Impact of Sensory and Collaborative Contexts on Usability Evaluation Methods Resources

Ebba Thora Hvannberg
University of Iceland
Dunhaga 5
107 Reykjavik, Iceland
ebba@hi.is

ABSTRACT
This paper describes the evaluation of two applications which are different in their sensory contexts and collaborative contexts. For each application, five method resources are analysed which were used for their evaluation to see what impact the contexts had on the method resources. The analysis concludes by discussing if the transfer across the contexts was successful. Formalising the transfer between contexts, the paper proposes three types of transfer function genres, generalisation/specialisation, adoption accompanied with adaptation and transfer between domains.

INTRODUCTION
With the ever increasing variety of contexts in which usability evaluation is carried out, researchers and practitioners propose modified or new methods that are tailored to specific contexts such as virtual reality to name one. To aid practitioners in selecting from the vast flora of methods, researchers have proposed frameworks with characteristics that become the basis for the method selection. Examples of such characteristics is the objective of the evaluation, e.g. precision (why), the outcomes and metrics of the evaluation (what), whether the evaluation is summative or formative (when), and which types of methods e.g. questionnaire or observation, to choose (how) [1, 2]. In addition to the above characteristics as a basis for selection, it has been well recognised that the resources available, e.g. people, time and money, may determine the type of evaluation applied [1].

Studying and designing new evaluation methods, researchers have proposed evaluation frameworks that can be used to select a certain method. To further narrow the scope of such frameworks, some have been targeted to specific types such as collaboration. The motivation or the process of transfer of methods within those frameworks is seldom seen. Methods evolve stimulated by some need and experience. Although, transfer has not been investigated extensively, there is clearly interest in learning more about it. For example, Antunes et al. [1] looked at 48 papers on reusable evaluation strategies of collaborative systems and saw some identifiable patterns of transfer between methods. The first is an adaptation of single-user evaluation methods to the context of collaborative systems. The second is an assimilation of methods and tools from other fields beyond technology development, e.g. ethnography. The third pattern identified by Antunes et al. [1] was that while the early methods were targeted for specific measures carried out in controlled situations, latter ones were tailored for more complex and broader contextual concerns.

Woolrych et al. [3] concluded that instead of seeing a method as a fixed entity, it would be more useful to view an evaluation method as having a number of resources being any pre-existing re-usable component of design or evaluation that can contribute to the formation of a method through practice.

This paper aims to analyse how usability evaluations are transferred between contexts with respect to five method resources. The transfer of method resources will be matched against three forms of transfers. The hope is that the study will increase our knowledge on transfer function.

‘OLD’ METHOD-RESOURCES IN NEW CONTEXTS

Two training applications
In this case study we will describe how usability evaluation has been carried out on two applications which are both training software. One is meant to train students in mathematics and the other to train crisis response and management. The former is targeted to blind students and the latter targets workers in several crisis management sectors, from fire fighters, police, rescue teams and medical personnel, responders and commanders, professionals and volunteers.

The mathematics application, Mathematics Cane, is a single user application, but the Crisis Management Training software (CMT) is a collaborative application since the work requires teams to coordinate and communicate on tasks. In both cases, the user population is specific and narrow, defined by their sensual abilities in the former case and work domain in the latter case. What they have in common, besides being training software, is that they apply non-visual user interface technologies which are meant to aid the user in completing the tasks. In
the case of the training software in mathematics for blind students, a haptic peripheral was used augmented with sounds, but in CMT soundscapes and voice communication characterised the virtual environment.

**Method resources**

For the purpose of this comparison, we have chosen to focus on five types of method resources [3]. The first two, participant recruitment and task selection, are generic for user testing. Training objective selection, the third method resource, is specific to training software and is a special case of evaluation criteria. The fourth one is data capture which we think is relevant when investigating innovative user interface technologies. The final method resource on our list is control of experiment:

- Participant recruitment
- Task selection
- Training objective selection (evaluation criteria)
- Data capture
- Control of experiment

Before describing the two application contexts, we will say a few words about the above method resources. Participant recruitment is about how users are invited and selected to take part in the usability evaluations. Invitation can be via e-mail, written letter, through a social network or from another population of users. The recruitment can be obligatory or optional and the selection can be according to criteria which are of varying restrictions.

Appropriate task selection and task description has been long considered vital [4, 5]. What is thought to be most important is that it matches the user group. Another criteria for selection is the scope and the span of the tasks, i.e., whether only a small fraction of the tasks offered by the application is chosen or a large one. The task selection will depend on the overall objective of the experiment. Sometimes, evaluations are carried out to test specific parts of a system which were found to be unsatisfactory in previous evaluations. The task selection will determine the length of the experiment and hence depend on how much time participants can devote to the experiment.

Whereas task selection results in a medium-grained set of tasks that users are instructed to carry out, applications can have an overall objective, as in the case of training software to train students in a skill or a competency. In contrast to evaluating if a user can complete a task it is desirable to see if he/she has achieved the objective of having this skill. This overall objective may or may not be considered a part of usability evaluation, but since evaluations of usability usually include effectiveness, it is natural to consider it.

Data capture means any tool or method with which we seize data from the usability evaluations to extract qualitative or quantitative data. What data is captured depends on the instruments applied, e.g., for heuristics evaluation, the main objective is to uncover problems. Data is captured manually by the expert but sometimes screen is captured of the execution of the task automatically so that the expert or a second expert may review the data and verify or carry out an independent review. Several other data capturing methods exists such logging of events, eye tracking and capturing physiological metrics.

Above, we discussed task selection and granularity of tasks. A related method resource is how these tasks are put in front of the users and what freedom they have to carry them out. In a tightly controlled experiment the tasks may be fine grained and put in front of the users in a systematic manner, but in a loosely controlled experiment, the users will be given overall problems which they are to solve. Thus, we note that the individual resources are not discrete but coupled together make up an integrated approach for an evaluation to take place.

**Contexts**

There are different contexts in which design and evaluations are carried out. These contexts are determined by diverse stakeholders, development methods, technologies, maturity of companies, access to users, bespoke or off-the-shelf software, to name a few.

We have chosen to describe two contexts which show the variability between the two applications under consideration (see Table 1). The contexts are the sensory context and the solo/collaborative context. The sensory context of the Mathematics application was haptic, sound and voice synthesis to accommodate blind students, and the sensory context of the CMT application was graphics, sound to accommodate navigation, situation awareness and reality of the noisy environment, and voice. The solo/collaborative context of the Mathematics cane was solo and collaborative for the CMT.

**Description of Original Application Context: Single User Training Software for Mathematics with Haptics, Sound and Voice Synthesis**

The Mathematics Cane is a software application developed by a master student,

**Table 1 Two applications characterised by the sensory context and the solo/collaborative context**

<table>
<thead>
<tr>
<th>Application</th>
<th>Sensory Context</th>
<th>Solo/Collaborative Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics</td>
<td>Haptic, Sound, Voice Synthesis</td>
<td>Solo</td>
</tr>
<tr>
<td>Crisis Management</td>
<td>Graphic, Sound, Voice</td>
<td>Collaborative</td>
</tr>
</tbody>
</table>
Magni Bör Birgisson [6], aimed to help young blind students to learn about 2D polynomials. It accepts polynomial equations and allows students, applying a haptic peripheral, to feel the shape of the polynomial. In addition to the kinaesthetic aid, a student can choose to have sound added when navigating the polynomials with the haptic peripheral. Furthermore, a grid can be superimposed with an extra force (like a tic) felt when passing over the grid crossings. The grid is to make it easier for the blind students to count as they examined the polynomials with the haptic peripheral. A research study was carried out to see which modes, haptic, sound or voice synthesis helped the students. In the remainder of this section, we will address each of the five method resources, starting with participant recruitment.

Participants were recruited from a group of blind and seeing students, six in each group, age 15-25. Since the user population of blind students in this age range is small, seeing students were included, but blindfolded. Furthermore, the aim was to see if there was any qualitative difference between the seeing and the blind students. There was no minimum threshold on mathematics competency, but in retrospect this may have been advisable.

For the tasks, four polynomials were selected with a variability of difficulty, from easy polynomials to challenging ones. Each polynomial was presented to users with and without sounds and grids with the objective to explore the different mode’s effectiveness.

The training objective of the study was to assess the students’ perception on polynomials by asking them to compare the haptically drawn polynomials to embossed ones on paper (3 options given). Hence, only a vague training objective was selected. The evaluation criteria of the research study included research questions on the usability of different sounds, haptic and voice while exploring the polynomials and if the Mathematics cane could be an improvement beyond the current technology.

Data was captured in several ways. An observer noted the time it took users to complete the tasks, intermediate steps of the tasks, such as when the students started to look at the embossed paper for comparison. Right and wrong solutions were recorded, deviations and problems. Using a five point Likert scale, the observer asked participants how easy they thought it was to solve a task, if the sound was useful and the grids. They were asked if they thought that they had solved the tasks efficiently. A software implemented log of users’ navigation through the polynomials was stored. In another research study, software was written to visualize the log from the navigation of the haptic movements [7].

Considering the fifth method resource, i.e. the controllability of the experiments, the research study had the tasks put in front of the user one after another, and thus, it was tightly controlled.

Description of New Application Context(s): Collaborative Training Software for CMT with Sound, Voice and Graphics

A training simulator set in a virtual environment has been built for training crisis management. A prototype of the simulator has been implemented to evaluate realistic soundscapes, voice communication and noise from the far end over the voice communication [8]. The simulator is multi-player, configurable with different crisis scenarios that typically occur in the transport sectors, such as in airports or railways. The objective of the training software is to train responders and commanders in triaging casualties, reporting the results in a command hierarchy, making resources available and transporting casualties to hospitals or shelters as appropriate based on the prioritisation of the injuries. Sounds, generated by objects, such as fire trucks, and chatter are implemented in the virtual environment. Three different communication metaphors, radio, mobile phone and face to face is made available in the simulator. The current prototype is limited to two persons speaking at any time over one of the metaphors. The remainder of the section describes the five method resources, participant recruitment, task selection, evaluation criteria, data capture and experiment control.

Six participants were recruited from a local fire- and rescue company to carry out training using the simulator in three pairs. The participants had been attending a course on simulator training software, but not the same as the one that was evaluated. Obviously, they were not recruited from a large population and they may have been self-selected in a way that the more computer-skilled employees may have shown more interest in the course and the study. There were no additional competency criteria, e.g. novice or experienced, or any particular roles, beyond working and receiving training.

Tasks were selected to evaluate the voice communication, perception to sound, navigation in the virtual environment, situation assessment, presence and flow. A scenario was written with a total of 14 tasks for two roles collaborating, an On Scene Commander and a Rescue Coordinator.

The training objective of the simulator is to train 12 competencies, and three of them were specifically trained in the above scenario: to assess alert, communicate and to communicate information. Seven research questions were posed for the research study, mainly on the effect of sounds, noise and the usability of voice communication using different metaphors.

Data was captured by an observer following each of the two users in the experiment and screen and audio captures, thus storing all graphics, sound and voice
communication. Afterwards, all voice communication was transcribed verbatim. Questions on effectiveness, efficiency and satisfaction were asked in a post-test review, asking these questions for each task. This was different than in the Mathematics cane study since doing it for each task would have broken the flow of the scenario.

The 14 tasks in the scenario were put in front of the user one by one by an observer. Hence the experiment was controlled tightly. This was not so successful and could be attributed to the requirement of letting participants engage in the virtual environment instead of being interrupted by the test facilitator saying what tasks to do next.

**Stories of Transfer: Triumph and Tragedy**

This section describes if the method resources needed to be transformed from one context to another. In an informal study that asked experts about possible transfers of evaluation methods, out of 20 stories we further analysed 12 of them. The outcome of that analysis gave three forms of transfer. First was the Generalisation/Specialisation transfer, where there is a generic method (resource) for a broad set of application but it is desirable to specialize it for a certain context. We have many examples of this transfer function, e.g., for Heuristics Evaluation, where the method resource is a list of Heuristics, e.g. specifically for virtual environments [9, 10].

Another example of transfer is an adoption, often accompanied with some adaptation [11], where a method resource is adopted from a discipline to be used in a new context. An example was: “As usability professional designing interactive and tangible children’s toys, having found laddering from the marketing domain I need to know what aspects need to be changed to apply the method to evaluate the affective responses to specific attributes of the prototypes.”

A third category of transfer is between domains, e.g., from gaming to learning. An example requirement to a method source transfer was “As a teacher (e-learning)designer, I want to explore animation techniques from certain games, so that I can increase learner’s engagement in exploring content in a reflected way.”

In the analysis below, we will see if the transfer of each of the method resources could be attributed to the above three forms of transfer. Each of the two application contexts (old or new) was characterised by two variables, the sensory context and the solo/collaborative context (Table 1). In our model, each variable can have two values. Although the participants in the two evaluation experiments are quite different in their skills and background, the participant recruitment procedure is similar in both cases and neither the sensory nor the collaborative contexts affected the recruitment procedure itself. In both cases, we set some criteria on whom we would select and asked for local help for the recruitment. A contact at a school helped with the recruitment in the case of the Mathematics and a contact at a local fire- and rescue company for the CMT. Hence, in both cases a facilitator close to the users helped with the recruitment. In both cases it was discovered during the tests that a further pre-screening of the participants for certain competencies would have increased their fit to the desired target group. The collaborative context had a slight impact in that in the CMT case we had to recruit two participants for each experiment. Hence, the transfer of this methods resource was successful and required specialization of the participant criteria which was directed by the skills and abilities of users of the target application domain.

The task selection, in both cases, depended on the desired outcome of the evaluation. For the CMT case the design of a scenario consisting of several tasks proved much more challenging since a realistic scenario had to be written covering the usability of the application and the specifics of the sensory channels (i.e. sounds, voice, and graphics) of the application. This transfer, from simple individual tasks of the Mathematics cane to more complex realistic scenarios including sound perception was partly unsuccessful. In the CMT case some users noted that a task was unrealistic and not according to crisis management procedures. Furthermore, it was hard to select tasks that tested the realism and appropriateness of sounds. Scripted training is bound to produce different results than free flowing open, non-scripted training. In particular, non-scripted testing is not repeatable and since participants are working together in the collaborative scenario they will influence each other’s actions [12]. We deem the type of transfer to be an adoption of techniques from other disciplines such as theatre, creativeness and flow [13] [14].

In neither experiment a formal training objective was assessed. The experiment with the Mathematics cane only evaluated if the users could recognise polynomials, but did not evaluate if the software helped the students gain more knowledge of polynomials. Three of the twelve competencies necessary for crisis management were addressed in the latter experiment and were determined by the collaborative contexts. The evaluation criteria for each application were designed for the research study and were heavily influenced by respective contexts. For the CMT and the Mathematics cane the sensory context influenced the evaluation criteria, i.e. sound, noise, voice and haptics but in different ways.

Data capture is the method resource which varies the most across the two contexts. Data was captured with special analytics software in the Mathematics cane but with regular screen capture and audio recording in the CMT. The analytics software made it easier to track navigation
along the polynomials [7], but no such tracking device was available for analysing navigation in the simulator. Instead, observers would need to mark specific places as visited as they reviewed the video afterwards. One of the challenges in evaluating the effect of sound is that it is not possible to track how a user perceives sound. This has to be judged through their behaviour. For example, possibly, a tracking device could help to conclude if sound and graphics had the desired effect, such as observing if a user is not moving close to a burning fire, and moving away from a fire engine before communicating. If sound had the desired effect, it should be noticed in users talking louder. This can be observed by listening to the audio, but it is better if it can be measured objectively with a sound level meter. We conclude that it takes considerable effort to transfer resources for data capture along the sensory context. The collaborative context influences data capture since synchronization is required between partners and it is important to note the time stamps of the users are synchronized. The form of transfer is rather complex, requiring perhaps all three forms of transfer, specialisation, transfer from other domains, such as games, and adoption accompanied with adaptation from the audio domain and eye tracking.

ANALYSIS OF CONTEXTUAL FACTORS

From the analysis of the success and failures of using methods resources across contexts, we conclude that the sensory context had an effect on the task selection and data capture, and the collaborative context had an impact on all method resources except participant recruitment (see Table 2). We only chose to indicate if a context had an impact on a method resource without quantifying the impact. For example, from our analysis of the two studies, we saw that the sensory context would have a large influence on data capture but collaborative context had a small influence on the same resource. We cannot conclude by analysing only one case of transfer, but there may be variability of impact across individual applications in the same context set.

Table 2 Contextual Factors Influencing Method Resources

<table>
<thead>
<tr>
<th>Method resources</th>
<th>Contextual Factors</th>
<th>Sensory context</th>
<th>Solo/Collaborative context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant recruitment</td>
<td>No impact</td>
<td>No impact</td>
<td></td>
</tr>
<tr>
<td>Task selection</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Training objective selection</td>
<td>No impact</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Data capture</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Control of experiment</td>
<td>No impact</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

The transfer of methods resources across contexts can be formalized with the following formula:


In cases where there is no impact as per Table 2 the Old Method Resource is equal to the Modified Method Resource. This assumes that the different contexts are independent and impact the method resources independently. However, if they are interdependent all the Contexts (set of Context-Old) for the source would have to be a part of the input simultaneously. This is to emphasise that when we cross context boundaries from old to new, project teams may use their collective experiences from multiple old contexts. The transfer function does not separate explicit knowledge, e.g. from the literature, as a variable, but it may be justified to have include it as a separate variable rather than tacit knowledge of the context. In addition to changed method resources, it may very well be that the set of method resources change. Thus, one can omit or add a method resource, e.g. participant recruitment, when crossing different values of a context.

It remains to be seen what kind of function the Transfer function is. We summarised three such transfer functions, generalisation/specialisation, adoption with adaptation, and transfer between domains. For each method resource we attempted to analyse if the two studies had applied these transfer functions. The fourth transfer function can be transfer across the sensory context as we saw in the example of the data capture where transfer takes place across sensory and collaborative contexts.

CONCLUSION

Technology, needs and development contexts are all factors that evolve rapidly and provide new opportunities to give enhanced experiences and satisfaction to users. This creates new requirements for design and evaluation methods. This paper has shed light on how this innovation of methods takes place. The analysis of methods resources showed that various factors influenced how the method resources were used or reused in each of the applications. Specifically, we looked at the impact of two contexts, the sensory and the collaborative contexts. We conclude that some of the method resources can be almost re-used directly, such as participant recruitment. The participant criteria had to be redefined and the local contact was different, but the main process was similar. The control of experiment was different for the CMT than the Mathematics cane and because of the influence by the solo/collaborative context that method resource had to be redesigned and could not be reused. Other method resources such as task selection and data capture had to be redesigned because of the influence of the sensory context and the solo/collaborative context. Because of the specialised domain of each application, especially the CMT, additionally, knowledge on processes and practices...
had to be taken into account when designing the scenarios of tasks for the evaluation.

Furthermore, the study showed that the method resources can be interdependent. For example, a detailed task scenario will require a high level of experimental control that produces different results than a non-scripted scenario.

In studying how methods transfer from one context to another we build new knowledge upon existing practices. While transfer of knowledge [15] is a powerful tool which has been used in many settings, it is challenging since two contexts need to be studied and the motivation for the transformation between them is hard to discover, what is the input, output and the transfer function which is most complex. This paper has given an example of such a study, where we have analysed evaluations of two training applications and looked at their different sensory and collaborative contexts and their impact on the methods resources. This has been only one of many examples needed to understand the transfer of methods resources across contexts.

ACKNOWLEDGEMENT
The author would like to thank the anonymous reviewers for valuable and constructive comments.

The research leading to these results has received funding from COST Action IC0904 and the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. [FP7-242474].

REFERENCES