

Position Paper: Evaluation in the FEEL Project, a Pilot Study

Peter Lönnqvist, Hillevi Sundholm, Fredrik Kilander & Patrik Werle

The FUSE group
Department of Computer and Systems Sciences
Stockholm University/KTH (The Royal Institute of Technology)
Kista, Sweden
{peterl, hillevi, fk, werle@dsv.su.se}

Introduction

The FEEL project [1] (FEEL IST-2000-26135) is part of the EU-funded proactive initiative “The Disappearing Computer” [2].

In the FEEL project, evaluation of notification modalities has been conducted in the iLounge at KTH. The acronym somewhat clumsily stands for: Non-intrusive services to support Focussed, Efficient and Enjoyable Local activities.

The main objective of the project is to deal with the problem of the intrusiveness of today's mobile and other technology, and how introducing the idea of non-intrusive services can enhance work in local environments realised partly by disappearing computer environments. The work in the project is based on the detailed analysis of a number of concrete scenarios where today technologies for mobile and distributed work. Based on the analyses of the concrete scenarios we will:

- design appropriate disappearing computer environments
- design mechanisms by which multiple services simultaneously can share and effectively use the disappearing computer environment
- build on the software agent paradigm for service design (for the above purpose)
- establish a software platform which handles software components on small artefacts transparently and manages open sets of services
- make usability studies for the disappearing computer environments

Method

The iLounge is a ubiquitous computing environment built at KTH in Kista, Sweden during spring 2002. Support for the construction and development has been given by the Wallenberg Foundation. The objective is to evaluate whether notification methods differ in how they disturb the ongoing work in a meeting room. We assume that certain ways of conducting messages to participants disturb the meeting between people in a greater way, i.e. if a person's mobile phone rings when he or she is talking to another person this will interrupt their conversation more than if he or she is notified about that there is a message for instance on a wall screen or other display in the environment. For this particular round of evaluation nothing has been adapted from existing applications to the environment; the applications used were created specifically for these tests.

Both qualitative and quantitative methods are used. Quantitative in the aspect of that the test is time-measured, comparing groups and whether they perform well on the test and if they have noticed their notifications. Qualitative in the way that we video record the participants during the test (with their consent), query them on their impressions from the test session and conduct an informal group interview (which is also video taped) after the session.

The evaluation is a pilot study for a coming study where work groups will use the iLounge during a longer period of time. In this study messages in the notifications were mock-ups in the form of text showed in different ways on different displays. Messages could be "To John: Do not forget to buy home milk, bread and butter!" Time constraints and technological ditto forced us to send arbitrary messages like these since the system does not yet collect real world data from email, phone messages and such. There are also not yet any automated agents running the background tasks of negotiation between messages and forwarding them to the correct displays. All such work was done by hand and through monitoring the ongoing activities in the room.

Quite high costs are associated with this work since it is conducted in a specially developed ubiquitous computing environment. For the tasks participants had gotten personal computers (laptops) and in the room there were two wall screen computers (SmartBoard [3] 72" touch sensitive screen cabinets built into a wall, each with a computer and projector attached), a table computer (flat lying 50" plasma screen with a SmartBoard overlay and computer attached, built into a table), four different video cameras at different angles and two omni-directional microphones sending their respective signals through mixers (we mix the four incoming video signals into one so we can watch from different angles on the same screen), a computer to write this data to disk and three different "backstage" computers to take care of different tasks in the system. There is also one more computer involved for the correction of the answers the participants give to the questions. All this is set up in the iLounge, our working environment where our and the architects goal has been to create a somewhat "magical" room. We want participants to feel that they leave the normal workspace and have another experience here. Much work has been done on lighting, coloring, materials, sound environment, etc.

Procedure/Case Study

Participants were recruited locally among the student population at KTH. These are our future real users of applications and environment so they were found as most appropriate for the task. Since the tests were to be performed very late in the term we also recruited participants among colleagues and friends at the department. Each participant was offered two cinema tickets for their participation and was informed that they were going to take part in a group task that would take approximately one hour and that they were to be video taped during the task. The final number of users for this pilot turned out to be 32 in an age span ranging from 19 to 60. The gender distribution was

m=18, f=14. The groups ranged from two to four since we sometimes had participants not showing up or being unable to slot them correctly into the time schedule.

When a group arrived they were welcomed and instructed to sit down around the table with the built-in touch sensitive screen and were then given instructions. The main task was Raven's Matrices, a non-lingual intelligence test, usually given to single persons. The task is to identify which one of a series of pictures completes another picture, all of which are shown at the same time. In our case our software enabled us to from the background send in all the pictures in one set to the table at which the participants were sitting. The participants could then start sorting the pictures through touching them and moving them around on the table. They could also drag the pictures to icons around the frame of the lying screen and thus instantaneously send them to another of the screens in the room, thus creating more space or a better overview. They were encouraged to do this. When reaching a result they were told to send their correct answer to one supporting computer (also through dropping it on an icon) and then got an answer back whether the it was correct or not. If their answer was incorrect they got their suggestion sent back and if it was right they were told to throw all pictures into the trash and were sent a new set.

The participants, when concurring that they had understood this task, were told that during the session they also would get messages sent to them, all on the same display at any given time. The displays differed from case to case and in one setting we did not send any messages in order to get a null-time, a hypothesized time for conduction the test without any messages disturbing the main task. The participants were told to memorize the messages and that we after the session would ask them what their messages had been. Due to the arbitrariness of the messages and (therefore probably increasing) memory slippage, we concluded that the number of messages should be kept very low (three messages per participant was chosen, which in some cases showed to be too much).

The displays for showing the notifications varied. We used a side-scrolling banner in the top area of the wall-mounted screens or on top of the users' personal terminals (laptops which only had this use in the experiment). This was in some cases done in conjunction with sounds when a message started scrolling by, and in some the displays remained silent.



Figure 1, Banner service displaying a message

In some cases we used a device we call the iClock. A screen that normally looks like an analog clock, but which can fade out the clock display and instead show a message. This was always displayed with a chiming sound. We had five different set ups

(including the mode with no notification), each set up were used by two groups. The last two groups got a mixed mode, where the participants received the notification in different ways.



Figure 2, The iClock first appearing as a normal clock and then displaying a message

The participants were told that there were no constraints in helping each other in noticing the messages but that they would be asked individually on what personal messages only they had gotten. After this the users got any number of test sets of Raven's Matrices to get acquainted with the interaction with table and walls and then the test itself was conducted.

After the fulfillment of the task the users got to have their say about their experience and were given a query where we asked about what their personal messages had been and if they had other comments. We also asked them to grade experience of stress and a number of other issues like their knowledge and use of computers.

What initially can be said about what we have learned from the study is that it to a high degree will influence our forthcoming design of services and modalities for their deployment. The participants have had very much to comment about, most of it in a positive way. Almost all liked much to work in this new kind of environment and also appreciated the way we want notifications to take on a new form to less disturb the ongoing work and communication among people. We still have to do work on the big amount of data collected but at a first-hand view we feel that we are going in the right direction. The main disadvantage with a method like this is mainly just that, the big amount of data that has to be analyzed. Issues that we currently discuss is what focus to take when doing the video analysis, since what we look for has to be constrained in some way to make the task at all possible. We also feel we have learned a lot from doing such a big pilot study like this, since for instance only the management of participants becomes an issue when there are as many users involved as in this case. For our forthcoming work we want to involve future users earlier in the design process. We aim at creating focus groups or the like with the students that will use the room in the future.

References

- [1] <http://www.feelproject.org>
- [2] <http://www.disappearing-computer.net>
- [3] <http://www.smarttech.com>