Context–Aware Notification for Wearable Computing

Nicky Kern and Bernt Schiele
Perceptual Computing and Computer Vision, ETH Zurich
{kern,schiele}@inf.ethz.ch

Abstract

In this paper we propose to use context information obtained from body–worn sensors to mediate notifications for a wearable computer. In particular we introduce a model which uses two axes, namely personal and social interruptability of the user in order to decide both whether or not to notify the user and to decide which notification modality to use. Rather than to model and recognize the complete context of the user we argue that personal and social interruptability can be derived directly from various sensors by the combination of tendencies. First experimental results show the feasibility of the approach using acceleration, audio, and location sensors.

1 Introduction

With the increasing number of wearable devices used by people in their everyday lives, there is an equally increasing number of applications that aim to grab the user's attention by various notifications. Be it arriving e–mails or telephone calls, upcoming meetings, changes in the stock market or navigation directions, the list of notifications on a wearable computer that can happen anywhere at any time in any situation is increasing. Clearly, there is a need to carefully handle and manage these increasing number of notifications in order to prevent wearable devices to become highly annoying. Importantly, management of notifications should take into account that the value of receiving a notification varies depending on the user's context.

Any notification has two sides for the user: on the one hand it has a value and on the other hand it comes at the cost of interrupting the user. It is well known that interruptions can decrease work performance considerably. In everyday life, the user's primary task is often unrelated to the wearable computer, such that interruptions can be highly annoying and may be even dangerous. For example while crossing a street, the user should not be distracted by a flashing head–mounted display notifying him of the arrival of an unimportant e–mail.

As pointed out before, the decision whether to notify or not depends on the user's context. In particular we differentiate five important factors of the current user context: the *importance* of the event, that is being notified, the *activity* of the user, the *social activity* (if the user is interacting with others and if so, in which way), the *social situation* ('in a restaurant', 'in a tram', etc.), and finally the *location*.

It is important to point out that none of these factors is sufficient alone to make the best possible decision. For example, a user in his office (a location) may well receive notifications if he is working alone (the activity of the user), but it is less appropriate to receive notifications while having a meeting with his supervisor in the same location (social activity).

The first factor of the five, i.e. the importance of the event, defines the value of the notification to the user. It is of quite different nature from the other four since it is mostly unrelated to the user's context. Since the automatic determination of the importance of events is a research topic on its own we assume that the importance is given for the purpose of this paper. Therefore this paper deals with the remaining four factors which are related to the cost of a notification event, i.e. the interruptiveness of that event.

There exist three principal ways to obtain those four factors. Firstly they can be supplied by the user. But since our goal is to reduce the load of the user, this option is clearly suboptimal and undesirable. Secondly, the factors could be inferred from usage patterns of the device. However it is hard to imagine that the four factors can be determined successfully in cases where the primary task of the user does not use the wearable device and is therefore not used. In our opinion the most promising possibility is the third, using sensors to capture and recognize information about those four factors automatically. Since the four factors span a wide space it is of course not possible to use one single sensor to perceive all of the factors. Hence we propose to use multiple, body–worn sensors to acquire the necessary context information.

In today's context-aware computing it is common to define or describe special situations or contexts which are directly related to the behavior of a 'context-aware' application. In the case of notification however this approach does not seem appropriate. Take again the example of the context 'working in your office'. Since the user might be alone or

Proceedings of the Seventh IEEE International Symposium on Wearable Computers (ISWC'03)
1530-0811/03 $ 17.00 © 2003 IEEE

might have a meeting with another person this context has a wide range of different notification requirements. Therefore, 'working in your office' does not seem to be the appropriate level of context abstraction for the management of notification. Further specifying the context would obviously help here but would require to specify many special cases for typically a large number of contexts.

In this paper we therefore do not aim to model and specify all potential contexts which might be encountered by the user. Rather we concentrate on a lower level of context abstraction which is more directly linked to the management of notifications. More specifically, we propose a notification system where the decision whether to notify or not is based on a set of low–level contexts, such as the 'user is having a conversation', without the intermediate step of recognizing the entire situation.

There are three principal contributions in this paper. Firstly we propose a model that allows to classify situations and notifications and to select the best notification modality (section 3). Secondly, a method is developed to combine multiple context sources in order to find the best notification (section 4). Finally we have investigated the acquisition of context information from audio (section 5) and acceleration sensors (section 6) and show the feasibility of the proposed methods with first experimental results (section 7). The last section (section 8) discusses the approach and future work.

2 Related Work

The importance of managing user interruption in HCI design is well–known. Until recently managing user attention and mediating notifications was mainly applied for highly specialized applications, such as Military Command Control [13] or Space Shuttle Monitoring [8], where the main task of the user is bound to the computer and requires the entire user attention. Cutrell et al. [4] have recently shown, that interruptions by instant messaging applications decrease human performance and have a negative effect on the memory required for resuming the original primary task.

Several people have tried to address notifications in a systematic way. McCrickard et al. [12] suggest to classify notifications according to three axes: the user's *interruption*, the required *reaction*, and the *comprehension* of the underlying event. However, they do not address the issue of interrupting the environment. Hanson et al. [7] address this issue explicitly. They introduce a model that allows to classify notifications according to their *publicity* and *subtleness*. They restrict themselves however to auditory and tactile notification cues.

Sawhney and Schmandt [14] address the issue of scalable, context–aware notifications for wearable computers. They restrict themselves to auditory context classification and audio notifications.

Few people have tried to estimate the user's interruptability using various sensors. Hudson et al. [9] published a Wizard of Oz study that explores how well the interruptability of office workers can be estimated using audio and video. According to their results the interruptability can be derived with some 75–80% accuracy. They do not address the issue of mobile interruptions.

Auditory Scene Classification. While auditory scene classification is not a new problem, it has seldom been used in practical applications. Büchler [2] gives a good survey on state of the art supervised techniques and their use for automatic hearing aid adaptation. He classifies into four classes *Speech*, *Noise*, *Music*, *Speech in Noise*, which are especially suited for the hearing aid application. Eronen et al. [5] have published an extensive supervised classification system and also performed a study how well humans perform in classifying auditory scenes. Clarkson et al. [3] employ a non–supervised approach to classify short duration sound objects and longer term sound scenes.

Since we use the output of the auditory scene analysis for further classification, we require more control over the auditory scene classification and hence opt for a supervised approach.

Activity Recognition. Classifying human activity based on acceleration data is a relatively new concept. Van Laerhoven [15] uses multiple, different sensors, to classify multiple user–specified contexts. Kern et al. [11] use multiple acceleration sensors to recognize and record the user's activity during a meeting.

3 Design Space of Notification

When considering notifications in a mobile setting, there are three principal aspects to consider. Firstly the current user activity and social situation, which together determine whether to notify or not, secondly the most appropriate notification means and finally the possibility of user intervention in exceptional situations. We devised a model, that allows to evaluate situations (section 3) and to serve for direct user control (section 3.3). A related model allows us to assess the intensity and modality of a notification (section 3.2).

3.1 Spanning the Design Space

Notifications are used to notify the user about an event that has happened, such as incoming phone calls or new e–mails. The event generally has a certain importance for the user. The notification itself has two sides for the user: on the one hand it has a value, because it conveys some important information about an event. On the other hand, it has a cost, because it interrupts the user in his current task. The value depends directly on the importance of the event to the user. As explained before in this paper, we focus on estimating the cost and assume the importance as given.

The cost depends directly on the degree of interruption. The degree of interruption depends on the interruptability

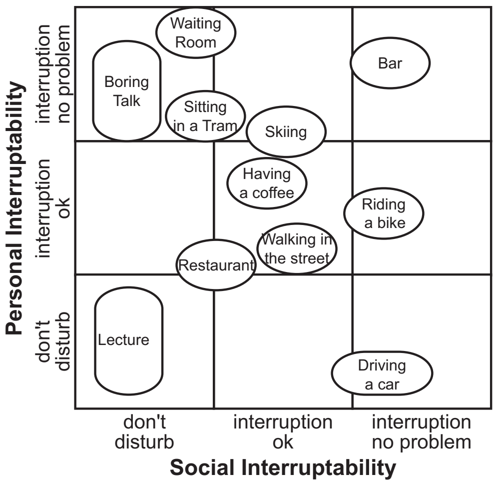

Figure 1. The Design Space of Notification

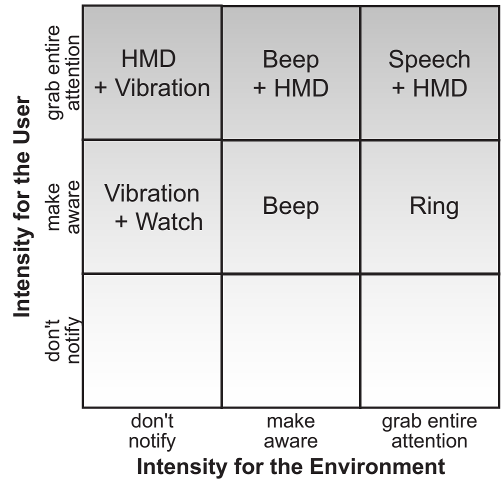

Figure 2. Classification of Notification Intensity of Different Modalities

of the user. This is highly depend on the current situation and activity of the user: e.g. while having a drink in a bar with some friends, a private call from the one missing friend would be appreciated, while the same call on the way home, while driving a car, would be much more interruptive. Therefore the *interruptability of the user* (referred to as *personal interruptability* in the following) is the key to evaluate the cost of a notification. It can be seen as a continuous variable that ranges from the extreme *Don't Disturb* over the intermediate range *Interruption OK* to the other end of the scale *Interruption No Problem*.

A notification does not necessarily reach the user only. An audio alarm can also be perceived by the environment — a potentially embarrassing situation, e.g. in a lecture. Thus we distinguish between the interruptability of the user (*personal interruptability*) and that of the environment (*social interruptability*). This allows us to choose a means of notification that interrupts the user only or interrupts both the user and the environment.

Since the social interruptability depends mainly on the social situation it is less dynamic than the personal interruptability, which depends strongly on the activity of the user. The social interruptability in a restaurant is the same whether the user is eating alone or not. However, the personal interruptability might be quite different and changing dynamically during the restaurant visit.

Our model consists thus of the two–dimensional space that is spanned by the personal interruptability and the social interruptability. Figure 1 shows the space with some example situations. The activity of 'Driving a car' requires much attention by the user, which has thus a low personal interruptability. It would however not disturb others, if he was notified, thus the social interruptability is high. For the situation 'Boring Talk' or 'Waiting Room' things are reversed: it would be highly unacceptable to notify the environment, e.g. using a loud ring, however an interruption of the user would probably be no problem if not appreciated.

3.2 Multi-Modality & Notification Intensity

Notifications are not binary, some allow to grab only part of the user's or the environment's attention. They can carry different amounts of information, they can be conveyed using different modalities, and some devices also allow to scale their interruptiveness (e.g. audio by changing the volume) [14]. The right notification modality should be chosen depending on the personal and social interruptability. We propose a scheme that allows to classify a notification's intensity and facilitates the matching from personal and social interruptability.

The intensity of a notification can be scaled from not notifying at all to trying explicitly to grab the entire user attention. We have to distinguish between the intensity for the user and the intensity for the environment. For example, a notification can be intense for the user, such as a flashing head mounted display (HMD), and completely imperceivable to the environment. We can thus classify notification modalities according to their intensity, as depicted in figure 2. Since it becomes harder to ignore a notification with increasing intensity, intense notifications can only be used when the user or the environment have enough attention available. This corresponds directly to the interruptability we are using in the Design Space of Notifications. Hence we can discretize the Design Space of Notification and make a simple one–to–one mapping to choose the best notification modality depending on the personal and social interruptability.

The intensity of a notification also varies with the amount of information conveyed. The amount of information can be varied from a simple binary pulse 'something has happened' to very complex information, such as the complete text of an e–mail. Depending on the device this can be a near–continuous space in which any amount of information can be conveyed or, for simpler devices, one of several distinct levels. Since the user needs more attention to deal with an information–rich notification, the intensity of a notification increases with the amount of information conveyed. In figure 2 the grey level indicates the amount of information that can be conveyed with a given notification.

Depending on the available devices this space has to be populated differently. In figure 2 we have depicted a selection of possible notification devices, that could be available to a wearable computer. Audio, as a classical notification means (a short beep, a loud ring, or a speech message), can be used in a wide variety of situations and modes, as long as the environment may perceive the notification. An HMD generally offers very intense notifications, so that it may only be used, when the user is highly interruptible. Notifications using an HMD are however not perceivable by the environment. Both amount of information conveyed and the intensity can be scaled using techniques such as size, colour, or animation of the notification. Finally a small vibrating device combined with a watch that has a built-in display, such as [10], can transfer a medium amount of information that is very easy to perceive and also to ignore.

3.3 User Control

An important aspect is that the user must be able to control the way he is notified. The Design Space of Notifications is simple enough that users can directly choose their personal and social interruptability by hand, if need be.

As opposed to the direct configuration of the notification modality, this allows to still select the best modality automatically, depending on the available devices. If the user left the HMD at home, he would not have to configure that he does not want notifications on the HMD. Instead the system can choose the best available notification modality automatically, depending on the available devices.

4 Estimating the Interruptability

In order to estimate the interruptability of user and environment, we propose to combine the inputs from all sensors and the prior that is given by the user. As pointed out in section 1 we do not model or recognize entire situations, but estimate the interruptability directly from the sensors, which give us a *tendency* for the interruptability.

When considering a single sensor only, we can already infer something about the user's and environment's interruptability. In the social situation 'Lecture', the user is probably personally little interruptible and quite certainly socially not interruptible. The sensor gives us a *tendency* which is the most probable interruptability. This tendency can be any kind of function that return the likelihood of interruptability on the entire Design Space of Notifications.

Figure 3 shows the tendencies we chose for the experiments in section 7. While activities such as 'Sitting' or 'Standing' can occur in nearly any situation, they cover a large area. However activities such as 'Stairs' or social situations such as 'Lecture' imply very specific interruptabilities. In the first case, the user needs much attention for finding his way, hence the personal interruptability is low, but he is generally in a public situation, where an interruption does not disturb. In the latter case, the user is devoting his attention to the lecture, thus he is little interruptible, and the environment must not be disturbed.

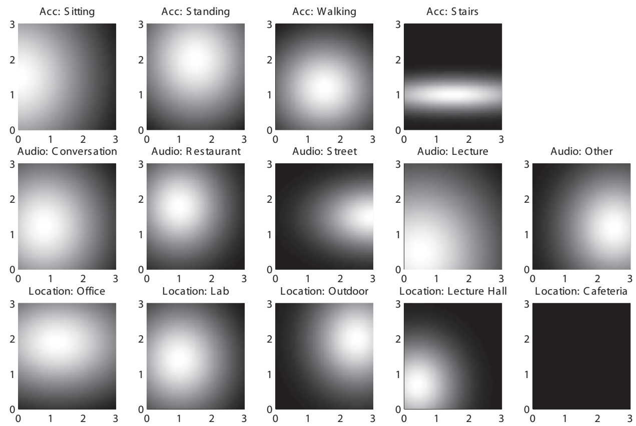

Figure 3. Tendencies for the sensors of section 7

Location can only give quite coarse priors about the current interruptability, especially the coarse location sensor we employ. Therefore the location tendencies all have a large variance and cover large portions the the Design Space of Notification. The space that is covered by the *Cafeteria* location is very large, covering not only the cafeteria but also substantial parts of the stair case. Hence the influence is practically neglegible for the interruptability estimation.

The next step after defining tendencies for all sensors is combining them to obtain a final estimation. Since our 'sensors' are not physical sensors, but rather classification sub–systems, we can obtain for every 'sensor–reading' a likelihood. We use this likelihood to combine the tendencies: they are weighted according to their classification likelihood and summed. This is a preliminary step for the final estimation and called the *sensor estimate*.

The final part to take into account is the user's prior. As explained in section 3.3 the user can set his personal and social interruptability manually. This can be seen as an additional tendency. However, it has to be treated separately, because it has to have the possibility of overriding all other tendencies. The sensor estimate is weighted with a confidence measure in its correctness, e.g. using its variance or entropy. The weighted sensor estimation is combined with the user's prior to produce the final estimate of personal and social interruptability.

Mathematical Formulation. This section describes how to encode the tendencies proposed above in a mathematical construct, that can be implemented on a computer.

As the tendencies are defined as any kind of function, there is an unlimited number of possible representations for them. They can be discrete on a regular grid of the design space of notification, or they can be continuous. The continuous ones can be simple linear functions or complex non–linear ones. Not all representations make necessarily sense for all sensors.

We chose for this paper to represent the tendencies by 2-dimensional Gaussian functions, because they combine a high expressiveness with few parameters that need to be set. Thus the tendency for sensor s at position x, y in the Design Space of Notification can be expressed as

$$T(x, y, s) = N(m_s, \Sigma_s)(x, y)$$

With mean vector m_s and, in our case, diagonal covariance matrix Σ_s. Figure 3 shows the tendencies we chose for the experiments in section 7.

Every sensor s recognizes the corresponding context at time t with likelihood $l(s, t)$. Thus we can compute the combined sensor estimate SE at time t in the Design Space of Notification

$$SE(x, y, t) = \sum_s T(s, x, y) \cdot l(s, t)$$

The final estimate of the interruptability at time t is the combination of the sensor estimate SE and the prior of the user:

$$\text{Interruptability}(t) = \text{argmax}_{x,y}(\delta \cdot SE(x, y, t) + U(x, y))$$

where $U(x, y)$ is the user's prior and δ the weighting factor for the sensor estimate. We currently assume δ as 1 and $U(x, y)$ as zero for all (x, y). We compute the argmax over the Design Space of Notification by sampling it on a regular grid.

5 Auditory Context Extraction

The goal of the auditory context extraction is to classify the social situation of the user. This problem is also called auditory scene classification. We classified the auditory scene into four social situations, that we found useful for notification, namely *street*, *restaurant*, *lecture*, and *conversation*. Here we introduce the approach and report experimental results. In section 7 we use it for the notification experiment.

Features. We use six features for the classification [2], based on the spectral centre of gravity, the tonality of the signal, the amplitude onsets, and the amplitude histogram width. Every audio sample is divided into 1sec *segments*. For all features but the tonality and the amplitude histogram width the segments are again divided into 30ms *frames*, for a sensible calculation of the spectrum. The classification is done on the segments, i.e. there is one feature vector every second.

The first two features are based on the *Spectral Centre of Gravity*. For the first, the centre of gravity is calculated on the spectrum of every frame. These centres are then averaged to obtain the feature for the segment. The second feature describes the temporal fluctuations of the centre of gravity. It is calculated using: $CGF = \log E[CG]/STDEV[CG]$, where CGF is the feature value for a segment, E[CG] is the expected centre of gravity over one segment and STDEV[CG] is the corresponding standard deviation.

The third feature describes the tonality of the signal. It is based on the normalized autocorrelation function (NACF) computed over a frame. If the maximum coefficient is sufficiently large, the signal is considered tonal. The feature value over one segment is the number of tonal frames.

The fourth and fifth feature are based on the Amplitude Onsets, which describe the activity in a certain band of the signal. The spectrum is divided into 20 Bark–bands [6]. The change in a bark band from one frame to the next defines the onset (onsets that are too small are set to zero). The first of the two features is the mean over a segment of the sum of all onsets in a frame. The second is the mean of the number of non–zero onsets per frame.

The sixths and last feature describes the width of the amplitude histogram. The histogram is computed on the maximum of 3ms sub–frames, scaled to DB units. The width between the 10– and 90–percentiles define the feature value.

Classification The classification is done using two–state ergodic Hidden Markov Models (HMM). Every class is described by a separate model. The class of the model with the highest a–posteriori likelihood is chosen as the final class for the preliminary experiments. For the notification experiment the likelihoods are used for combining the tendencies.

Experiments To show the performance, we evaluated the above classification scheme on a test–database. The database itself is composed of 54 samples of each 1min length recorded in 44kHz stereo. It contains 17 *Street* samples, 15 *Restaurant*, 12 *Lecture*, and 10 *Conversation* samples. They were recorded with a Sony ECM–TS125 clip–on microphone attached to the collar of the user. The samples are recorded at 44kHz sampling rate using 16 bit precision in stereo.

Table 1 summarizes the recognition results on the test database. To factor out the influence of a specific test or training set, we used 5–fold cross validation. The overall recognition result of 83.17% is satisfying. However, the two classes 'lecture' and 'conversation' are often confused. While they are similar (both contain indoor speech) it should be possible to differentiate one from the other, e.g. using signal strength. In this case it is however probably

	Recognized Class			
	street	restaurant	lecture	conv.
street	**82.35**	17.65		
restaurant	6.67	**86.67**	6.67	
lecture			**91.67**	8.33
conv.			28.00	**72.00**

Table 1. Evaluation of the audio context extraction on the test database using 5-fold cross validation (overall 83.17% recognition rate)

due to the fact, that the database is not sufficiently diverse for these two classes.

6 Activity Context Extraction

The goal of the activity context extraction is to classify the user's physical activity using an acceleration sensor. We classified the activity into *sitting*, *standing*, *walking*, *walking upstairs*, *walking downstairs*, and *running*. Here, we describe the approach and present experimental results. In section 7 we use it for the notification experiment.

Sensing. We use a Smart-It sensor board [1] to capture the sensor data. It contains a single 2–axis acceleration sensor. It is placed above the user's right knee on the thigh. The first axis points downward and the second in forward direction. The data is sampled at 50 Hz. The sensor board is attached to a Compaq IPAQ. It both records the data and allows for online annotation of the data using an interactive application.

Classification. The classification of the activity is done using Bayes' rule. The mean and variance for every axis over the last 50 samples are used as features. They are assumed to be independent. The a–priori–likelihoods are computed using the annotated data. The resulting probability distribution functions are represented as 100 bin histograms.

Experiments. To show the performance of the activity recognition, we have evaluated it on a test sample. The sample is 12.6min long and consists of the activities sitting (13.3% of the data), standing (27.4%), walking (34.7%), upstairs (8.8%), downstairs (7.0%), and running (8.8%).

Table 2 summarizes the recognition results on the test sample. While in general the recognition seems to work, the activities *upstairs* and *downstairs* are problematic. They are often confused with *walking*. This is probably due to the small amount of training data for both activities *and* to the similarity to the movement of the activity *walking*.

Figure 4 shows the recognition results over time. The system oscillates between walking and up/downstairs, which is expected given the previous results. Furthermore, the recognition has a small systematic delay over the entire sample. This is caused by the window over which the features are calculated (currently 50 samples corresponding to one second). However, reducing the window size introduces more noise to the features and thus lowers the recognition performance.

	Recognized Activity					
	sit	stand	walk	up–stairs	down–stairs	run
sit	**98.5**	1.5				
stand	1.5	**94.1**	4.4			
walk		2.3	**94.7**	1.6	0.3	1.1
upstairs		0.1	14.7	**85.2**		
downstairs			45.8		**53.4**	0.8
run	0.1		6.8			**93.1**

Table 2. Evaluation of the activity context extraction (overall 86.5% recognition rate)

7 Experiments

To show the feasibility of our approach we have conducted an experiment that includes both audio and acceleration context extraction, as well as location context from wireless LAN access points, in order to estimates the social as well as the personal interruptability of a user from them.

7.1 Experimental Setup

We recorded a 37.78min stretch of audio and acceleration data, in which we covered a broad variety of different situations. Sensors were attached to a laptop which recorded and synchronized the data. The audio and acceleration context was annotated manually during the recording using an externally attached Compaq IPAQ. Right after the recording misannotations were removed and the interruptabilities manually annotated by hand (see figure 5).

The recording contains a large variety of different situations and activities. Figure 5 shows the encountered situations at the very bottom. We started the recording in our lab and had a short walk on the street, during which we had two brief discussions. After coming back we attended a lecture for a short while and had a coffee in the computer science students' restaurant. We then went to our secretary and to our office, having a short chat with the secretary and our co–worker. On the way back to the lab we met a colleague on the corridor. After a short conversation we finished the recording in the lab.

For the context recognition we employ the audio and acceleration recognizers as described in sections 5 and 6 and location from the current wireless LAN access point. We have to change audio and acceleration recognition slightly, so that the frequencies of audio recognition, acceleration recognition and interruptability estimation matched. See figure 3 for the tendencies of every sensor.

For the audio context recognition we used a slightly different set of classes, *Conversation*, *Lecture*, *Restaurant*, *Street*, and *Other*. Classification is done every ten seconds,

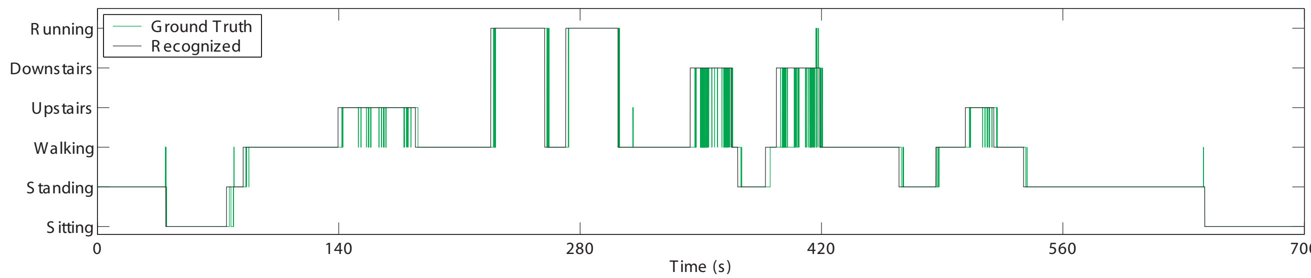

Figure 4. Activity Recognition over Time

which proved sufficient for the relatively slow changing auditory scene. The classification results are duplicated, so that one classification is available every second. The results are normalized to 1.

For the acceleration context recognition we changed the set of classes slightly. For this experiment we used *Sitting*, *Standing*, *Walking*, and *Stairs*. The former classes *Upstairs* and *Downstairs* are merged, because for notification there is no difference between walking up– and downstairs. The classification is done at a frequency of 55Hz. The results are down sampled to 1Hz and normalized to 1.

For the location context we use the wireless LAN access point to which the computer is currently associated. Due to the high density of access points, there was generally more than one access point per physical location in our recording. We have thus grouped the actual access points into the five groups *Office*, *Outdoor* (no access point), *Lecture Hall*, *Lab*, and *Cafeteria*. The current access point is sampled every second. The thus acquired location data is used both as 'ground truth' and as sensor value for the later experiments.

7.2 Discussion

Figure 5 shows an overview of the experimental results. The interruptabilities we use are in the range $[0; 3]$ (as in figure 1). The axis at the bottom indicates the situation the user was in. While the first and third plot show the ground truth of social resp. personal interruptability, the second and fourth plot show the interruptability estimation error for the social and personal interruptability respectively.

It is important to note, that although the space and the estimation of the interruptabilities are continuous, the selection of notification modalities requires a discretization of the space. Using the discretization in three bins as shown in figure 2, we can tolerate errors up to 0.5 in either direction, because the interruptability will be discretized into the same bin.

In order to verify that using tendencies is a sensible approach, we used the context ground truth as sensors and estimated the interruptabilities. The dark lines in plots two and four indicate these results. The small variance of the estimation error of 0.20 for the social interruptability and 0.13 for the personal interruptability show that using tendencies is a feasible approach. The error of the social interruptability estimation is below 0.5 and thus sufficiently precise for 88.5% of the time resp. 96.3% for the personal interruptability.

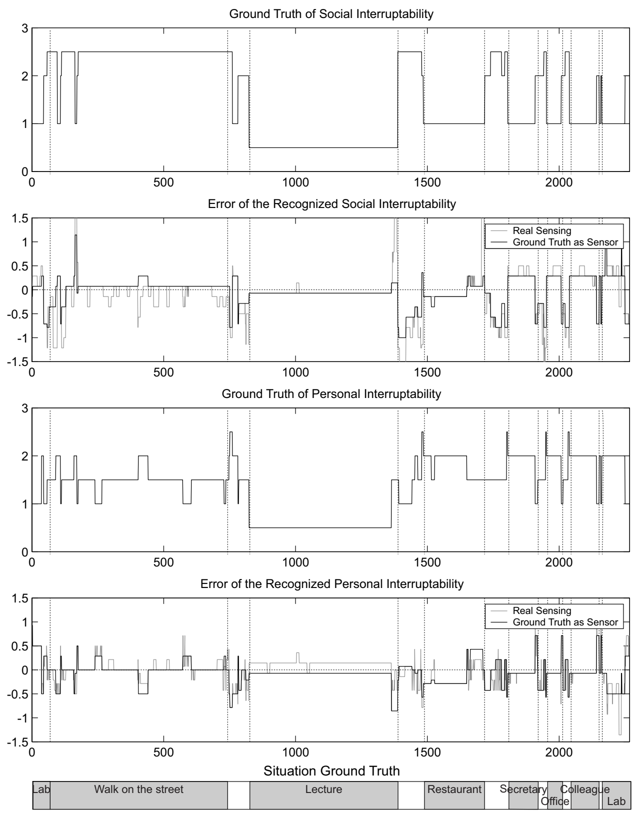

Figure 5. Ground Truth Interruptability and Estimation Error

The estimation error of the personal interruptability using real sensors is indicated in the fourth plot in grey. The sensing introduces little additional error. The variance of the estimation error is with 0.16 close to the one where we used the ground truth as 'sensor'. The error is 96.2% of the time below 0.5 and thus sufficiently small for modality selection. The outlier at second 2230 is due to a misclassification in the auditory scene.

The second plot in figure 5 shows the estimation error for the social interruptability using real sensors. When comparing the error variance of 0.58 with the error variance of 0.20 when using the ground truth as sensors leads to the conclusion, that the sensing introduces considerable noise. The error is for 86.0% of the time below 0.5 and thus sufficiently precise. It is important to note, that the error tends to be negative. This means that the system would rather have chosen not to notify the environment instead of notifying too aggressively. As opposed to the contrary, notifying too much instead of too little, this is the 'better' kind of error.

In order to determine which part of the sensing to improve, we consider the respective recognition rates of audio and acceleration context recognition. While the acceleration context is recognized in 91.9% of the time, the audio context is context is recognized only 65.5% of the time. Given this and the fact, that the social interruptability mainly depends on the social situation, we conclude, that audio context recognition needs to be improved in order to improve the estimation of the social interruptability.

Although the location information we use is quite coarse, it provides useful additional priors. Without the location information, the error of the social interruptability estimation is below 0.5 77.0% of the time, as opposed to 86.0% when using the location. The impact on the personal interruptability estimation is smaller: the estimation error is below 0.5 94.61% of the time without location and 96.16% of the time when using location.

8 Conclusion and Outlook

Managing notifications is an important problem for future wearable computing devices and applications. They should be mediated automatically in order to avoid user annoyance. To this end we propose to use context acquired from multiple, body–worn sensors.

We have introduced a model that classifies notifications according to the *personal* and *social interruptability* of the user. It allows both to handle notifications in a systematic manner and to enable the automatic selection of the best notification modality.

We do not model the entire user context but estimate the interruptabilities directly from the sensor readings. Every sensor contributes a *tendency* about the user's current personal and social interruptability. The user's prior is integrated as a special tendency that can override the other ones.

Preliminary experiments have shown the feasibility of the approach experimentally. We used context information from acceleration and audio sensors. The estimation of the personal interruptability provides a sufficiently precise estimate 94.6% of the time. The estimation of the social interruptability is not quite as reliable, due to the less reliable audio context recognition.

Of course, several issues need to be addressed further. Since our results are only preliminary, further investigation will be necessary. The audio context recognition should be improved, in order to improve the estimation of the social interruptability. In general, other types of sensors could supply interesting context information. To investigate the influence of various kinds of sensors and contexts, we are planning more experiments as well as user studies.

References

[1] M. Beigl, T. Zimmer, A. Krohn, C. Decker, and P. Robinson. Smart-its - communication and sensing technology for ubicomp environments. technical report issn 1432-7864. Technical report, TeCo, University Karlsruhe, Germany, 2003/2.

[2] M. Büchler. *Algorithms for Sound Classification in Hearing Instruments*. PhD thesis, ETH Zurich, 2002.

[3] B. Clarkson, N. Sawhney, and A. Pentland. Auditory context awareness via wearable computing. *Workshop on Perceptual User Interfaces*, November 1998.

[4] M. Cutrell, M. Czerwinski, and E. Horvitz. Notification, disruption, and memory: Effects of messaging interruptions on memory and performance. In *Proc. of Interact*, 2001.

[5] A. Eronen, J. Tuomi, A. Klapuri, D. Fagerlund, T. Sorsa, G. Lorho, and J. Huopaniemi. Audio–based context awareness — acoustic modeling and perceptual evaluation. In *Proc. Intern. Conf. on Acoustics, Speech, and Signal Processing, ICASSP*, 2003.

[6] B. Gold and N. Morgan. *Speech and Audio Signal Processing*. John Wiley & Sons, 2000.

[7] R. Hansson, P. Ljungstrand, and J. Redström. Subtle and public notification cues for mobile devices. *Lecture Notes in Computer Science, Ubicomp*, 2201:240–247, 2001.

[8] 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.

[9] S. Hudson, J. Fogarty, C. Atkeson, D. Avrahami, J. Forlizzi, S. Kiesler, J. Lee, and J. Yang. Predicting human interruptibility with sensors: a wizard of oz feasibility study. In *Proceedings of the conference on Human factors in computing systems*, pages 257–264. ACM Press, 2003.

[10] N. Kamijoh, T. Inoue, C. M. Olsen, M. Raghunath, and C. Narayanaswami. Energy trade-offs in the ibm wristwatch computer. In *ISWC*, pages 133–140, 2001.

[11] N. Kern, B. Schiele, H. Junker, P. Lukowicz, and G. Troester. Wearable sensing to annotate meetings recordings. In *Proc. ISWC*, 2002.

[12] D. S. McCrickard, R. Catrambone, C. Chewar, and J. T. Stasko. Establishing tradeoffs that leverage attention for utility: Empirically evaluating information display in notification systems. *International Journal of Human–Computer Studies*, 8(5):547–582, May 2003.

[13] 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.

[14] N. Sawhney and C. Schmandt. Nomadic radio: Scalable and contextual notification for wearable audio messaging. In *Proc. ACM CHI*, pages 96–103, May 1999.

[15] K. van Laerhoven, K.Aido, and S. Lowette. Real-time analysis of data from many sensors with neural networks. In *ISWC*, pages 115–123, 2001.