A Study on Users' Preference on Interruption When Using Wearable Computers and Head Mounted Displays

Marcus Nilsson[1], Mikael Drugge[1], Urban Liljedahl[2], Kre Synnes[1], Peter Parnes[1]

Luleå University of Technology
Department of Computer Science & Electrical Engineering
[1]*Division of Media Technology*, [2]*Division of Computer Science and Networking*
SE–971 87 Luleå, Sweden
{*marcus.nilsson, mikael.drugge, urban.liljedahl, kare.synnes, peter.parnes*}*@ltu.se*

Abstract

An important part of making a wearable computer unobtrusive is the user interface and the way it interrupts the user. The amount of interruption that have to be made on the users primary task should be minimized. Usage without interruption is of course best for the performance of the primary task but is often not possible as many tasks that the wearable computer is involved in do need interaction with the user. It is therefore important to understand what type of interruption that will benefit the user most. The question about which interruption that causes the least amount of stress for the user is as important as which interruption that gives the best performance. This paper continues on previous work where performance of the users was measured for different types of interruption. In this paper subjective data is analyzed to understand how to build user interfaces for wearable computers and head mounted displays that considers stress and other subjective variables.

1 Introduction

The Wearable computer is designed to help the user without being obtrusive. People have worked on making it smaller, finding areas for comfortable placement on the body [1] and finding good user interface to name a few examples. How to interrupt the user is an area that is becoming more important in pervasive computing[8] and when working with interfaces for devices that tries to be unobtrusive, such as a wearable computer. The best would of course be if no interruption had to be made at all, but in many cases it will be impossible to achieve results without interrupting the user. It is therefore important to understand how to interrupt a wearable computer user in the best way.

A user could for example be in a chat session with some-one while at the same time being occupied in the real world with another task. In this scenario the interruptions will be expected, as messages in a chat will often arrive within certain timeframes. Another type of interruption can be an important phone call that the user must answer. This kind of interruption will be unexpected and may need the users immediate attention.

The latter kind of notification can be further exemplified by emergency situations in which the user must be notified about an issue and resolve it, yet still be able to continue with doing real world tasks.

For example, a medical doctor at an emergency site or a fire fighter in a disaster area may need to perform their normal work in the real world, but at the same time they must also be kept informed about the progress of other workers and possibly assist with guidance through a wearable computer. Since the user views both of these tasks as important, it is vital to assess how the virtual task can be presented for a user while minimizing interference with her real world task, but at the same time minimize stress and other long run variables that decrease performance.

Furthermore, since the wearable computer is meant to act as an assistant for its user in everyday life, (e.g. as exemplified by the remembrance agent[14] and the shopping jacket[13]), it is important to increase the knowledge on how interruption of users should be done. As wearable computers become more common it is important to develop tools to capture data for usability studies [6]. It is important that these studies not only see to what performance the user achieves with the wearable computer, but also looks at subjective data as what interface and system creates more stress and which system and interface is preferred by the user.

The research question this brings forward is what type of interruption is preferred by users in dual task situations and what type of interruption and notification creates the least stressful environment? There is also a question if there ex-

ist a conflict between the interruption that gives best performance and what users will prefer?

Our main hypothesis is that the type of notification will have a disparate impact on the user's workload, and that performance and stress for the user will be affected differently depending on how she is allowed to handle the interruptions.

This paper builds on earlier work by the authors that is published in [4] where the objective data for a user study was analyzed. This paper takes the subjective data that was gathered in the same user study and analyses it and also compares the results to the results from the objective data in previous work.

The organization of the paper is as follows. Section 2 presents the primary and secondary task used in the study. Section 3 explains the previous work done on objective data. Section 4 discusses the study made and how the subjective data was collected, and section 5 presents the results. Finally, section 6 concludes the paper together with a discussion of future work.

1.1 Related Work

In [10], McFarlane presents the first empirical study of all four known approaches to the problem of how to coordinate user interruption in human-computer interaction and multiple tasks. His study is done with respect to how to interrupt the user within the context of doing computer work without increasing that person's cognitive workload. A more detailed description of this study is given in [9].

The study presented in our paper repeats the experiment done in [10], but focuses on the interruption of a wearable computer user involved in real world tasks. We are thus able to compare the results from both studies to see whether they differ and how the user is affected by performing the tasks in a wearable computing scenario.

In [5], the use of sensors in order to determine human interruptibility is presented. While this is most certainly useful and would be highly valuable to have in a wearable computer environment, our study instead focuses on when the interruption is of such importance that it cannot be postponed. That is, regardless of how involved the person is in real world tasks, the interruption must still take place even if that would be intrusive and may affect performance negatively. As an example of when this would occur, imagine having two tasks of equal importance, where one task cannot be put on hold for a very long time at the expense of the other.

In [3] an experiment is presented where a person asks questions to a user playing a game, thereby interrupting him and forcing him to respond before continuing playing. The study shows what happens if the asker is given clues about the user's workload, as that should allow him to ask questions at more appropriate times and withhold them during critical periods in the game. In a wearable computer environment, sending live video and audio streams from the wearable computer user to a person at a remote location could convey this information. However, there are privacy concerns with this approach, and it may also be the case that a person being able to assess the situation does not initiate the interruption — it may be machine initiated or triggered by events beyond human control. For such occasions, we believe interruption will still occur even during critical periods of time, and thus it is still desirable to know what methods of interruption will disturb the user the least.

A related study is Maglio's study of peripheral information [7] where the user's cognitive workload is measured when working on one task while getting unrelated peripheral information. The study does not consider the use of wearable computers, but is interesting as the use of peripheral information could be a good way to notify users of such computers. In contrast to our study, the users did not act on the notification given.

The study made by Brewster [2] shows that sound is important in single tasks when the visual capabilities of the device are restricted. Our study also investigates the effect of sound but in a scenario with dual tasks.

2 The Dual Task

The setup used in this study is taken from [10] and redone to fit in a scenario where the secondary task is presented in a head mounted display which is common on wearable computers. The experiment has two distinct parts, the primary task and the secondary task. The tasks should be of equal importance to the subject. The primary task is constantly running while the secondary task interrupts the user with different type of methods. To decrease the chance that subjects would bias the importance of any of the tasks it was made clear by both by written and spoken instructions that both tasks are of equal importance.

2.1 Primary Task

The experiment has a real world task represented as a trivial yet challenging computer game[1] which the user plays on a laptop computer. The objective of the game is to bounce jumping diplomats on a stretcher three times so that each diplomat lands safely in a truck. A screenshot from the game can be seen in figure 1.

For simplicity, each diplomat jumps and bounces in an identical trajectory so that the stretcher needs only be placed in any of three fixed positions. If the user misses a diplomat that person is lost and cannot be saved. The number of saved

---

[1]Original code by Dr. Daniel C. McFarlane.

IEEE
COMPUTER
SOCIETY

Figure 1. The bouncing diplomats game.

Figure 2. The matching task.

and lost diplomats is recorded during the game in order to get statistics about user performance.

The total number of jumping diplomats in a game is held constant, and they appear randomly throughout the game. As the time for each game is kept constant as well, this randomness means that at times there may be few or no diplomats while at other times there may be several of them that need to be saved. Thus, the user gets a varied task that requires attention and is difficult to perform automatically.

2.2 Secondary Task

The interruption task consists of a matching task[2] shown in the user's semi-transparent head-mounted display. When the task appears, the user is presented with three objects of varied colors and shape as shown in the example screenshot in figure 2. The top object is used as reference and the user is informed by a text in the middle of the screen to match this object with one of the two objects at the base. The matching can be either by color or by shape, and only a single object will match the reference object.

As the color and shape is determined at random, the user should not be able to learn any specific pattern or order in which they will appear. No feedback is given to the user after selecting an object regardless of whether the matching is correct or wrong, in order to avoid additional stress and distraction for the user.

2.3 Treatments

In order to investigate the different methods of interrupting the user, five different treatments were used where each of them tests a certain aspect of the interruption.
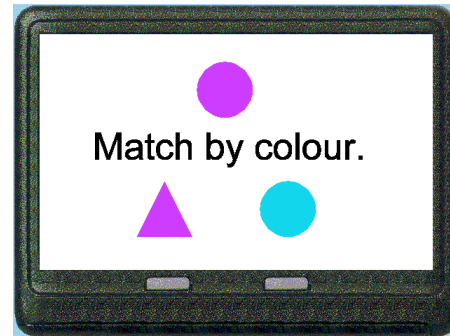
_____

[2]Original code by Dr. Daniel C. McFarlane.

1. **Game only** Control case where only the bouncing diplomats game is played for a given period of time. The user will never be interrupted in this treatment.

2. **Match only** Control case where only the matching task appears at random during a given period of time, the length of it identical to that for *Game only*. The user will not be presented with the bouncing diplomats game during this time.

3. **Negotiated visual** User plays the bouncing diplomats game. Matching tasks are announced visually by flashing a blank matching task for 150 ms in the head-mounted display. The user can choose when to present and respond to it, and also to hide it again e.g. in case of a sudden increase in workload in the game.

4. **Negotiated audible** Identical to *Negotiated visual* but the matching tasks are announced audibly by playing a bell-like sound for about half a second each time a new matching task is added.

5. **Scheduled** User plays the bouncing diplomats game. Matching tasks are accumulated over a period of time and the entire queue is presented at regular intervals. The user cannot negotiate when the matching tasks are presented, and neither can they be hidden once they have appeared. The only way for the user not to have the tasks presented is to respond to every task in the queue, after that there will be no interruption until the next interval round.

It should be noted that in [10], six different treatments were used; in addition to the two control cases (*Game only* and *Match only*) and the *Scheduled* treatment were *Immediate*, *Negotiated* and *Mediated*. Due to the nature of what this study tests those treatments were abandoned or modified because of the following reasons:

- *Immediate* presents the matching task immediately when it appears, forcing the user to respond to it as

the game is replaced with the matching task. However, as the user is involved in real world tasks there is no such enforcement as he can simply choose to ignore the matching task while continuing in the real world. Thus, the treatment is reduced to a variant of *Negotiated*, and therefore it was abandoned.

- *Negotiated* was extended so that an audible announcement was added in addition to the visual announcement, thus splitting up the treatment in the two separate treatments *Negotiated visual* and *Negotiated audible*. These treatments are identical to the original *Negotiated* treatment, with the exception that the game is still playable even when a matching task is present. Some wearable computers can only notify the user through audio[15] and the best type of notification can vary dependent on the context [16, 11], in this case wearable computers. It is therefore important to study if there exists a difference between audio and visual notifications when considering the user's cognitive workload and stress when using wearable computers.

- *Mediated* measured the workload based on the number of diplomats currently being bounced. For real world tasks the workload may depend on numerous factors that can be difficult to take into account outside of a lab environment, so a better approach is then to monitor the user's response to the workload. Since a wearable computer is used, biometric data (e.g. heart and eye blink rate) can be retrieved to derive the user's focus and stress level. However, this is in itself a complex study outside the scope of this paper, and therefore the treatment was abandoned.

The two control cases, *Game only* and *Match only*, provide a baseline for the performance of the user. For the remaining treatments, *Negotiated visual*, *Negotiated audio* and *Scheduled*, they will all interrupt the user and may thereby affect the performance.

3   Previous Work

The measurements made are to see the effect on performance when using different methods for interrupting the wearable computer. The measurements chosen were the same as in [10], in order to allow for an easy comparison between the two sets of results. The result from the objective measurements are examined more in depth in a previous paper [4] and is in this paper covered only briefly for comparison with the subjective data. The objective measurements are;

Diplomats saved. Number of jumping diplomats saved.

Matched wrong. Number of matching tasks answered wrong.

Percent done wrong. Percentage of matching tasks done answered wrong.

Matches not done. Number of matching tasks not answered before treatment ended.

Average match age. Length between onset of matching task until it was responded to.

The original study also measured the number of times the subject changed between game and matching task. However, as the user in our study can switch mentally between tasks without using the keyboard, this measurement is not valid unless other equipment (e.g. gaze tracking) is used.

3.1   Results

The data collected in the treatments where tested for significant differences between the treatments. A multiple measurement ANOVA was made over all the treatments to be sure that significant difference existed between the different groups of data. The ANOVA showed that a significant difference existed between the treatments. A post hoc analyses with paired t-test and Bon Ferroni correction was made between the different treatments in pair to find which had significant difference. The result of these test can be seen in table 1

Table 1. Pair wise t-tests of treatments.

Measurement	Vis. / Aud.	Aud. / Sched.	Sched. / Vis.
Diplomats saved	0.2152	0.4131	0.1952
Matched wrong	0.1256	0.2315	0.0286
Percent done wrong	0.0959	0.3575	0.0464
Matches not done	0.0471	**0.0002**	**<0.0001**
Average match age	0.1258	**<0.0001**	**<0.0001**

The measurements show only a significant difference in matching not done and average match age. This difference can probably be related to the fact that some matching tasks where queued and showed up to late in the end of the treatment so that the subject did not have time to finish all of them.

Even if there is no significant difference in other variables then average match age and matches not done a slightly better result could be seen on the scheduled treatment. A slightly better performance for audible could also be seen between the audible and visual negotiated treatment. More tests must be made before anything can be said about this for sure though.

4 The Study

A test group of 20 subjects were recruited from several sources including a larger testbed in northern sweden called "Testbed Botnia"[3]. Participants in the user study were given a cinema ticket for their help. It was clearly stated that this ticket would be given even if the subject did not fulfill the whole study. The subjects that applied through "Testbed Botnia" also received points that can later be exchanged for discounts on merchandise by "Testbed Botnia".

Upon arrival, a test leader informed each subject about the purpose of the study and how it would be performed. Each treatment was described in general terms, much like the description in section 2.3, but the exact number of diplomats or matching tasks was not disclosed. The instructions for a specific treatment was also repeated in the pause preceding each of them. Pilot studies indicated this repetition was useful as it served to remind the subject of what to expect before proceeding. It also seemed to help in making the atmosphere in the lab environment less strict and not as tense, thereby making the subjects feel more comfortable and willing to comment on the experiment.

Just before the experiment was started the subject put on the head-mounted display. As the display is rather sensitive to the viewing angle, a sample image was shown in the display to help the subject align it properly. The same image was also shown in each pause in the test session so as to give the subject a chance to adjust it further if needed.

4.1 Test Session

The test is a within subjects design with the single factor of different treatments used as independent variable. The participants were randomly divided into 5 groups; in each group, the order in which the treatments were presented differed to avoid bias and learning effects. The order of the treatments in the different groups was chosen to comply with a Latin square distribution.

4.2 Session Length.

The experiment started with the user being able to try and learn every treatment before data was gathered to avoid learning effects. Under the learning treatments the subject was free to ask questions at any time. The learning treatments took about 30 minutes with question and instructions. The time for the treatments was twice as long for the case of being able to gather enough data. This makes the total time for the experiment around 90 minutes

---

[3]http://www.testplats.com

4.3 Apparatus

The apparatus used in the experiment consists of a Dell Latitude C400 laptop with a 12.1" screen, Intel Pentium III 1.2 GHz processor and 1 GB of main memory. Connected to the laptop is a semi-transparent head-mounted display by TekGear called the M2 Personal Viewer providing the user with a monocular full color view in 800x600 resolution. In effect, this head-mounted display gives the appearance of a 14" screen floating about a meter in front of the user's eye. As the display is semi-transparent the user can normally look right through it without problems, but when the interruption task is presented the view with that eye is more or less obscured.

The bouncing diplomats game is shown on the laptop's 12.1" screen in 800x600 resolution, while the matching task is shown in the head-mounted display in 800x600 resolution. The actual screen space taken up by the game and matching task is 640x480 pixels, the rest of the area is colored black.

User input is received through an external keyboard connected to the laptop. In the game, the user moves the stretcher left and right by pressing the left and right arrow keys, respectively. The matching task is controlled by pressing the "Delete" key to select the left object, and "Page Down" to select the right object. In the *Negotiated* treatments, pressing the up arrow presents a matching task under condition the queue is not empty, while pressing the down arrow hides any matching task currently presented. As shown in figure 3, the natural mapping of keys as they appear on an ordinary keyboard should make control fairly intuitive for the user.
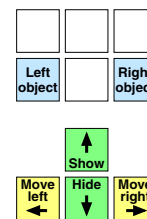

Figure 3. Keys for controlling the tasks.

The laptop was elevated 20 cm over the table so that the subject when sitting down faces it approximately straight ahead. By elevating the laptop the head-mounted display was also more naturally aligned so that the laptop's screen would be covered, this was done intentionally in order to try and force the user to look through the head-mounted display at all time. Although an option is to let the head-mounted display be positioned below or above the user's normal gaze, the enforcement of looking through it was chosen because such situations are assumed to occur in real life with this kind of display. Our pilot studies also indi-

cated the chair and external keyboard allowed the subject to sit comfortably and control the tasks without strain. Figure 4 shows the complete setup.

Figure 4. User study setup.

4.4 Collecting Subjective Data

It is important to compare what the subject felt and thought under the experiment with the objective data of the subjects' performance under the test. Questions about how the user felt under different treatments were therefore given to understand not only what treatment gave best performance but also what treatment made the user less stressed and most comfortable in the dual tasks situation. The questionnaires are taken from [9] and are extended to fit this setup and also translated to Swedish.

4.4.1 Questionnaire Before Treatments

The first questionnaire given before the first treatment was to determine what person the subject is. Demographic question about gender, age and education was given but also questions about computer and gaming experience. Questions were also asked to get the subjects' thoughts about her performance in the areas of dual tasks and interruptions.

Question about colorblindness was also asked, since the matching task depends on being able to match corresponding colors. Two colorblind subjects participated in the study, but they had no problems differentiating between the colors used in the matching task.

4.4.2 Questionnaire After Treatments

The second questionnaire was to get a picture about the subjects more subjective thoughts about the treatments. To understand this the subject were given question about experienced stress and how she liked different part of the experiment. She was also asked to rank the different treatments in order of preferences. The ranking was made dependent on which she liked best for different criteria's, the experienced stress, the feeling of interruption and how easy it was to

foretell interruption. She was also given highly subjective questions, such as which treatment was the least complex one to perform, even though the number of matching tasks and jumping diplomats were kept constant in all treatments.

5 Results

Subjective data was gathered through questionnaires described in section 4.4 but also verbally through comments that the subject expressed under the treatments. The ranking of the subejcts are presented in table 2, where answers are grouped dependent if they choose negotiated treatments before the scheduled treatment or scheduled treatment before the negotiated treatments. The ranking that was made showed that a majority of the subject preferred the scheduled treatment before the notification treatments. A majority alos thought that the scheduled treatment was better for solving the double tasks. As the extra work involved in notification was said to be easy by the subjects the answer to why schedule was preferred must be somewhere else. When looking on the ranking on treatments dependent on how stressed the subject felt we can see a relation to what treatment the subjects' preferred. Subjects felt more stressed by negotiated treatments then by the scheduled treatment even if they felt more interrupted by the schedule treatment. It is also important to point out that the audio notification was selected as giving more stress than visual notification. The opposite was true for the feeling of interruption where subjects thought that visual notification felt more interrupting. A reason for why subjects found themselves more stressed in negotiated treatments can be seen in the ranking of how busy the subject felt when interrupted in the treatments. In this ranking, negotiated and in majority audible notification was selected as the treatment that interrupted when the user felt mostly busy. Scheduled was also selected by a majority to be the treatment that made it most easy to continue the primary task after an interruption.

The subjects were also asked to rank treatments dependent on how distracting the treatment was and small majority thought negotiated was more distracting, the different type of notification was divided equally on first and second place or second and third place when schedule was ranked first.

The preferred treatment when considering only the game task was divided equally on negotiated treatments and scheduled treatments. When considering the matching task negotiated had a small majority.

5.1 Analysing the Groups

The subjects were divided into groups dependent on the ranking that they had made of treatments for different criteria's. Two groups were made for every ranking question and

Table 2. Ranking of treatments

Ranking criteria	Negotiated first	Scheduled first	Other
Liked best	7	10	3
Best for the dual task	8	10	2
Best for the matching task	8	7	5
Best for the game task	8	8	4
Most stressful	9	6	5
Most interrupting	7	8	5
Most distracting	7	6	7
Easiest to predict interruptions	6	10	4
Most occupied when interrupted	9	5	6
Easiest to continue after interruption	7	11	8

were divided between the subject that had selected both negotiated treatments before the scheduled treatment and the subjects who had selected scheduled first. A small fraction of the subjects had ranked the scheduled treatment as second and therefore had negotiated treatments both first and third, these results where discarded in this test as the division into groups was to determine the difference between subjects preferring negotiated treatments and subjects preferring scheduled treatments. The other questions, where users graded different aspect of themselves and the treatments, was then analyzed to see what aspects were different for the separate groups. The result of the analysis between the groups can be found in table 3. Questions that had no significant difference are left out because of space requirements. The question in the table are these:

A How distracting was the visual notification?

B How good are you in typing on keyboard without looking?

C How would you grade your skill with computers?

D How easy was it to predict interruption in the scheduled task?

E Grade your experience in writing on computers?

F Grade your performance in writing on computers?

G How much computer experience do you have?

All question with a significant difference between the groups had a higher average for scheduled treatments then for negotiated treatments.

The question that separates the groups in most cases are if subjects thought the visual notification was distracting. The group with the highest score for the distraction of the visual notification is the group that have selected schedule before notifications. It is interesting to see that audible notification had no significant difference between the groups.

This may have some relations to what was discussed in section 5, that audible notification was selected as the most stressful treatment and therefore was viewed more equally by subjects.

The question about being able to type without looking at the keyboard was a question that had significant difference between the groups when ranking the best treatment for her performance in the matching task. This can be related to that a user with high ability to type without looking at the keyboard, does not need to look down at the keyboard when switching task and is therefore more comfortable with having both tasks shown at the same time.

It is interesting that some of the subjects did indeed choose negotiated treatments before the scheduled treatment as there was no way to predict the interruption in the negotiated treatments. The experience in computers may have helped the subject to understand the treatments better and in that see the possibility to predict the interruption. This is also supported in that there is a significant difference between the groups in the question how easy it was to predict interruptions in the scheduled treatment. The other two questions that have significant difference between the groups are about typing on the computer and are probably related to the difference in computer skill which is also have significant difference.

The significant difference in computer experience when looking at the groups for the ranking of treatments after how easy it was to continue after a treatment can have relation to the fact that working with computer today often incorporates interruptions with instant messages and e-mail arriving. The subject with higher computer skill and experience may therefore have liked the possibility to oversee both task at the same time as was possible in the scheduled treatment.

5.2 Subjective Comments

Three subjects reported that the use of sound in *Negotiated audible* lost its meaning when it was played at the same time as a diplomat was bounced. The sound was merely in-

Table 3. A table over the questions that had significant difference between subjects who preferred negotiated treatments and those who preferred scheduled treatment for different criteria

Question	A	B	C	D	E	F	G
Best for dual task	0,0178	-	-	-	-	-	-
Best for Matching	0,0140	0,0404	-	-	-	-	-
Most stressful	0,0231	-	-	-	-	-	-
Most interrupted	0,0072	-	-	-	-	-	-
Most distracting	0,0214	-	-	-	-	-	-
Easiest to predict	0,0317	-	0,0393	0,0248	0,0465	0,0101	-
Easiest to continue	0,0195	-	0,0046	-	-	-	0,0112

terpreted as a "bouncing sound" and not as an indication that there was a new matching task to perform, even though participants were fully aware of the actual meaning of the sound. This suggests that for certain tasks, care must be taken not to let the sound coincide and relate to the task — especially if the two tasks are meant to be disjoint.

Two subjects reported that hearing a sound was more difficult to relate to in a temporal sense compared to seeing a visual flash. At times the subjects made an attempt to show the matching tasks, only to realize that no new tasks had been added. Apparently the chronological order of when a sound is played can be more difficult to determine compared to when a visual flash is shown, at least when the task to be informed about is also done in the visual domain. Whether the same situation would occur for a task in the audible domain remains an open question.

Subjects also said that the scheduled treatment was much better as no extra work had to be done to see the secondary task. Some also suggested that it would be an improvement to take away the scheduling which would create the treatment immediate which further suggest that immediate should have been part of the study.

5.3 Comparing Subjective Data with Objective Data

The slightly better performance of scheduled treatment together with the fact that subjects liked this treatment better suggest that the decision to take away the immediate treatment may not have been right. This is because these both treatments share the characteristics that the user does not need to do anything to show the secondary task. This was expressed by some subjects as the reason for selecting scheduled as the best treatment which was commented on in section 5.2.

5.3.1 Impact on User Interfaces for Wearable Computers

When considering new user interfaces for wearable computers that uses interruption it can be tempting to use the one which shows best performance of for users from an objective perspective. Results from this study of subjective data suggest though that such an approach may lead to an interface that is less accepted by the user. Instead it is important to test not only the performance but also gather subjective data and evaluate it.

Both subjective data and objective data suggest that when designing new user interfaces for wearable computers it may be better to present the secondary task directly instead of having some kind of notification. This may suggest that the wearable computer today is intrusive in itself and notifications only add to this intrusiveness. In that case better equipment in the future may show different results than this study has shown. The result may instead show that wearable computers and see through head mounted displays do not make the transparent secondary task that is shown feel obtrusive and can be ignored if the primary task is at the moment intensive. The extra work with negotiated treatments is therefore only extra work and should be avoided. But both the subjective and objective data suggest a difference in type of notification and other notification types may therefore show better performance and/or be preferred by users.

No significant difference could be found in computer experience but in the groups for predicting interruption and the ease of continuing the primary task after interruption. This suggest that interfaces with interruption for wearable computers does not need to focus, at least for interruption, on the computer experience of the group of users. As wearable computers have different characteristics then ordinary computers this may be shown to be different when testing on more experienced wearable computer users.

6 Conclusions

In this paper an analysis is shown of the subjective data from a study of different methods for interrupting a wearable computer user. The subjective data is then compared to the objective data gathered in the user study. A slightly better result was present for scheduled treatment in all measurements but average age and matches not done. Subjective data also showed that subjects preferred the scheduled treatment. Negotiated treatment had a better performance overall but subjective data suggest that much care have to be taken into not creating a stressful environment.

It was suggested that an audible notification increased the performance of the matching tasks, while at the same time not affecting the game task negatively compared to the visual treatment. However, as it was shown by the subjective data, audible notification was more stressful, suggesting that audible may not be the best in the long run.

When considering new user interfaces for wearable computers that works with interruption it is therefore important to understand that there can be a conflict between best performance and subjective considerations. Notification should be considered with caution if the secondary task does not need very quickly attention by the user.

All in all, this indicates that both main hypotheses posed in the introduction are true; a user's performance and workload is affected by how interruptions are allowed to be handled, and the type of notification used will have a further impact.

6.1 Future Work

The studies have shown that type of notification have an impact on the stress that the user feel and also the performance of the user. More studies need to be done to see if some type of notification can give a less stressful environment but still have a better overall performance.

It would also be interesting to see if an immediate approach would give better result than the other treatments.

To repeat a similar experiment in a real world scenario, would be good to see how the results gathered in this paper would affect users when not being in a lab environment.

Interruption in social context, where the wearable computer user is in conversation with other collocated persons, have been shown to be an important area [12] and is something that should be studied more thoroughly as this is a situation that will be common for the wearable computer user in the future.

In the experimental setup, the subjects were enforced to look through the head-mounted display. An alternative is to have the display placed to either side, above or below the subject's normal gaze. By not obscuring the game it should be easier to selectively focus on either task, but on the other hand that may make one task easier to ignore.

7 Acknowledgments

References

[1] Kerry Bodine and Francine Gemperle. Effects of functionality on perceived comfort of wearables. In *Seventh IEEE International Symposium on Wearable Computers*, 2003.

[2] S.A Brewster. Sound in the interface to a mobile computer. In *HCI International'99*, pages 43–47, 1999.

[3] Laura Dabbish and Robert Kraut. Coordinating communication: Awareness displays and interruption. In *CHI 2003 Workshop: Providing Elegant Peripheral Awareness*, 2003.

[4] Mikael Drugge, Marcus Nilsson, Urban Liljedahl, Kåre Synnes, and Peter Parnes. Methods for interrupting a wearable computer user. In *Eighth IEEE International Symposium on Wearable Computers*, 2004.

[5] Scott E. Hudson, James Fogarty, Christopher G. Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny C. Lee, and Jie Yang. Predicting human interruptibility with sensors: A wizard of oz feasibility study. In *Proceedings of Conference on Human Factors in Computing Systems (CHI 2003)*, pages 257–264. ACM Press, 2003.

[6] Kent Lyons and Thad Starner. Mobile capture for wearable computer usability testing. In *Proceedings of IEEE International Symposium on Wearable Computing (ISWC 2001)*, 2001.

[7] Paul P. Maglio and Christopher S. Campbell. Trade-offs in displaying peripheral information. In *CHI*, pages 241–248, 2000.

[8] D. Scott McCrickard, C. M. Chewar, Jacob P. Somervell, and Ali Ndiwalana. A model for notification systems evaluation–assessing user goals for multitasking activity. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 10(4):312–338, December 2003.

[9] Daniel C. McFarlane. Interruption of people in human-computer interaction, 1998. Doctoral Dissertation. George Washington University, Washington DC.

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

[11] Maxim Moldenhauer and D. Scott McCrickard. Effect of information modality on geographic cognition in car navigation systems. In *Proceedings of the IFIP TC.13 Conference on Human-Computer Interaction (INTERACT '03)*, September 2003.

[12] Marcus Nilssson, Mikael Drugge, and Peter Parnes. Sharing experience and knowledge with wearable computers. In *Workshop on Memory and Sharing of Experiences*, April 2004.

[13] Cliff Randell and Henk Muller. The shopping jacket: Wearable computing for the consumer. In Peter Thomas, editor, *Personal Technologies vol.4 no.4*, pages 241–244. Springer, 2000.

[14] Bradley J. Rhodes. The wearable remembrance agent: A system for augmented memory. In *Proceedings of 1st International Symposium on Wearable Computers (ISWC'97)*, 1997.

[15] Nitin Sawhney and Chris Schmandt. Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Computer-Human Interaction*, 7(3):353–383, 2000.

[16] Patrik Werle and Carl Gustaf Jansson. Achieving non-intrusive notifications in a ubiquitous service environment. In *RVK 02, RadioVetenskap och Kommunikation, [i. e. Radio Science and Communication.]*, June 2002.