Using the "HotWire" to Study Interruptions in Wearable Computing Primary Tasks

Mikael Drugge[1], Hendrik Witt[2], Peter Parnes[1], Kåre Synnes[1]
[1] *Media Technology, Luleå University of Technology, SE-97187 Luleå, Sweden*
[2] *TZI, Wearable Computing Lab., University of Bremen, D-28359 Bremen, Germany*
mikael.drugge@ltu.se, hwitt@tzi.de, peter.parnes@ltu.se, kare.synnes@ltu.se

Abstract

As users of wearable computers are often involved in real-world tasks of critical nature, the management and handling of interruptions is crucial for efficient interaction and task performance. We present a study about the impact that different methods for interruption have on those users, to determine how interruptions should be handled. The study is performed using an apparatus called "HotWire" for simulating primary tasks in a laboratory experiment, while retaining the properties of wearable computers being used in mobile, physical, and practical tasks.

1. Introduction

In stationary computing users concentrate mainly on one task to be performed with the computer. Wearable computing, however, typically expects users to accomplish two different tasks. A primary task involves real world physical actions, while the secondary task is often dedicated to interacting with a wearable computer. As these two tasks often interfere, studying interruption aspects in wearable computing is of major interest in order to build wearable user interfaces that support users during work with minimized cognitive load.

1.1. Motivation

Limitations of human attention have been widely studied over decades by psychological science. What we commonly understand as attention consists of several different but interrelated abilities [5]. In wearable computing we are particularly interested in divided attention, i.e. the ability of humans to allocate attention to different simultaneously occurring tasks. It is already known that divided attention is affected by different factors such as task similarity, task difference, and practice [3]. The question of when to interrupt a user can be decided by estimating human interruptability [4], while the question of how depends on the methods used. Although studying divided attention has already provided detailed findings, applying and validating them for wearable computing is still a challenging issue. Once approved, they can be used in wearable user interface design to adapt the interface to the wearer's environment and task. Furthermore, being able to measure such attention enables the specification of heuristics that can help to design the interface towards maximal performance and minimal investment in attention [8]. Here, however, a major problem is the simulation of typical real world primary tasks under laboratory conditions. Such simulation is needed to analyze coherence between attention on a primary task and user performance in different interaction styles.

In this paper we present a study of different ways to interrupt a user performing a physical task. We will investigate the correlations between cognitive engagement, interruption type, and overall performance of the users.

1.2. Outline

The remainder of the paper is structured as follows: Section 2 reviews related work to the presented interruption study. Then, in section 3 we describe the experiment conducted including the different interruption methods tested. Section 4 explains the user study itself and the apparatus used for primary task simulation. The results are discussed in section 5, while the apparatus itself is evaluated in 6. Finally, section 7 concludes the paper.

2. Related Work

In [6], McFarlane presents the first empirical study of all four known approaches to coordinate user interruption in human-computer interaction with multiple tasks. The study concerns how to interrupt users within the context of doing computer work without increasing their cognitive load. The method applied in the laboratory experiments was based

1-4244-0598-X /06/$20.00 ©2006 IEEE.

on a simple computer game that requires constant user attention, while being randomly interrupted by a color and shape matching task. As a continuation of McFarlane's original interruption study for the scope of wearable computing, in [2] a head-mounted display (HMD) was used to display the matching tasks. It was found that the scheduled approach gave the best performance, while using notifications came second although with shorter response time. As wearable computers are closely connected to the user, performance is not the only factor to be considered — the user's preferences on interruption also need to be taken into account. In [7] it was found that audio notification appeared to give slightly better performance although users considered it more stressful, compared to visual signals that on the other hand were more distracting for the primary task. Although the mentioned work was able to relate human-computer interaction findings to wearable computing, the conducted laboratory experiments only use virtual primary tasks in form of computer games. This does not entirely encompass the properties of wearable computers being used in mobile and physical tasks, indicating that a follow-up study is needed to complement the earlier studies.

3. Experiment

The experiment addresses how different methods of interrupting the user of a wearable computer affects that person's cognitive workload. The scenario involves the user performing a primary task in the real world, while interruptions originate from the wearable computer and call for the user to handle them. By observing the user's performance in the primary task and in the interruption task, conclusions can be drawn on what methods for handling interruptions are appropriate to use. In order to measure the user's performance in both types of tasks, these must be represented in an experimental model. This section describes each task and how they are combined in the experiment.

3.1. Primary Task

The primary task needs to be one that represents the typical scenarios in which wearable computers are being used. Primary tasks in wearable computing are often physical tasks, i.e. tasks that require users to work with their hands on real world objects while being mobile (e.g. assembly or inspection tasks). For the purpose of our study, the task has to be easy to learn by novice users to reduce errors in the experiment caused by misunderstandings or lack of proficiency. The time to make the user proficient and fully trained should also be short enough to make a practice period just before the actual experiment sufficient, so that the user's performance will then remain on the same level throughout the experiment. To simulate such a task in a controlled laboratory environment, we decided to use the "HotWire" experimental setup [9].

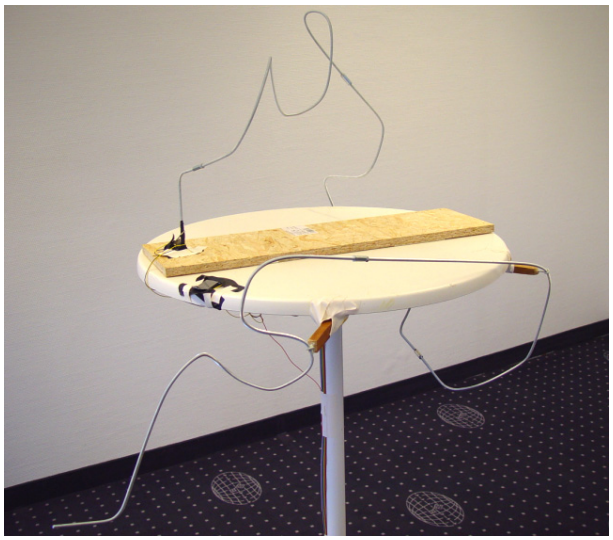


Figure 1. The HotWire apparatus used.

The HotWire apparatus was developed for simulating primary tasks that satisfy the requirements discussed above. It is based on a children's game commonly known as "The Hot Wire". It consists of a metallic wire bent in different shapes that is mounted on both ends to a base plate, plus a special tool with a grip and a metallic ring. The idea of the game is that a person has to pass the ring from one end of the wire to the other end without touching the wire itself. If the wire is touched with the ring while being on the track an acoustic feedback indicates an error. For our apparatus, shown in figure 1, we constructed the bent metallic wire out of differently shaped smaller segments each connected via windings to another segment. This allows the difficulty or characteristic of the primary task to be varied by replacing or changing the sequence of connected segments.

3.2. Interruption Task

The secondary task consists of matching tasks presented in the user's HMD. An example of this is shown in figure 2. Three figures are shown of random shapes and colors, and the user must match the figure on top with either the left or the right figure at the bottom of the display. A text instructs the user to match either by color or by shape, making the task always require some mental effort to answer correctly. There are 3 possible shapes (square, circle, triangle) and 6 colors (red, yellow, cyan, green, blue, purple), allowing for a large number of combinations. Tasks are created at random so that on average a new task appears every five seconds, and if the user is unable to handle them soon

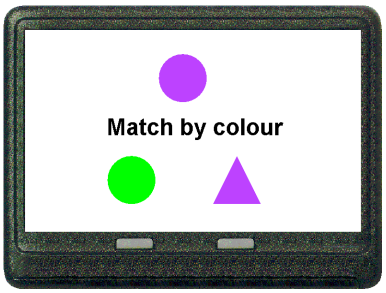


Figure 2. Matching task presented in HMD.

enough they will be added to a queue of pending tasks.

3.3. Methods for Handling Interruptions

The methods used for managing the interruptions are based on the four approaches described in McFarlane's taxonomy in [6]. During all of these methods, the user performs the HotWire primary task while being subject to interruption. The methods used are as follows

- **Immediate:** Matching tasks are created at random and presented for the user in the instant they are created.
- **Negotiated:** When a matching task is randomly created, the user is notified by either a visual or audible signal, and can then decide when to present the task and handle it.
- **Scheduled:** Matching tasks are created at random but presented for the user only at specific time intervals of 25 seconds, typically this causes the matching tasks to queue up and cluster.
- **Mediated:** The presentation of matching tasks is withheld during times when the user appears to be in a difficult section of the HotWire. The algorithm used is very simple; based on the time when a contact was last made with the wire, there is a time window of 5 seconds during which no matching task will be presented. The idea is that when a lot of errors are made, the user is likely in a difficult section so no interruption should take place until the situation is better.

In addition to these methods, there are also two base cases included serving as reference. These are as follows

- **HotWire only:** The user performs only the HotWire primary task without any interruptions, allowing for a theoretical best case performance of this task.
- **Match only:** The user performs only the matching tasks for 90 seconds, approximately the same period of time it takes to complete a HotWire game. This allows for a theoretical best case performance.

Taken together, and having two variants — audio and visual notification — for the negotiated method, there are seven methods that will be tested in the study.

4. User Study

A total of 21 subjects were selected for participation from students and staff at the local university — 13 males and 8 females aged between 22–67 years (mean 30.8). The study uses a within subjects design with the method as the single independent variable, meaning that all subjects will test every method. To avoid bias and learning effects, the subjects are divided into counterbalanced groups where the order of methods differs. As there are seven methods to test, a Latin Square of the same order was used to distribute the 21 participants evenly into 7 groups with 3 subjects in each.

A single test session consists of one practice round where the subject gets to practice the HotWire and matching tasks, followed by one experimental round during which data is collected for analysis. The time to complete a HotWire game naturally varies depending on how quick the subject is, but on average pilot studies indicated it will take around 90–120 seconds for one single run over the wire. With 7 methods of interruption to test with short breaks between each, one practice and one experimental round, plus time for questions and instructions, the total time required for a session is around 40–45 minutes.

4.1. Apparatus

The apparatus used in the study is depicted in figure 3, where the HotWire is shown together with a user holding the ring tool and wearing a HMD. The HotWire is mounted around a table and approximately 4 meters in length. To avoid vibrations because of its length, the wire was stabilized with electrically isolated screws in the table. An opening in the ring allowed the subject to move the ring past the screws while still staying on track. To follow the wire with the tool, the user needs to move around the table over the course of the experiment. The user may also need to kneel down or reach upwards to follow the wire, furthermore emphasizing the mobile manner in which wearable computers are used. Figure 4 illustrates the variety of body positions observed during the study.

In the current setup, the user is not wearing a wearable computer per se, as the HMD and tool is connected to a stationary computer running the experiment. However, as the wires and cabling for the HMD and tool are still coupled to the user to avoid tangling, this should not influence the outcome compared to if a truly wearable computer had been used. In particular, we also used a special textile vest the users have to wear during the experiment that was designed and tailored to unobtrusively carry a wearable computer, as

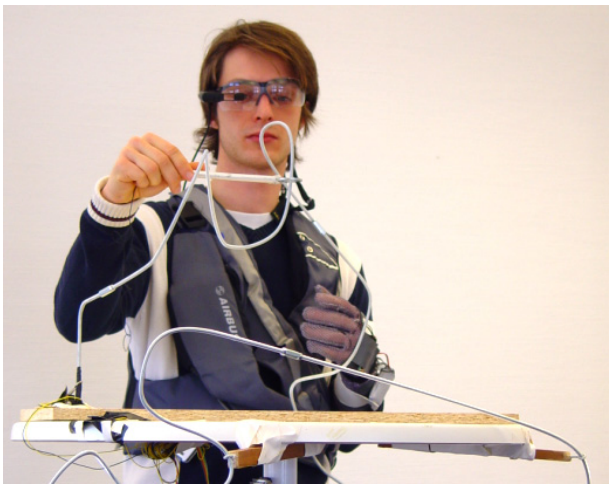


Figure 3. Experiment performed by a user.

well as all needed cabilings for a HMD without effecting the wearers freedom in movement. For having an even more realistic situation we put a OQO micro computer in the vest to simulate also the weight a wearable computer equipment would have outside the laboratory environment.

The matching tasks are presented in a non-transparent SV-6 monocular HMD from MicroOptical. A data-glove used in earlier research [1] is worn on the user's left hand serving as the interface to control the matching tasks. To ensure maximum freedom in movement of the user, the data-glove uses a Bluetooth interface for communication with the computer. By tapping index finger and thumb together, an event is triggered through a magnetic switch sensor based on the position of the user's hand at the time. Using a tilt sensor with earth gravity as reference, the glove can sense the hand being held with the thumb pointing left, right or upwards. When the hand is held in a neutral position with the thumb up, the first of any pending matching tasks in the queue is presented to the user in the HMD. When the hand is turned to the left or to the right, the corresponding object is chosen in the matching task. For the negotiated methods, the user taps once to bring the new matching tasks up, and subsequently turns the hand to the left or right and taps to answer them. For the immediate and mediated methods where matching tasks appear without notification, the user need only turn left or right and tap. Because of the novelty of the interface, feedback is required to let the user know when an action has been performed. In general, any feedback will risk interfering with the experiment and notifications used, but in the current setup an audio signal is used as it was deemed to be the least invasive. In order not to confound the user, the same audio signal was used regardless of whether the user answered correctly or not.

(a) Standing (b) Kneeling (c) Bending

Figure 4. Different body positions observed.

5. Results

After all data had been collected in the user study, the data was analyzed to study which effect different methods had on user performance. For this analysis, the following metrics were used

- **Time:** The time required for the subject to complete the HotWire track from start to end.
- **Contacts:** The number of contacts the subject made between the ring and the wire.
- **Error rate:** The percentage of matching tasks the subject answered wrong.
- **Average age:** The average time from when a matching task was created until the subject answered it, i.e. its average age.

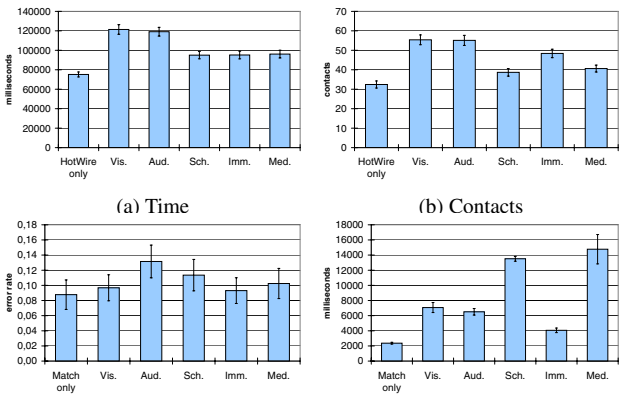
The graphs in figure 5 summarizes the overall user performance by showing the averages of the metrics together with one standard error.

A statistical repeated measures ANOVA was performed to see whether there existed any significant differences among the methods used. The results are shown in table 1. For all metrics except the error rate, strong significance ($p<0.001$) was found indicating that differences do exist.

Metric	P-value
Time	**<0.001**
Contacts	**<0.001**
Error rate	0.973
Average age	**<0.001**

Table 1. Repeated measures ANOVA.

To investigate these differences in more detail, paired samples t-tests were performed comparing the two base cases (HotWire only and Match only) to each of the five interruption methods. The results are shown in table 2. To ac-



(a) Time (b) Contacts

(c) Error rate (d) Average age

Figure 5. Averages of user performance.

comodate for multiple comparisons, a Bonferroni corrected alpha value of 0.003 (0.05/15) was used when testing for significance.

Metric	Vis.	Aud.	Sch.	Imm.	Med.
Time	<**0.0001**	<**0.0001**	<**0.0001**	**0.0002**	**0.0003**
Contacts	<**0.0001**	<**0.0001**	**0.0022**	<**0.0001**	**0.0004**
Error rate	0.7035	0.1108	0.0668	0.8973	0.4979
Average age	**0.0012**	**0.0001**	<**0.0001**	0.0194	0.0046

Table 2. Base case comparison t-tests.

All of these differences are expected; the completion time will be longer when there are matching tasks to do at the same time, and the error rate is likely to increase because of that reason. Also, the average age is expected to be longer than for the base case since the user is involved with the HotWire when matching tasks appear, and both the scheduled and mediated methods will by definition cause matching tasks to queue up with increased age as a result. That no significant differences in the matching tasks' error rate was found was unexpected, intuitively there should be more mistakes made when the subject is involved in a primary task. However, when looking at the data collected, most subjects answered the tasks as good in the interruption methods as they did in the base case of match only. Since there was nothing in the primary task that "forced" the subjects to make mistakes, as e.g. imposing a short time limit on the tasks would certainly have done, the subjects mainly gave accurate rather than quick and erroneous answers. All in all, this comparison of methods with base cases shows that in general, adding interruptions and a dual task scenario with a physical and mobile primary task will be more difficult for the subject to carry out successfully.

Following, the five interruption methods were then compared to each other using a paired samples t-test, the results of which is shown in table 3. As can be seen, a number of significant differences were found between the interruption methods. We will now analyze each of the metrics in turn to learn more about the characteristics of each method.

Time	Vis.	Aud.	Sch.	Imm.	Med.
Vis.	-	0.6859	<**0.0001**	**0.0001**	<**0.0001**
Aud.	0.6859	-	**0.0003**	<**0.0001**	<**0.0001**
Sch.	<**0.0001**	**0.0003**	-	0.9773	0.8157
Imm.	**0.0001**	<**0.0001**	0.9773	-	0.7988
Med.	<**0.0001**	<**0.0001**	0.8157	0.7988	-

Contacts	Vis.	Aud.	Sch.	Imm.	Med.
Vis.	-	0.9434	**0.0002**	0.1508	**0.0006**
Aud.	0.9434	-	<**0.0001**	0.0240	**0.0002**
Sch.	**0.0002**	<**0.0001**	-	**0.0038**	0.4217
Imm.	0.1508	0.0240	**0.0038**	-	**0.0031**
Med.	**0.0006**	**0.0002**	0.4217	**0.0031**	-

Error rate	Vis.	Aud.	Sch.	Imm.	Med.
Vis.	-	0.2744	0.4335	0.9041	0.8153
Aud.	0.2744	-	0.5258	0.3356	0.1039
Sch.	0.4335	0.5258	-	0.5852	0.6118
Imm.	0.9041	0.3356	0.5852	-	0.7668
Med.	0.8153	0.1039	0.6118	0.7668	-

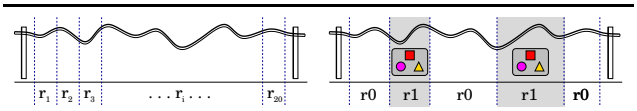
Average age	Vis.	Aud.	Sch.	Imm.	Med.
Vis.	-	0.5758	**0.0001**	0.0470	0.2180
Aud.	0.5758	-	<**0.0001**	0.0170	0.1411
Sch.	**0.0001**	<**0.0001**	-	<**0.0001**	0.3256
Imm.	0.0470	0.0170	<**0.0001**	-	0.0061
Med.	0.2180	0.1411	0.3256	0.0061	-

Table 3. Pairwise t-tests of methods.

5.1. Time

With regards to the completion time, the interruption methods can be divided into two groups; one for the two negotiated methods (visual and audio), and one for the remaining three methods (scheduled, immediate and mediated). There are strong significant differences between the two groups, but not between the methods in the same group. The reason for the higher completion time of the negotiated methods is because of the extra effort required by the user to present matching tasks. As this additional interaction required to bring the tasks up is likely to slow the user down, this result was expected. An important finding was, however, that the overhead (24.8 seconds higher, an increase of 26%) was much higher than expected. A lower overhead was expected, considering the relative ease — in theory — of holding the thumb upwards and tapping thumb and finger together to present the matching tasks, but in practice the subjects found this to be difficult when doing it simultaneously as the HotWire primary task. The data-glove itself accurately recognizes the desired gestures when done right, but the problem is that the subjects experience problems because their sense of direction is lost when doing the physi-

cal task, something we noticed when watching videos of the subjects in retrospect. Relating to our findings in [2], where the primary task was less physical as the user sat in front of a computer and interacted using a keyboard, we see that even seemingly simple ways to interact can have a much higher impact when used in wearable computing scenarios. Therefore, we argue that using a more physical primary task can increase the validity of user studies in wearable computing.



(a) Fixed-length (b) Interruption-based

Figure 6. Segmenting the track for analysis.

5.2. Contacts

Looking at the number of contacts between the ring and the wire, i.e. the number of physical errors the subject made in this primary task, we can discern three groups for the methods. The two negotiated methods form one group, where the additional interaction required to present matching tasks also cause more contacts with the wire. The scheduled and mediated methods form a second group with the lowest number of hotwire contacts. The immediate method lies in between and significant differences for this method were only found for the scheduled and mediated methods. It is of interest to know what causes these differences, if it is interference with the subject's motorical sense because of the dual tasks, or some other underlying factor.

As can be seen, there is a correlation between the completion time and error rate, which can be interpreted as indicating that the number of contacts made depends mainly on the time spent in the HotWire track, and is not affected by the different interruption methods per se. To analyze this further, the rate r of contacts over time was examined.

$$r = \frac{contacts}{time}$$

When comparing this rate between all interruption methods, no significant differences were found. This can be expected because of the correlation between time and contacts made. However, since there are both easy and more difficult sections of the HotWire, such a naive way of computing the overall contact rate risks nullifying these changes in track difficulty. To examine the error rate in detail and take the HotWire track itself in account, assuming the user moved the ring with a constant speed on average, we divided the track in 20 segments (see figure 6(a)) and compared the rate r_i per segment i between the methods[1]. However, no significant differences could be found here either. This suggests that our experiment was unable to uncover the impact of the interruption method as a whole, if such an effect exists, on the amount of contacts made in the HotWire.

Assuming that solely the appearance of matching tasks in the HMD cause more contacts being made, we decided to test this hypothesis. The contact rates were divided in two categories; $r0$ indicated the rate of contacts over time when no matching task was present in the HMD, while $r1$ indicated the rate of contacts over time with a matching task visible (see figure 6(b)). The rates $r0$ and $r1$ then underwent a paired samples t-test for each of the interruption methods, to see whether the means of these two kind of rates differed. According to the hypothesis, having a matching task present in the HMD should increase the contact rate $r1$ compared to the rate $r0$ when no matching task is present. Surprisingly, no significant difference was found. This can be taken as indication that either no difference exists, or more likely, that the number of contacts made by our HotWire apparatus is too random so that the smaller underlying effects of having a matching task present become lost in this noise. As our initial version of the HotWire apparatus [9] could reveal these differences with stronger significance in pilot studies, it suggests the version used in this larger study simply became too difficult. Since the user now needed to walk around the track and change into different body positions, this would cause more random contacts being made than with a version where the user stands still, thereby causing so big variance in the data collected that small differences caused by the matching task or interruption method cannot be found.

To determine whether the methods influence the subject overall and make him or her more prone to make errors, we compared first the rate $r1$ between different methods, and then $r0$ in the same manner. For $r1$, when there was a matching task shown, the mediated interruption method had the lowest contact rate (0.38) while immediate had the highest rate (0.69), yet with p=0.04 this is not significant enough to state with certainty when Bonferroni correction is applied. For $r0$, however, the mediated interruption method still had the lowest contact rate (0.33), while the two negotiated methods had the highest (both 0.48), and this difference was observed with significance $p<0.003$ confirming the hypothesis that the mediated method will help reduce this number. This finding shows that the algorithm we used for the mediated method can make the user perform the primary task slightly better in between interruptions, compared to letting her negotiate and decide for herself when to present the matching tasks.

1 To get a more accurate segmentation, the ring's position on the track would need to be monitored over time, something our current apparatus does not yet support.

5.3. Error rate

The error rate for the matching tasks exhibited no significant differences regardless of method. One reason for this is likely that a majority of the subjects answered all matching tasks correctly, (the median was zero for all methods except negotiated), while four subjects had very high consistent error rates (20$\sim$70%) through all methods, including the base case, that contributed to a high variance. In other words, the matching task may be a bit too easy for most people, while some can find it very difficult to perform.

A difference found compared to [2] is that the error rates for negotiated audio and visual have been exchanged so that audio, rather than visual, now exhibits worse performance. Although this cannot be said with statistical certainty in either case, it may indicate that differences do exist between subjects and their preference, and likely also by the kind of primary task being done.

5.4. Average age

Naturally, the average age is expected to be the highest for the scheduled method, since the matching tasks are by definition queued for an expected 12.5 seconds on average. This was also found with strong statistical significance ($p<0.0001$) for all methods but mediated. With an average age of 13.5 seconds on average, and an expected age of 12.5 seconds, this means the user only spends on average 1 second to respond to the queued matching tasks. Comparing this to the immediate (4.1 sec) and negotiated (6.5 and 7.1 sec) methods, this is significantly ($p\leq0.0002$) faster, likely because the need to mentally switch between primary and matching task is reduced because of the clustering.

Mediated on the other hand exhibited such high variance in its data, about an order of magnitude larger than for the other methods, so no real significant differences could be shown. The reason for this high variance is because the mediated algorithm was based on a fixed time window, and for some users who made errors very frequently this time window was simply too large so that the queued matching tasks showed up very seldom.

6. Evaluating the apparatus

Since the HotWire is an apparatus for evaluating wearable user interfaces, it is important to determine how suitable it is compared to other laboratory setups. In [2] a computer game and keyboard was used in a non-mobile setting where the user sat still during the course of the study, and we will use this as reference setup for the comparison.

The task of matching was the same in both studies, with minor differences in the frequency of appearance and the HMD used to present them in, as well as the physical means to interact with the task. As can be seen, the metrics that are comparable across the studies — the error rate and the average age — had a better significance in the former study. This would indicate that our current setup is less likely to uncover differences, if such exist, compared to the former non-mobile setup. Reasons may be that our study used a shorter time span for each method and that a novel interaction method was used, thereby increasing the variance of the data collected and diminishing the significance by which differences can be observed.

The primary task cannot easily be compared across studies; in the former study the number of errors was bounded and time was kept constant, whereas in our new study both errors and completion time are variable and unbounded. The former study thus had the errors as the only metric, whereas the HotWire offers both errors and time as metrics of performance. However, what can be seen is that in the former study no real significant differences could be found for the error metric between methods. With the HotWire, strong significant differences were observed in a majority of the tests for both the error and time metrics. This shows that differences do indeed exist between the interruption methods, and that these can more easily be uncovered by the apparatus we used. Therefore, as the HotWire apparatus is more mobile, physical, and more realistically represents a wearable computing scenario, we argue that using this in favour of the stationary setup is better for evaluating and studying wearable user interfaces.

Considering the fact that very few significant differences could be observed when looking into closer detail on the errors over time, as discussed in section 5.2, this basically indicates that there are more factors that need to be taken in account for research in wearable interaction. Ease of interaction, mobility, walking, changing body position, using both hands to handle the dual tasks — all of these factors cause errors being made in the primary task, while the effects of interruption and the modality used have less impact. Thus, we argue that the HotWire can aid in focusing on the problems most relevant in wearable computing interaction, as details that are of less importance in the first stages are clearly not revealed until the important problems are dealt with. In our study, we used a data-glove that is conceptually simple to operate — the user can select left, right, or up — yet even this was shown to be too difficult when operated in a more realistic wearable computing scenario.

7. Conclusions

The recommendation when implementing efficient interruption handling in wearable computing scenarios is to examine the needs of the primary and secondary task, and choose the method which best adheres to these as there are specific advantages and drawbacks with each method. The

HotWire study both confirms and complements the findings in [2] and [7] applied in a wearable computing scenario. Overall, the scheduled, immediate, and mediated methods result in fewer errors than the negotiated methods. Scheduled and mediated methods cause a slower response to the matching tasks, whereas immediate allows for quicker response at the cost of more errors in the primary task. The algorithm used in the mediated method was, despite its simplicity, able to reduce the error rate in the primary task in between the matching tasks compared to the negotiated method. Therefore, it can in certain situations be better to utilize context awareness and take the primary task in account, rather than explicitly allowing the user to decide when matching tasks should be presented. The new metric of completion time indicates that a significant overhead on the primary task is imposed when subjects get to negotiate and decide when to present the matching tasks, which also results in a larger number of errors being made. The cause of this was unforeseen difficulties in the interaction, even though a conceptually simple data-glove was used to control the matching. This suggests that efforts should primarily be focused on improving the interaction style and ease of use, while the actual methods used for interruption is of secondary importance.

The architectural implications of the different methods will still be relevant to consider in any case. Assuming the wearable computer is part of a more complex system where interruptions originate from elsewhere, the immediate and negotiated methods both require continuous network access so that the task to handle can be forwarded to the user immediately. On the other hand, the clustering of tasks that result from the scheduled and mediated methods may only require sporadic access, e.g. at wireless hot-spots or certain areas in the working place with adequate network coverage.

The HotWire apparatus itself demonstrated that many findings from non-mobile interruption studies could be confirmed, while also pointing out that there are inherent differences in wearable computing due to mobility and performing physical primary tasks. These differences cause some findings to stand out stronger than other, and as the apparatus more accurately resembles a realistic wearable computing scenario, this will better help guide research in wearable interaction to the areas where most focus is needed in the first stages of development. Since this represents a compelling (and worst case) scenario involving very high cognitive and physical workload, the results can likely be applicable in application domains with more relaxed constraints such as business and consumer use.

7.1. Future Work

For more accurate and in-depth analysis of the data collected from the HotWire, the user's position around the track would need to be monitored to know where contacts are being made and what causes them. This would show if the contacts are primarily caused by difficult sections on the track, or from the interruption task or interaction device used. Furthermore, the algorithm in the mediated method was able to demonstrate benefits despite being trivial. It would therefore be interesting to evaluate different algorithms for this kind of context awareness, that through very simple means can be applied in real life scenarios and still have a positive effect.

8. Acknowledgments

This work has been partly funded by the European Commission through IST Project wearIT@work (No. IP 004216-2004) and also partly funded by the Centre for Distance-spanning Healthcare and the Centre for Distance-spanning Technology at Luleå University of Technology. We thank Dr. McFarlane for providing us with the source code to his original experiments and giving us permission to modify it for our studies.

References

[1] M. Boronowsky, T. Nicolai, C. Schlieder, and A. Schmidt. Winspect: A case study for wearable computing-supported inspection tasks. In *5th IEEE International Symposium on Wearable Computers (ISWC'01)*, 2001.

[2] M. Drugge, M. Nilsson, U. Liljedahl, K. Synnes, and P. Parnes. Methods for Interrupting a Wearable Computer User. In *Proceedings of the 8th IEEE International Symposium on Wearable Computers (ISWC'04)*, November 2004.

[3] M. W. Eysenck and M. T. Keane. *Cognitive Psychology: A Student's Handbook*. Psychology Press (UK), 5th edition, 2005.

[4] N. Kern, S. Antifakos, B. Schiele, and A. Schwaninger. A Model for Human Interruptability: Experimental Evaluation and Automatic Estimation from Wearable Sensors. In *Proceedings of the 8th IEEE International Symposium on Wearable Computing (ISWC'04)*, November 2004.

[5] N. Lund. *Attention and Pattern Recognition*. Routledge, East Sussex, UK, 2001.

[6] D. C. McFarlane. Coordinating the interruption of people in human-computer interaction. In *Human-Computer Interaction - INTERACT'99*, pages 295–303. IOS Press, Inc., 1999.

[7] M. Nilsson, M. Drugge, U. Liljedahl, K. Synnes, and P. Parnes. A Study on Users' Preference on Interruption When Using Wearable Computers and Head Mounted Displays. In *Proc. of the 3rd IEEE International Conference on Pervasive Computing and Communications (PerCom'05)*, March 2005.

[8] T. Starner. Attention, memory, and wearable interfaces. *IEEE Pervasive Computing*, 1(4):88–91, 2002.

[9] H. Witt and M. Drugge. Hotwire: An apparatus for simulating primary tasks in wearable computing. In *CHI '06: Extended Abstracts on Human Factors in Computing Systems*, April 2006.