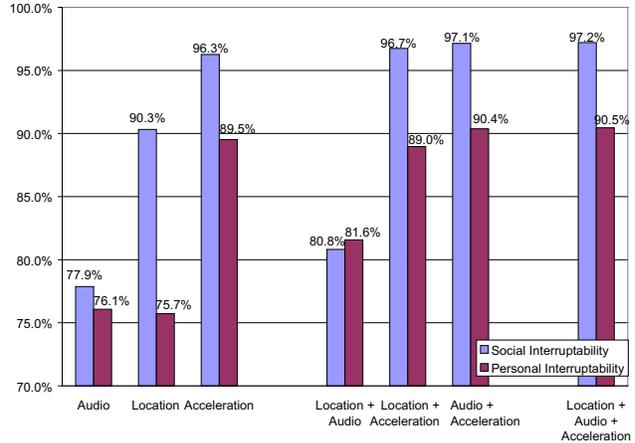


Figure 7. Comparison of sensor combinations (‘acceleration’ implies all 12 sensors)

Experimental Setup. To show the performance of our system we have recorded a set of 3.5h of data in two parts of 1h 23min and 2h 06 min each. The data was continuously annotated (online) using the wrist-mounted display. The first part contains a walk from ETH’s CS to its EE department (10min) and back (5min) with a detour via the library (10min) and the student restaurant (10min). The user had several discussions on the way to and at the EE department. The second part contains the author working in his office (10min), having a short bike ride (20min) and then going to dinner with 3 colleagues in a restaurant (1h in the restaurant, 20min for the way back and forth).

The audio features are 10 cepstral coefficients, computed over 30ms windows and averaged over 1 second. The average and variance over one second were used as the features for every acceleration sensor. The closest wireless LAN access point was sampled every second as location estimate. By concatenation of these three vectors we obtain a feature vector every second. All experiments use a resolution of the interruptability space of 6x6 and blur the ground truth with a Gaussian of variance 0.5. All results are obtained using 50 automatically extracted low-level contexts. All results are 5-fold cross-validated. We use the same error measure as in Section 3.3.

General Results. As a first evaluation we have computed the recognition results for every single sensor and all possible combinations (counting all acceleration sensors as a single one). Figure 7 summarizes these results.

The first important result is, that the algorithm performs similarly on this data set as on the one used in our previous work (Section 3) when using all sensors (see the right-most set of bars in Figure 7). The recognition score of the social interruptability is 4.5% higher (97.2% instead of 92.7%), while the score for the personal interruptability has dropped by about the same rate (90.5% instead of 97.7%). Even

though the score of the personal interruptability is lower, these are very good results, considering that the data set is considerably longer (3.5h instead of 37 min) and more complex.

In general the recognition of the personal interruptability seems harder than that of the social interruptability. We believe that the personal interruptability changes more quickly than the social interruptability and is thus harder to recognize. Another issue is that a change in the personal interruptability is not necessarily reflected in the output of sensors such as audio, acceleration, and location. Introducing sensors that measure the state of the user, such as galvanic skin resistance, heart rate, etc., could alleviate this problem.

When comparing the performance of single sensors each among each other, acceleration clearly performs best, with 96.3% and 89.5% for the social and personal interruptability, respectively. Audio gives a relatively low performance of only 77.9%/76.1% for social and personal interruptability. Location performs well with 90.3% for the social interruptability, but has the lowest score among all three for the personal interruptability (75.7%).

Considering combinations of sensors, obviously the combinations *acceleration+location* and *acceleration+audio* perform best, *acceleration+audio* giving slightly better results (97.1%/90.4% instead of 96.7%/89%) The combination *audio+acceleration* seems to level out the performance of the two single sensors: together, they perform slightly better than each of the single sensors alone (80.8%/81.6%). However, the good performance of the location sensor for the social interruptability is equally leveled out. This could be because audio has more weight due to the larger number of features (10 instead of 1).

The combination of all three sensors still performs best. However, the results are only slightly better than the combination *acceleration+audio*. Given that the position estimation we get from the WLAN access point is quite coarse, it would be interesting if better results can be achieved with a finer positioning system.

Reducing the Number of Acceleration Sensors. As it will not always be possible to place 12 acceleration sensors on the user, we started to investigate which subsets of sensors are useful.

In a first attempt, we calculated the scores for the 4 6-sets of sensors upper body, lower body, right half, and left half of the body. Overall, the recognition results remain in the same order of magnitude as with all sensors. The personal interruptability is least affected. With 89.2%-89.9% the recognition score is only at most 1.3% smaller than for all sensors. The recognition scores for the social interruptability are very good with 92.3%-96.4% and still better than those for the personal interruptability. Interestingly, the right side of the body seems to be less informative than the left. We credit this to the fact that the user, being right-handed, per-

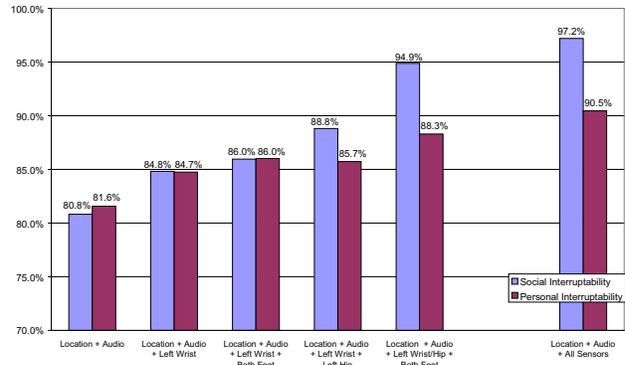


Figure 8. Recognition Scores for Audio, Location and Small Sets of Acceleration Sensors

forms more diverse actions with his right side and that thus the data has greater variance on the right side.

The second investigation is driven by the possibility to place the sensors (see Figure 8). We chose four locations: the left wrist, left hip and both shoes. The left wrist is the place where most people wear their watches where an acceleration sensor can easily be embedded [7]. The shoes are an ideal place to place self-powered acceleration sensor nodes [10]. The hip sensors can for example be embedded in a belt. The results for *audio+location* and *audio+location+all acceleration* are given for comparison.

Adding a single acceleration sensor on the wrist to audio and location sensors already increases the recognition score from 80.8%/81.6% to 84.8%/84.7% for social and personal interruptability respectively. Adding the sensors on the shoes or on the hip increases the score only by another 1.0%-4.4%.

By combining all three sensor positions (wrist, hip, and shoes) the recognition score increases considerably, especially for the social interruptability (to 95.9%, which is only 2.3% worse than the result for all sensors). The personal interruptability with a score of 88.3% also comes within 2.2% of the result for all sensors. This is the best 4-set we found.

Experience Sampling. For evaluation of the system on real-world data, we have recorded a large data set of two entire days (of 8 hours each) of acceleration, audio, and location data. We have sampled the users interruptability at regular intervals. During the two days of data, we have annotated 2.5 hours continuously, which are used for training the tendencies. We have interrupted the user 2-3 times per hour using an audio alarm, and have asked him to annotate his interruptability for that specific moment. We have thus obtained 54 interruptability samples.

Out of the 54 interruptability samples, the social interruptability was correctly classified 53 times (corresponding to 98.1% of the cases). For the personal interruptability however, only 44 samples, or 81.5%, were correctly classi-

fied. 9 out of these 10 samples are all from the same activity (namely, ‘having a coffee’) and in fact constitute all occurrences of that activity in the 54 samples. This suggests that the low performance for the personal interruptability is due to an under-representation of ‘having a coffee’ in the training set. Adapting the model online by relearning low-level contexts and tendencies could eliminate that problem.

6 Conclusion and Outlook

Estimating the interruptability of users is an important problem, especially for future mobile and wearable applications. We have introduced a model of interruptability in which we distinguish the users’ personal and social interruptability. We show that this distinction makes sense and that this model can also be used for automatic recognition from wearable sensor data.

As the first contribution, we have verified in a user study, that the distinction between social and personal interruptability is necessary. 24 subjects were asked to rate the social (for a ringing phone only) and personal interruptability (for a ringing and a vibrating phone separately) for 47 different situations (twice per situation). The results show, especially for the vibra condition, that indeed there is a significant difference between the social and personal interruptability. Furthermore, the study has shown that people are quite different in the way they assess the interruptability. Thus, future systems should be able to adapt to the user.

As second contribution, we introduce a scalable scheme for estimating the social and personal interruptability of the user from sensor data obtained from wearable sensors, which scales with the number of contexts and sensors, allowing for online adaption of the system.

The third contribution is a novel hardware platform that overcomes the limitations of our previous work, which only used 1 2D acceleration sensor. The novel platform now uses 12 3D acceleration sensors worn on the users body. A microphone worn on the users collar and a location estimate are also included.

We have evaluated the approach on three different data sets. The comparison of old and new approach shows that the novel approach performs better than the previous one (increasing the recognition rate from 86%/92.7% to 96.2%/97.75% for social and personal interruptability). The second data set is 3.5h long and continuously annotated. The results are very similar (97.2%/90.5%), even though the data is considerably more complex and varied. We have tried to identify smaller sets of acceleration sensors and found the set *audio+location+left wrist+hip+both feet* to perform nearly as well as all sensors together. Finally, we have recorded 2 entire days of data (of 8h each), in which we have sampled the user’s interruptability 2-3 times per hour. While the social interruptability is well recognized

(with 98.1% recognition score), the personal interruptability achieved only 81.5% recognition score. We believe this is due to insufficient training examples. This issue should be solvable by learning the model online from the experience samples.

References

- [1] M. Beigl, T. Zimmer, A. Krohn, C. Decker, and P. Robinson. Smart-its - communication and sensing technology for ubi-comp environments. technical report issn 1432-7864. Technical report, TeCo, University Karlsruhe, Germany, 2003/2.
- [2] J. Fogarty, S. E. Hudson, and J. Lai. Examining the robustness of sensor-based statistical models of human interruptability. In *Proc. ACM CHI*, 2004.
- [3] E. Horvitz and M. Barry. Display of information for time-critical decision making. In *Proc. 11th Conf. on Uncertainty in Artificial Intelligence*, pages 296–305. Morgan Kaufmann, 1995.
- [4] J. Hudson, J. Christensen, W. Kellogg, and T. Erickson. ‘i’d be overwhelmed, but it’s just one more thing to do’: Availability and interruption in research management. In *Proc. ACM CHI*. ACM Press, 2002.
- [5] S. Hudson, J. Fogarty, C. Atkeson, D. Avrahami, J. Forlizzi, S. Kiesler, J. Lee, and J. Yang. Predicting human interruptability with sensors: a wizard of oz feasibility study. In *Proc. ACM CHI*, pages 257–264. ACM Press, 2003.
- [6] S. Intille, E. M. Tapia, J. Rondoni, J. Beaudin, C. Kukla, S. Agarwal, L. Bao, and K. Larson. Tools for studying behaviour and technology in natural settings. In *Proc. UbiComp, LNCS*, volume 2864, pages 157–174, October 2003.
- [7] N. Kamijoh, T. Inoue, C. M. Olsen, M. Raghunath, and C. Narayanaswami. Energy trade-offs in the ibm wristwatch computer. In *Proc. ISWC*, pages 133–140, 2001.
- [8] N. Kern and B. Schiele. Context-aware notification for wearable computing. In *Proc. ISWC*, pages 223–230, White Plains, NY, USA, October 2003.
- [9] N. Kern, B. Schiele, and A. Schmidt. Multi-sensor activity context detection for wearable computing. In *Proc. EU-SAI, LNCS*, volume 2875, pages 220–232, Eindhoven, The Netherlands, November 2003.
- [10] J. Kymisis, C. Kendall, J. Paradiso, and N. Gershenfeld. Parasitic power harvesting in shoes. In *Proc. ISWC*, pages 132–139. Computer Society Press, October 1998.
- [11] R. W. Obermayer and W. A. Nugent. Human-computer interaction (hci) for alert warning and attention allocation systems for the multi-modal watchstation (mmws). In *Proceedings of SPIE*, volume 4126, 2000.