

Methods for Interrupting a Wearable Computer User

Mikael Drugge¹, Marcus Nilsson¹, Urban Liljedahl², Kåre Synnes¹, Peter Parnes¹
Luleå University of Technology

Department of Computer Science & Electrical Engineering

¹*Division of Media Technology, ²Division of Computer Science and Networking*
SE-971 87 Luleå, Sweden

{mikael.drugge, marcus.nilsson, urban.liljedahl, kare.synnes, peter.parnes}@ltu.se

Abstract

A wearable computer equipped with a head-mounted display allows its user to receive notifications and advice that is readily visible in her field of view. While needless interruption of the user should be avoided, there are times when the information is of such importance that it must demand the user's attention. As the user is mobile and likely interacts with the real world when these situations occur, it is important to know in what way the user can be notified without increasing her cognitive workload more than necessary. To investigate ways of presenting information without increasing the cognitive workload of the recipient, an experiment was performed testing different approaches. The experiment described in this paper is based on an existing study of interruption of people in human-computer interaction, but our focus is instead on finding out how this applies to wearable computer users engaged in real world tasks.

1. Introduction

As time goes by, wearable computers can be made smaller, increasingly powerful and more convenient to carry. When such a computer is network enabled within a pervasive computing environment, its user is able to access a wide range of information while at the same time allowing herself to be notified over the network. Such notification can either be expected like in a conversation, or it can come unexpectedly in which the recipient has no way of anticipating the information — neither its content nor its time of arrival. While interrupting the user needlessly should be avoided in general, this latter kind of notification can be exemplified by emergency situations in which the user must be notified about an issue and resolve it, yet still be able to continue functioning in 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 both of these tasks are viewed as important by the user, it is vital to assess how the virtual task can be presented for a user while minimizing interference with her real world task.

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[9] and the shopping jacket[8]), 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 [4]. This should be done so that the future design of wearable computers can go from building complex and specialized hardware to developing user interfaces that support the interaction with the user.

The research question this brings forward is how to interrupt the user of a wearable computer without increasing her cognitive workload more than is absolutely necessary. Considering a wearable computer built out of standard consumer products with basic video and audio capabilities, what ways are there to present information to the user? In what ways can a user be notified that new information exists and needs to be dealt with, and which is the most preferable method for doing so?

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

The organization of the paper is as follows. Section 2 presents the experiment with the tasks and treatments used. Section 3 discusses the method used for conducting the experiments, and section 4 presents the results. Finally, section 5 concludes the paper together with a discussion of future work.

1.1. Related Work

In [7], 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 on 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 [6].

The study presented in our paper repeats the experiment done in [7], but focuses instead 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 [3], 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 [2] 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, this information could be conveyed by sending live video and audio streams from the wearable computer user to a person at a remote location. However, there are privacy concerns with this approach, and it may also be the case that the interruption is not initiated by a person being able to assess the situation — 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[5] 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[1] 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. Experiment

The experiment addresses how different methods of interrupting the user of a wearable computer will affect that person's cognitive workload. The interruption in this case originates from the wearable computer and calls for the user to interact and then carry on with the real world task as before. In order to measure the user's performance in both types of tasks, these must be represented in an experimental model. This section describes the general idea of each task and how they are combined in the experiment, the setup is based on that used in [7].

2.1. Real World Task

The experiment has a real world task represented as a trivial yet challenging computer game¹ 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.



Figure 1. The bouncing diplomats game.

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

¹ Original code by Dr. Daniel C. McFarlane.

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. Interruption Task

The interruption task consists of a matching task² shown in the user's semi-transparent head-mounted display. When the task appears, the user is presented with three objects of varied colour 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 colour or by shape, and only a single object will match the reference object.

As the colour 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.

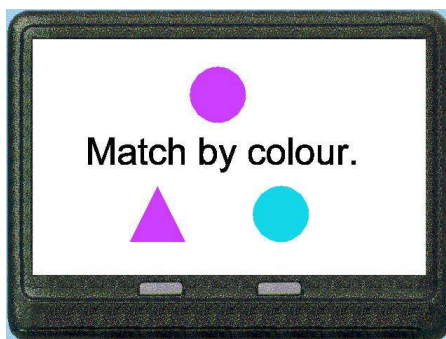


Figure 2. The matching task.

2.3. Combining the Tasks

While the user is playing the bouncing diplomats game, he will be interrupted by matching tasks appearing at random intervals. The tasks are either presented without user intervention or announced by use of visual or audible notification. For the announced tasks, the user negotiates and decides when to present them. When a task is shown, the user may choose to respond to it by selecting an object or ignore it while continuing with the game. If the task is not

handled fast enough, new matching tasks will be added to a queue (hidden from the user) which must eventually be taken care of.

To prevent the user from deliberately ignoring the interruption task throughout the entire game, the user is informed in advance that both tasks are of equal importance from an experimental standpoint. Although personal opinions about the importance of tasks may differ — e.g. saving the jumping diplomats may be perceived as being more important than matching objects — pilot testing did not reveal any such bias in our case.

2.4. 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.

- 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 can not 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 [7], 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:

² Original code by Dr. Daniel C. McFarlane.

- *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. Since some wearable computers can only notify the user through audio[10], it is important to see if there exists a difference between audio and visual notifications when considering the user's cognitive workload.
- *Mediated* measured the workload based on the number of diplomats currently being bounced. For real world tasks the workload may depend on numerous factors which 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. User Study

A total number of 20 subjects were recruited among students and a larger testbed called "Testbed Botnia" (<http://www.testplats.com>) where the user study was announced together with a set of questions. Individuals wishing to partake in the study responded to the questions to express their interest. Based on their answers, a heterogeneous group of 16 males and 4 females aged between 12 and 39 years were selected for participation. As members of the testbed the participants receive points for each study they partake in and can later exchange those points for merchandise. Due to the test session's length of 90 minutes, they were also given a cinema ticket as compensation for their participation in the study. They were also informed they would receive this ticket unconditionally

even if not completing the full study for some reason.

Upon arrival, each subject was informed by a test leader 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 2.4, but the exact number of diplomats or matching tasks was not disclosed. The instructions for a specific treatment were 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.

Before the test, the subject was asked to fill in a questionnaire with general questions about their computer skill and ability to work under stress. Demographic questions about their age, gender, education and whether they were color blind were also given; the latter being relevant since the matching task depends on being able to match corresponding colours. Two colour blind subjects participated in the study, but they had no problems differentiating between the colours used in the matching task.

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.

After the test, the subject filled in another questionnaire with questions about the test, e.g. how they had experienced the treatments and their rating of them in order of preference. They were also given highly subjective questions, such as which treatment (excluding the control cases) was the least complex one to perform, even though the number of matching tasks and jumping diplomats were kept constant in all treatments.

3.1. 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.

The test session consists of each round of treatments being done twice; one practice round and one experimental round. During the first round the subject is given a chance to learn about the five treatments — the data from this round is not included in the final results. At the end of the practice

round, each subject is sufficiently trained for the experimental round; here the five treatments are done once more but this time the data will be included in the final results.

Session Length. Pilot studies indicated that subject learning had stabilized after about 4.5 minutes, so during the first round each treatment was done only once. Even though learning stabilized early, the subjects were still required to practice each of the five treatments in order to learn them in detail. The total effective length of a treatment is 4.5 minutes, when including the pause the actual length becomes about 5 minutes. The practice round with five treatments thus takes 25 minutes to complete; adding 5 more minutes for questions makes the practice round take about 30 minutes in total.

In the experimental round, each treatment is done twice so as to get enough statistically valid data. Each treatment is divided in two with a short pause in between to give the user time to relax and get rid of fatigue. Thus, each treatment takes $2 * 4.5 = 9$ minutes to complete, with pauses included the time is about 10 minutes in total. The experimental round will thus take 50 minutes to complete all five treatments. Adding 10 minutes for the subject to be instructed and fill in the questionnaires before and after the test makes the entire session take about 90 minutes to complete.

Number of Diplomats and Matching Tasks. During the practice round a total of 38 jumping diplomats and 40 matching tasks were used per treatment. In the experimental round, these numbers were raised to 59 diplomats and 80 matching tasks per treatment. The numbers were chosen to be the same as in [7] to allow for direct comparisons between the studies. None of the subjects expressed any negative opinion about this increase; on the contrary it seemed the added difficulty served as extra motivation.

3.2. 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 colour 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 coloured 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.

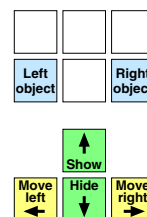


Figure 3. Keys for controlling the tasks.

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 indicated 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. Results

The measurements chosen were the same as in [7], in order to allow for an easy comparison between the two sets of results. The graphs in figure 5 show the average value, together with one standard error, of the measurements below.

Diplomats saved. Number of jumping diplomats saved.

Matched wrong. Number of matchings 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.

When doing measurements on the same variables and the same subject under different conditions it is important to accommodate for this in the analysis. A repeated measures ANOVA was therefore used on the data to see if any significant differences were present between the treatments. The results of these tests can be seen in table 1, indicating that the means for the measurements are not all equal.

| Measurement | P-value |
|--------------------|---------|
| Diplomats saved | <0.0001 |
| Matched wrong | 0.0022 |
| Percent done wrong | 0.0014 |
| Matches not done | 0.0003 |
| Average match age | <0.0001 |

Table 1. Repeated measures ANOVA.

4.1. Comparison with Base Cases

When performing a post-hoc statistical paired samples t-test comparing the two base case treatments, *Game only* and *Match only*, with the remaining three treatments, a number of significant differences were shown to exist. This asserts the assumption that interrupting the user will have a detrimental effect on that person's performance. In table 2, a summary of these comparisons is shown, indicating whether there is a significant difference between the base cases and treatments. To accommodate for multiple comparisons, a Bonferroni adjusted alpha value of 0.008 (0.05/6) is used when testing for significance.

The only measurements which were not significantly different from the base case was "Matches not done" for the

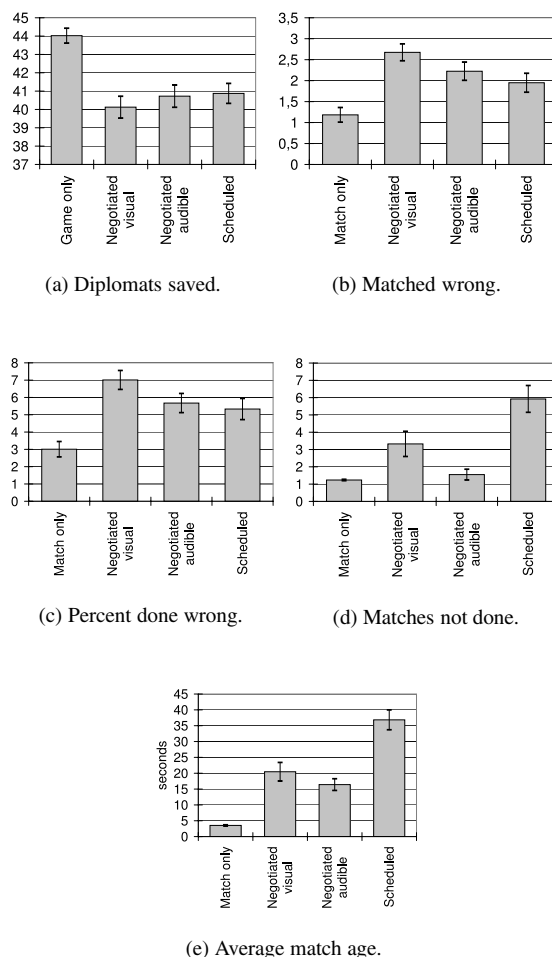


Figure 5. Average measurements.

two *Negotiated* treatments, and "Matched wrong" together with "Percent matched wrong" for the *Scheduled* treatment. The reason for the former is that the subjects often completed roughly the same number of matching tasks as in the base case treatment. This suggests that allowing subjects to negotiate when to present the matching task does not cause it to be omitted more than what would have been the case had the matching task been the only task present. The latter indicates that in *Scheduled*, the subject can better concentrate on the matching tasks. The significant difference for "Matches not done" compared to the *Scheduled* treatment is most likely caused by matching tasks being queued but not presented before the treatment is over.

4.2. Pairwise Comparison of Treatments

The three treatments *Negotiated visual*, *Negotiated audible* and *Scheduled* were compared to each other using a

| Measurement | Base case | Vis. | Aud. | Sched. |
|--------------------|--------------|---------|--------|---------|
| Diplomats saved | <i>Game</i> | <0.0001 | 0.0013 | 0.0012 |
| Matched wrong | <i>Match</i> | 0.0021 | 0.0014 | 0.0671 |
| Percent done wrong | <i>Match</i> | 0.0011 | 0.0013 | 0.0406 |
| Matches not done | <i>Match</i> | 0.1408 | 0.4189 | 0.0072 |
| Average match age | <i>Match</i> | 0.0074 | 0.0020 | <0.0001 |

Table 2. T-tests of base cases vs. treatments.

paired samples t-test. Table 3 shows a summary of this indicating whether a significant difference exists between each pair of treatments. A Bonferroni corrected alpha value of 0.008 is used when testing for significance.

| 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 |

Table 3. Pairwise t-tests of treatments.

As shown in table 3, there were no significant differences in terms of diplomats saved or matching tasks done answered wrong. This means that our test was not sensitive enough to uncover any differences, if such exists, between the treatments for these measurements. However, the “Average match age” measurement is significantly different between the *Scheduled* and the two *Negotiated* treatments. For the two *Negotiated* treatments, the difference is not significant enough ($p = 0.1258$). Nevertheless, the performance of certain subjects together with their comments indicate that there may still be an underlying difference that was not fully uncovered by our study. When relating to what is shown in the graph in figure 5(e); the average age of a matching task is less for *Negotiated audible* than for *Negotiated visual*. Thus, the use of sound may be a stronger reminder that there are matching tasks to perform, compared to using a visual signal. Furthermore, the graph in figure 5(d) shows that the number of tasks not done is also less for *Negotiated audible* than for *Negotiated visual*. While it is not marked as significant in the table ($p = 0.0471$), it still suggests that a difference may exist. This strengthens the indication that sound can have a higher impact on subjects when it comes to reminding them to perform the matching tasks.

As an audible announcement seems to be stronger than a visual one, it is of interest to know how this affects the number of diplomats saved. Referring to the graph in figure 5(a), there is a minor advantage of audio over visual with nearly one more diplomat saved in *Negotiated audible*, but this difference is not significant enough ($p = 0.2152$) to draw any conclusions from. Also, referring to the graphs in

figures 5(b) and 5(c) shows an advantage of audio over visual when it comes to reducing the number and percentage of matching tasks answered wrong, but these are also not significant enough ($p = 0.1256$, $p = 0.0959$). Further studies are needed to see whether the advantage of audible over visual announcements have a positive effect also for these.

The *Scheduled* treatment left significantly more matching tasks undone at the end of a treatment compared to the negotiated treatments. The reason is that when tasks are presented just before the end of the treatment, a large number of them may be in the queue and are not answered before the time runs out. The other measurements were, however, better in *Scheduled* than in the negotiated treatments. This suggests that our decision to skip the *Immediate* condition was erroneous, and that it is likely to have exhibited the benefits of *Scheduled* without the drawback of high average age.

4.3. Comparison with Original Study

In general, the subjects’ in our study scored better results than in in the original study[7]. This is most likely caused by the different setting in which our study was done; as the two tasks could run simultaneously without the matching task blocking input for the game, the user could quickly switch mentally back and forth between them. The user could answer the matching tasks while still seeing the game in the background, and could thus more easily detect when the game task needed attention. The number of diplomats saved was around 10% higher for *Game only*, and one third higher for our two *Negotiated* treatments. This did not, however, affect the matching task negatively; the number of tasks answered wrong was around 45–55% less, suggesting our setup was less prone to leave subjects making wrong decisions. The number of tasks not done was 40–72% less for both negotiated treatments in our study, while in *Scheduled* it was 56% higher. Likely the subjects in the original study were more cautious to switch to the matching task, while in *Scheduled* they had to finish answering them before proceeding with the game. In our study they could switch freely between the dual tasks, explaining this difference. Our average match age was 2 seconds higher for *Match only*, 5 seconds higher for *Negotiated visual*, yet only 1 second higher for *Negotiated audible*. For *Scheduled*, the average age was 26 seconds higher since the subjects could still play the game while the queue of tasks was present.

Audio notification was not used in the original study, but appeared to give a slightly better result than for visual, suggesting that the type of notification can be significant.

4.4. Subjective Comments

In addition to the quantitative data presented in the previous sections, there is also some qualitative data of inter-

est. This data was given either by word of mouth or as written comments in the questionnaires the subjects filled in.

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 interpreted 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.

5. Conclusions

We have presented a study investigating the interruption of a wearable computer user, some of the methods to achieve this and what effects they will have on the user. The results indicate that the scheduled treatment gave the best results, with the drawback of a considerably higher average age before tasks were answered. The negotiated treatments, where the user could decide when to handle the interruptions, were more useful when considering the overall performance of the user; they had a much shorter average task age with only slightly worse performance compared to the scheduled treatment. 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, a more detailed study is required to assert the significance of this observation. All in all, this indicates that both hypotheses posed in the introduction are true; a user’s performance is affected by how interruptions are allowed to be handled, and the type of notification used will have a further impact.

5.1. Future Work

As the user has no direct feedback about the number of interruption tasks currently in the queue, it may be interesting to investigate how such feedback would affect the results. Would the user appreciate seeing this number to plan ahead or would it merely have a detrimental effect?

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.

6. Acknowledgments

This work was funded by the Centre for Distance-spanning Technology (CDT) under the VINNOVA RadioSphere and VITAL Mål-1 project, and by the Centre for Distance-spanning Health care (CDH). Original code by Dr. Daniel C. McFarlane (mcfarlane@acm.org), developed at the Naval Research Laboratory’s Navy Center for Applied Research in Artificial Intelligence (<http://www.aic.nrl.navy.mil>), Washington DC under sponsorship from Dr. James Ballas (ballas@itd.nrl.navy.mil). We thank Dr. McFarlane for providing us with the source code for the game and matching task and giving us permission to modify them for our study. We thank Dr. David Carr as well as the anonymous reviewers for insightful comments and advice given. The authors finally wish to thank all the volunteers who participated in our study.

References

- [1] S. Brewster. Sound in the interface to a mobile computer. In *HCI International’99*, pages 43–47, 1999.
- [2] L. Dabbish and R. Kraut. Coordinating communication: Awareness displays and interruption. In *CHI 2003 Workshop: Providing Elegant Peripheral Awareness*, 2003.
- [3] S. E. Hudson, J. Fogarty, C. G. Atkeson, D. Avrahami, J. Forlizzi, S. Kiesler, J. C. Lee, and J. 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.
- [4] K. Lyons and T. Starner. Mobile capture for wearable computer usability testing. In *Proceedings of IEEE International Symposium on Wearable Computing (ISWC 2001)*, 2001.
- [5] P. P. Maglio and C. S. Campbell. Tradeoffs in displaying peripheral information. In *CHI*, pages 241–248, 2000.
- [6] D. C. McFarlane. Interruption of people in human-computer interaction, 1998. Doctoral Dissertation. George Washington University, Washington DC.
- [7] 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.
- [8] C. Randell and H. Muller. The shopping jacket: Wearable computing for the consumer. In P. Thomas, editor, *Personal Technologies vol.4 no.4*, pages 241–244. Springer, 2000.
- [9] B. J. Rhodes. The wearable remembrance agent: A system for augmented memory. In *Proceedings of 1st International Symposium on Wearable Computers (ISWC’97)*, 1997.
- [10] N. Sawhney and C. Schmandt. Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Computer-Human Interaction*, 7(3):353–383, 2000.