Remote usability evaluations With disabled people

Conference Paper · January 2006

DOI: 10.1145/1124772.1124942 · Source: DBLP

4 authors, including:

Helen L. Petrie
The University of York

Fraser Hamilton

Remote Usability Evaluations with Disabled People

Helen Petrie
Department of Computer Science
University of York
Heslington York YO10 5DD
United Kingdom
Helen.Petrie@cs.york.ac.uk

Fraser Hamilton, Neil King, Pete Pavan
Designed for All Ltd
Prince Consort House, 109 – 111 Farringdon Rd
London EC1R 3BW
United Kingdom
fraser|neil|pete@designedforall.com

ABSTRACT
Finding participants for evaluations with specific demographics can be a problem for usability and user experience specialists. In particular, finding participants with disabilities is especially problematic, yet testing with disabled people is becoming increasingly important. Two case studies are presented that explore using asynchronous remote evaluation techniques with disabled participants. These show that while quantitative data are comparable, the amount and richness of qualitative data are not likely to be comparable. The implications for formative and summative evaluations are discussed and a set of principles for local and remote evaluations with disabled users is presented.

Author Keywords
Remote evaluation, usability research, usability testing and evaluation, disabled users

ACM Classification Keywords
H.5.2 [User interfaces]: Evaluation/methodology, User-centered design, Theory and methods; H.1.2 [User/Machine Systems]: Human Factors.

INTRODUCTION
When developing systems for specialist user groups with specific demographics, it can be difficult to find sufficient numbers of potential users of the system who can come to a facility to take part in usability evaluations. Traveling to users can be an expensive and impractical alternative. For example, Brush, Ames and Davis [2] found it difficult to find sufficient professional users of an urban planning tool locally, and as the users were distributed between Washington, Oregon, Utah, Texas and Hawaii, traveling to the users would have been prohibitively expensive and impractical. Instead they conducted both local and synchronous remote usability evaluations and found the results comparable. They found that the number of usability issues elicited, their type and severity, to be similar in both situations. Even when developing systems for general rather than specialist use, it can be difficult to recruit a sufficiently diverse sample of users locally for usability and user experience evaluations. In both these scenarios, disabled users may be particularly difficult to recruit locally. Increasingly, developers are aware that they need to ensure the usability of mainstream systems for disabled people, for a variety of reasons: legal (for example due to legislation such as the Americans with Disabilities Act and Section 508 of the Telecommunications Act in the USA, the Disability Discrimination Act in the UK, and similar legislation in many other countries), economic or moral. This is a separate issue from developing systems specifically to meet the needs of disabled users (often referred to as assistive technologies), which also require evaluation to ensure their usability for the target audience [16, 17].

In recent years we have conducted many accessibility and usability evaluations with disabled users and have often struggled to recruit locally a sample that meets particular required characteristics. We have traveled widely in the U.K. to undertake evaluations and on a number of occasions we have experimented with different types of remote evaluations to address this problem. In this paper we will report on two evaluations partly undertaken remotely that we have conducted with disabled users. Each has a different methodology and in each we were able to compare the effectiveness of local and remote evaluations. We had hoped to conduct one synchronous and one asynchronous evaluation, but for reasons explained below, both were asynchronous evaluations. Before presenting these case studies, we will discuss the different types of remote evaluations that can be undertaken, some of the issues of conducting usability evaluations with disabled users and the possible advantages and disadvantages of conducting remote evaluations with disabled users.

TYPES OF REMOTE USABILITY EVALUATION
"Remote evaluation" can refer to a wide variety of different techniques. A decade ago, Hartson, Castillo, Kelso, Kamler and Neale [7] outlined a range of different possibilities, some of which have now become commonplace, while others remain uncommon. These include:

Portable evaluation: uses a portable evaluation unit, which is taken to the users in their normal environments. This is now part of the standard set of techniques used by usability specialists.

Local evaluation at remote site: the system to be evaluated is transferred to a remote site, for example from the client's offices to a usability facility. This is now so common that it is probably not considered a remote evaluation technique by practitioners.

Remote questionnaire/survey: a system can be augmented to display a questionnaire to collect data from users. The appearance of the questionnaire is triggered by an event (such as task completion) during usage. This is now used in market research, but not frequently in usability research.

Remote control evaluation: involves data being transferred from the users' normal environments to a usability facility for synchronous or asynchronous observation and analysis. By using tools such as Morae [24] we can now expect these techniques to become more frequently used.

Video conferencing as an extension of the usability facility: involves the users undertaking the evaluation remotely while communicating with the evaluator via a video conference connection. This allows a communication situation that is very close to two way face-to-face communication patterns. It can now be augmented by screen sharing software, so that the evaluator can see (and record) what is happening on the user's screen.

Instrumented remote evaluation: an application can be augmented to collect usage data occurring in the user's normal environment. This technique is now widely used, but practitioners are aware of the problems of analyzing and interpreting the large amounts of data that are produced [8, 9, 11].

Semi-instrumented remote evaluation: uses selective data collection triggered directly by users performing tasks in their normal interaction with the system. Users are trained to identify usage events having significant negative or positive effects on their task performance or satisfaction. In particular, a negative effect is usually an indicator of a usability problem. Information about these events is transmitted to developers. This has not become widely used, probably due to the complexities of training users to identify the appropriate events and ensuring that they collect the data.

CLASSIFICATION OF REMOTE USABILTY EVALUATIONS

We would classify remote evaluations on several major dimensions (which are not necessarily orthogonal to each other):

Synchronous or asynchronous: Do the participant and the evaluator need to participate at the same time, even if they are in different locations?

Participant independence: Is there contact between the participant and the evaluator? If the evaluation is synchronous, do the participant and the evaluator communicate, or can the participant undertake the evaluation independently?

Is the evaluation process intrusive to the participant's interaction with the system or unobtrusive?

Does the participant require training to undertake the evaluation, and if so, can the training be undertaken remotely?

A further important consideration is whether the remote evaluation technique is suitable for, and useful in, formative or summative evaluations. In formative evaluations the object is to collect information about design flaws and inform re-design. One needs to be confident that appropriate information for this purpose can be collected in a remote evaluation situation. In formative evaluations it is particularly important that participants feel free to criticize a system and to avoid evaluator bias, and this may be easier if they are in the privacy of their own environment, rather than the potentially more threatening situation of the usability facility [1]. For summative evaluations, one aspect that may be of interest is whether the user can install and run a system on their own machine. Using a remote evaluation will simulate this situation much more accurately than a procedure that takes place in the usability facility. However, in summative evaluations one may wish to collect metrics of the usability of a system, and one needs to be confident that such quantitative measures can be collected in a remote evaluation situation. These are some of the issues we wished to explore by comparing local and remote evaluations.

DESIGNING FOR DISABLED PEOPLE

As mentioned above, designing for disabled people is becoming an increasingly important topic for a variety of reasons, but especially due to recent legislation in many countries promoting the rights of disabled people. A number of philosophies and methodologies have been developed to support this process. Firstly there has been the development of the universal design/design for all/universal usability philosophy. As Vanderheiden and colleagues have noted [25, 26], there are probably as many definitions of these concepts, all closely related, as there are people discussing them, and considerable misunderstanding. Many developers worry that they will be expected to produce a system that will be usable by every user, regardless of capability, and that they might have to seriously compromise their overall design to achieve this aim. Clearly this would be in no-one's interest, and with the increasing ability to personalize interfaces to meet the requirements of different users [21], this is not necessary.

Secondly, there have been numerous sets of guidelines to help developers produce systems that are accessible and usable by disabled people. These include very general guidelines, such as those produced by the Center for Universal Design [3], to the very specific guidelines for Web user agent [27], authoring tool [28] and content developers [29]. However, it is not clear whether providing guidelines is an effective method for ensuring usable designs. The HCI community debated this issue in the late 1980s, when guidelines for interface design, such as those by Smith and Mosier [23] were evaluated and found to be difficult to use [4, 6, 14, 15]. In our experience with designs for disabled people, developers need to have a conceptual framework in which to situate disabled-related guidelines, which they often do not have due to lack of experience with disabled people and their technologies. For example, in teaching Web developers we have found that they are often already familiar with the relevant guidelines, but are startled when they see the actual effects of violating guidelines in a demonstration by disabled people. Seeing a blind person try to access a badly designed web page often produces a "eureka" moment of understanding for web developers.

Partly as a consequence of the problems with relying only on guidelines to drive design for disabled people, and partly due to the need to develop practical methodologies that instantiate the universal design philosophy, a number of authors have developed frameworks for extending user-centered design methodologies to include disabled participants. For example, Newell and colleagues have developed "user sensitive inclusive design" [16] and the User Project developed the UserFit methodology [22]. In addition, interesting techniques have been developed to allow developers themselves to make an initial assessment of how usable a design might be for people with particular disabilities before conducting user evaluations [13].

One of the many challenges of designing for disabled people is the fact that there are many possible disabilities, of the sensory, physical and cognitive systems, and that these can occur in combinations, rather than singly. How do we design for a person with both a hearing and sight loss, or a blind person with only one hand? This is particularly important as with aging, everyone is likely to acquire multiple disabilities, and although they may each be relatively minor, their combined effects may be major. The situation may not be as complex as it may seem however, when one considers the functional consequences of technology use for different disabilities, the number of issues is not as great as one might imagine [18]. Indeed when we conducted an extensive study of the use of the Web by people with a range of sensory, physical and cognitive disabilities, there were far more overlaps in the issues raised for the different disability groups than we expected [5, 19]. However, this again highlights the need for more empirical data from users with disabilities, and therefore the need for evaluations with disabled users as these relationships are not yet well understood.

ADVANTAGES AND DISADVANTAGES OF REMOTE EVALUATIONS WITH DISABLED USERS

Usability and user experience specialists often complain of the difficulty of recruiting people with disabilities for their evaluations. This is generally because they lack contacts with disabled people that could be overcome by developing relationships with organizations of disabled people which would allow them to conduct local evaluations. However, there are a number of specific advantages of conducting remote evaluations with disabled people. For example, disabled people with use assistive technologies, both hardware and software, such as screenreaders for blind people and screen magnification programs for partially sighted people [20], single handed keyboards, and sip and puff systems for people with physical disabilities. Usability facilities are very unlikely to have every type of assistive technology required, and individuals may configure their assistive technologies in ways that are difficult to re-create in the laboratory for each participant. People with some disabilities may have difficulties traveling to the facility, either because of the nature of their disability or the inaccessibility of public transport services and the facility itself may not be accessible.

However, one should also realize that there are also possible disadvantages of conducting remote evaluations with disabled people to overcome these problems. One of the greatest difficulties for usability and user experience specialists is understanding how the use of assistive technology and people's disabilities affects their interaction with a system (see the discussion of the "eureka" moment for Web developers, above). Careful observation of people using their assistive technologies and discussion with them greatly increases understanding in this area. This is much more difficult to achieve in remote evaluations.

We will now present two case studies of evaluations where we have been able to compare results from local and remote evaluations with disabled participants. In both cases these were asynchronous remote evaluations, with the participants undertaking the evaluation independently (although in the first case study, a post-study interview was conducted synchronously by telephone). They were also both classic usability studies in the sense that participants were asked to undertake specific tasks and reply to questions about their experience. However, one evaluation was a formative evaluation and the other was a summative evaluation. In one case, participants were trained locally in how to conduct the evaluation, whereas in the other, participants were sent instructions and expected to train themselves.

CASE STUDY 1: FORMATIVE EVALUATION OF THE TEDUB SYSTEM WITH BLIND PEOPLE

Introduction

The TeDUB system (Technical Diagram Understanding for Blind People) is designed to automatically convert certain types of technical diagrams into a spoken format and allow blind people to explore them on a personal computer. In the project which developed the TeDUB system, three classes of technical diagram were used: electronic circuit diagrams, UML diagrams, and architectural floorplans. An iterative, user-centred design lifecycle was used in the project, with three major prototypes developed, one with each of the diagram types. For the first two types of diagrams (electronic circuit diagrams and UML diagrams), finding blind people who had some understanding of the content of the diagrams was extremely difficult. For example, amongst blind people in the computing industries, of whom there are considerable numbers, there is interest in using UML. However, because it is a very visual format, few blind people have yet had any experience with it. Therefore to find sufficient numbers of people for the evaluations, a combination of local evaluations and traveling to the participants' localities (all over the U.K.) was used. For the final evaluation, we decided to undertake a remote evaluation with some participants and explicitly compare the results with a local evaluation. We planned to make this a synchronous remote evaluation, with a screen sharing facility and telephone communication. Unfortunately, security features of the university computer network thwarted all attempts to conduct a synchronous evaluation, so an asynchronous evaluation was undertaken.

The TeDUB system can either take a digital representation of a diagram (particularly in the case of UML diagrams) or conduct image processing on a visual image and extract the appropriate conceptual information from the diagram [10, 12]. It uses a commercial screenreader (JAWS V4.5/V5.0) to convert the resulting textual information into synthetic speech for blind users. Interaction is via a standard computer keyboard and a computer gaming joystick. These devices enable the user to move around the diagram, say a floorplan of an office building, and construct a mental model of the building and its component rooms using information presented through speech, non-speech sound (both 2D and 3D audio presentation) and forcefeedback (from the joystick).

A range of functions was developed to assist users when navigating through a diagram, including:

Text View: presents the user with a hyperlinked textual description of the diagram;

Map View: enables the user to move the joystick around the whole area of the diagram, presenting the room name at the current location of the joystick;

Spatial Navigation View: enables the user to move the joystick and establish the location of connected rooms;

Walkthrough: lets the user walk around the floorplan navigating by compass directions;

Route Planner: calculates a room-by-room pathway between 'to' and 'from' destinations, as selected by the user;

Literal Representation: enables the user to establish the shape of a room by restricting the joysticks sweeping movement to an area the shape of the room.

Here we report on the evaluation of the third major prototype, for use with architectural floorplans.

Methodology

Design

All participants were asked to install the TeDUB system at home (using the system manual and installation guide) and to complete an introductory tutorial in their own time before the evaluation took place. A three week period was allowed for this.

Then half the participants came to our usability laboratory and undertook the evaluation with a researcher. The other participants were sent the same evaluation protocol and asked to undertake the evaluation by an agreed time, so a researcher could then telephone to conduct the post-evaluation interview.

Participants

8 participants took part in the evaluation, 4 in the local evaluation and 4 in the remote evaluation. 7 were men and 1 was a woman. Their ages ranged from 19 to 49 years, with a mean age of 32 years. All participants lacked any functional vision; three had been blind since birth and the others had lost their sight before the age of 3 years.

Materials

All participants were sent the following materials prior to the evaluation:

- A Saitek Cyborg 3D Force joystick
- TeDUB system software and installation guide
- A set of TeDUB architectural floorplans
- The TeDUB system manual
- The TeDUB tutorial explaining the interface and architectural floorplans with practice tasks
- Consent and confidentiality forms.

For the evaluation, the following materials were used:

Evaluation tasks: a series of tasks was developed to cover all the major features of the TeDUB system. These tasks were woven into a single scenario in which the participant explored a building.

Post-evaluation interview schedule: This consisted of a series of rating scales and open-ended questions to capture participants' reactions to the major functions of the TeDUB system and their overall impression of the utility of the system and how it could be improved.

Procedure

Pre-evaluation phase: All participants were sent the initial system materials three weeks before their evaluation was scheduled and asked to install the hardware and software and work through the introductory tutorial. Each participant was encouraged to contact the TeDUB researchers with any questions regarding the installation and tutorial, but to attempt to learn how to use the system on their own.

Evaluation Phase:

Local evaluation procedure: Participants came to the usability laboratory. Participants attempted the tasks while 'thinking aloud' and the evaluator recorded the problems uncovered.

Remote evaluation procedure: Participants were sent evaluation materials. Participants attempted evaluation tasks by an agreed time. While doing the tasks, participants were asked to make notes on problems they encountered and these were sent to the evaluator.

Post evaluation Phase*:* Evaluator either conducted the post evaluation interview face-to-face (local evaluations) or phoned participants (remote evaluations).

Results

All participants used the initial system material (manual and tutorial) under the same conditions (remotely). However they rated its usability under different conditions during the post-evaluation phase (either in person to the researcher during the local evaluations or over the phone during the remote evaluations). This difference provides an initial useful indication of any effect of the difference in information elicitation from the participants. The mean ratings of both groups are identical for the system manual and installation guide and very similar for the tutorial (see Table 1), with no significant differences between local and remote evaluation situations.

Table 2 shows the mean ratings given for eight functions of the TeDUB system in the local and remote evaluation conditions (due to technical problems, participants in the remote evaluation did not evaluate the Bookmark function). Only one difference between the two evaluation situations was statistically significant (measured on independent sample t-tests between the two conditions), the Map View function. This was used in a task in which participants were asked to identify the spatial location of all of the rooms by using the audio cues incorporated into the Map View function. However, in three cases in the local evaluations researchers noted that at least one room was not correctly identified, due to its close proximity to another room. The participants were made aware of this problem and asked to investigate further, which resulted in the following comment from one participant:

This feature is no good when you have a big drawing … you can not distinguish between objects due to the constant beeping …

Due to the nature of the remote evaluations, the researchers could not establish if the remote participants had correctly identified all of the rooms, and the qualitative data provided by these participants showed no clear evidence either way. Therefore, it is plausible that the significantly more positive mean rating given to the Map View function by the remote participants was due to their lack of awareness of the problems of which the local participants were aware. This is a particular issue for the TeDUB system, but highlights the type of intricacies of a system that may be missed in a remote evaluation.

Documentation	Local evaluation	Remote evaluation
System Manual and Installation Guide	4.3	4.3
Tutorial	4.3	4.2

NB: Scale 1 = not at all useful to 5 = very useful.

Table 1: Mean ease of use ratings for initial documentation

Both remote and local evaluations provided considerable quantities of qualitative data, but as illustrated with the two examples below, the local evaluations provided far richer data as the researchers were able to record problems that the participant may not have been aware of, and were in a position to prompt the participant to explore these problems, comment on them, and analyse what had caused them.

For example, one of the local participants found the naming of objects and presentation in the Text View confusing:

	Local evaluation	Remote evaluation	Sig Diff
Spatial Nav View	3.8	3.8	No
Annotation	4.0	3.6	No
Text View	3.6	4.0	No
Map View	3.6	4.3	Yes
Walkthrough	3.3	3.3	No
Route Planner	4.0	3.6	No
Bookmark	4.0	-	No
Search	4.3	4.6	No

NB: Scale 1 = not at all useful to 5 = very useful

Table 2: Mean usability ratings for TeDUB functions

I take it each thing in brackets is the thing they are talking about *[Door to Hall and Living room]… [Door to Bathroom]* it is not clear what the relationship is… I cannot technically understand these relationships … … it just doesn't work …

In contrast, the issues identified by the remote participants were generally high level, and often subjective, rather than specifically articulated usability problems. For example, commenting on the same feature of Text View, a remote participant noted:

there are a lot of problems with the descriptions … I don't like the Text View …

Clearly, this comment does not tell the researchers anything specific about what the problems that the participant actually encountered with the Text View function.

CASE STUDY 2: SUMMATIVE EVALUATION OF WEBSITES BY DISABLED PEOPLE

Introduction

As part of a very extensive investigation of website accessibility undertaken for the British Disability Rights Commission [5, 19], it was decided to conduct user evaluations of a large number of websites with people with a range of disabilities relevant to web access. Our target was to evaluate 100 websites with 50 disabled people in two months. We quickly realized that the only way to complete these evaluations within the time frame available and with the human resources available to the project, was to have the majority of the evaluations conducted remotely and without specialist support.

Methodology

Design

100 websites were selected to cover a range of sectors, technologies used on the websites and levels of technical accessibility. Each participant evaluated 10 websites, in both local and remote evaluation situations. They attempted two tasks per website, making a total of 1000 tasks undertaken in total. Tasks were chosen to be typical of what people would expect from that website. For example, if it was an e-banking site, the task might be to find the current interest rate for a savings account. The tasks were chosen by one researcher and then independently validated by another researcher to ensure that they were appropriate and achievable.

Participants all came to the laboratory and evaluated 2 websites with a researcher taking them through a protocol and instructing in how to use the protocol on their own. They then evaluated one website on their own to practice the protocol. They then evaluated a further 7 websites in their own computing environments.

Participants

51 disabled people took part in this evaluation. Their ages ranged from 18 to over 60 years and 33 were men and 17 were women. The participants also represented a wide range of levels of experience with computers and the web, and types of assistive technology used.

Five groups of disabled people were included, being those who:

- are totally blind or lack any functional vision (who use screen readers to access the Web);
- are partially sighted (may use screen magnification programs to access the Web);
- have physical disabilities that affect their access to the Web (e.g. lack of control of arms and hands, tremor, lack of dexterity in hands and fingers).;
- are profundly deaf (including users of British Sign Language) or hard or hearing;
- have specific reading problems such as dyslexia.

Procedure

Initial local evaluation procedure: For each website the participant was asked to freely explore the site for a few moments, to familiarize themselves with it. They were then asked to undertake each task, while 'thinking aloud' to the researcher. The researcher recorded all problems encountered. When the task was completed, the researcher asked them to rate how easy it was to complete the task on a 1 (very difficult) to 7 (very easy) scale. When both tasks were completed, the researcher asked overall questions about the website and asked the participant to rate how well they thought the website had taken their disability into consideration.

Practice for remote evaluation: The procedure of how to undertake the remote evaluations was explained to participants (this procedure was very similar to the evaluation they had just undertaken). To practice the remote evaluation procedure, the participants were then left to evaluate a website on their own. The participants recorded any problems they encountered as they undertook the tasks, and after completing the tasks, completed the ratings as before. In this instance they were also asked to rate how confident they were that they had successfully completed the task. When this practice of the remote evaluation procedure was complete, the researcher returned and discussed the practice evaluation and answered any questions.

Remote evaluation procedure: participants undertook evaluation of 7 websites in their own time over a period of two to four weeks.

Data analysis

For the local evaluation, the researcher could observe whether the participant had successfully completed the task. However, in the remote evaluation, it may not have been clear to the participant (as we have seen in the example of the Map View problem in the TeDUB case study). For this study we wanted to improve on the methodology of the TeDUB study and make a good estimation of whether participants in the remote evaluation had successfully completed tasks. So for each task we asked participants to rate their confidence that they had been successful on a scale from 1 (not at all confident) to 7 (very confident). Then we took only those tasks where they were very confident (rating or 6 or 7) that they were successful or very confident that they were not successful and used those tasks for the success rate analysis. This was possible because the number of tasks completed in the remote situation was so high (700). An analysis was also undertaken using an even stricter criterion (only taking the tasks with the top/bottom confidence ratings) and this made no difference at all to the results. Of course, one cannot be totally sure that a participant was not completely mistaken and thought they had successfully completed a task when they had not, but we think this is sufficiently unlikely to have had a major impact on the results.

Results

The task success rate was 76.4% for local evaluations and 75.3% for remote evaluations, not a significant difference (chi-square = 0.08, df = 1, n.s.). In addition, there were no significant differences between the different disability groups in the success rates in local versus remote evaluations (although there were very large differences between the disability groups in their success rates, with blind participants finding it significantly more difficult to successfully complete the tasks [3]).

The mean number of problems reported per website was 6.5 for local evaluation and 1.6 for remote evaluations, a significant difference ($t = 16.45$, $df = 14$, $p < 0.01$, this analysis is based on an analysis of data from 16 participants from all disability groups, rather than the total sample of 50 people). Indeed, the number of problems reported per website in the local evaluations is over four times as many as in the remote evaluations. In addition, as we saw in the TeDUB case study, the problems reported provided much richer detail on the nature and cause of the problem.

	Local	Remote
Ease of task	4.8	5.1
Consideration of disability	3.6	4.1

Table 3: Mean ease of task and consideration of disability ratings

The participants were also asked to rate how easy the task was for them to undertake, and whether the site had taken their disability into consideration. The mean ratings for these two measures for the local and remote evaluation situations are shown in Table 3. There were no significant differences between the ratings in the two evaluation situations.

DISCUSSION AND CONCLUSIONS

The two case studies show that quantitative data collected in asynchronous remote evaluations with disabled users is comparable with that collected in local evaluations. There was only one significant difference in one of the quantitative measures we collected (the rating of the Map View for the TeDUB system) and there appeared to be a specific reason for that. However, the amount and richness of the qualitative data were very different between the two evaluations, even in a situation as in the DRC case study, in which we trained participants in how to undertake the remote evaluation procedure.

These results have implications for the type of evaluation undertaken. If the evaluation is formative, the quantity and richness of the qualitative data is probably very important in understanding how to improve the system, and thus the remote evaluation technique may not be appropriate. However, if the evaluation is summative, and one is interested in whether participants can install and run the system themselves and how they rate the key functions of the system, then a remote evaluation may be quite appropriate.

A further consideration is that currently usability and user experience specialists often lack detailed understanding of how people with disabilities use their assistive technologies. It is our contention that only by directly and carefully observing disabled people interacting with technologies and discussing their experiences with them will we develop deep understanding of these issues. This is also vital for those who undertake the implementation work. Some of the most fruitful experiences for us in conducting evaluations with disabled participants have been when the implementation team have observed the evaluations with disabled people live. This has given them a real feeling for what it is like to use their application with assistive technologies. They also then get a chance to discuss the problems and possible solutions with disabled people.

Currently achieving this rich interaction between participants, researchers and developers is very difficult in remote evaluation situations, even with the latest technologies. However, with high quality video conferencing, broadband connections and remote recording systems, it should be possible to conduct remote evaluations that capture the richness of local evaluations. Further research needs to investigate exactly what information the researcher is using in the local evaluation situation, for example non-verbal behaviour by the participant, and ensure that this is captured and transmitted between locations to allow for totally effective remote evaluation.

We conclude by presenting a set of principles for local and remote evaluations with disabled users.

PRINCIPLES FOR CHOOSING A LOCAL OR REMOTE EVALUATION FOR DISABLED PARTICIPANTS

Be aware of the advantages and disadvantages of remote and local evaluations: Both evaluation types appear to gather comparable quantitative data for summative evaluation. However, remote evaluations enable users to conduct the evaluations using their own assistive technologies with familiar configurations. But local evaluations collect more data which is also richer compared with remote evaluations. This may make local evaluations particularly appropriate for formative evaluations.

Local evaluations: it is essential to find out what assistive technologies (to the specific version) participants use, acquire it, and test it thoroughly prior to the test session.

Ask disabled participants if they can get to the location (e.g. using public transport). If public transport is not feasible, work with them to make alternative arrangements.

During local evaluations, allow time for participants to configure in-house assistive technologies and other equipment. Bear in mind that an exact replication of all users' configurations is likely to be difficult, if not impossible.

Be sensitive to the needs of disabled people. Many organizations can offer training in how to work with people with particular disabilities. When in doubt, talk to the participants themselves about their requirements, they are usually much less embarrassed than you are.

Remote evaluations: When planning remote evaluations ensure that appropriate training material is sent to users in advance and that support is available to users who may have difficulty installing the system or doing the evaluation.

Although more difficult to organize than asynchronous evaluations, synchronous remote evaluations may allow usability experts and other remote observers to gather more qualitative data on specific problems. For asynchronous remote evaluations, consider following up the evaluation with a telephone call to gather more qualitative data on specific problems.

ACKNOWLEDGEMENTS

We would like to thank all participants in both evaluations who contributed greatly to our research.

The TeDUB Project (IST-2001-32366) was supported by the Commission of the European Communities. We are grateful for the collaboration with the Consortium, which consisted of City University (UK), F. H. Papenmeier GmbH & Co. KG (Germany), FNB (Netherlands), Laboratory for Semantic Information Technology (Bamberg, Germany), National Council for the Blind of Ireland, Technologie-Zentrum Informatik (Bremen, Germany), Unione Italiana Ciechi di Verona (Italy), and The University of Manchester (UK). We would also like to thank Anne-Marie Burn and Alex Fitzgerald for their contributions to the TeDUB Project.

The web accessibility research was conducted as part of the Formal Investigation into Website Accessibility undertaken by the Disability Rights Commission. We would like to thank Commissioners Michael Burton and Christopher Holmes, as well as Stephen Beesley and Nick O'Brien for their support and collaboration.

REFERENCES

1 Bradner, E. (2004). Keeping your distance. *User experience*, 3(6), 11.

2 Brush, A.J., Ames, M. and Davis, J. (2004). A comparison of synchronous remote and local usability studies for an expert interface. *Proc. CHI '04.* New York: ACM Press.

3 Center for Universal Design. *Principles of Universal Design*. Available at: www.design.ncsu.edu:8120/cud/univ_design

4 DeSouza, F. and Bevan, N. (1990): The use of guidelines in menu interface design: evaluation of a draft standard. In D. Diaper, D. Gilmore, G. Cockton and B. Shackel, (Eds.), *Proceedings of INTERACT 90 - 3rd IFIP International Conference on Human-Computer Interaction*. Cambridge, UK.

5 Disability Rights Commission (2004). *The web: access and inclusion for disabled people.* London: The Stationery Office.

6 Hammond, N., Gardiner, M., Christie, B., and Marshall, C. (1987). The role of cognitive psychology in user-interface design. In *Applying cognitive psychology to user-interface design*. Chichester: John Wiley.

7 Hartson, H.R., Castillo, J.C. Kelso, J. and Neale, W.C. (1996). Remote evaluations: the network as an extension of the usability laboratory. *Proc. CHI '96*, New York: ACM, 228 – 235.

8 Hilbert, D.M. and Redmiles, D. F. (1998). An approach to large-scale collection of application usage data over the Internet. *Proc. 20th International Conference on Software Engineering*. IEEE Press.

9 Hong, J.I., Heer, J., Waterson, S. and Landay, J.A. (2001). WebQuilt: a proxy-based approach to remote web usability testing. *ACM Transactions on Information Systems*, 19(3), 263 – 285.

10 Horstmann, M., et al. (2004). Automated interpretation and accessible presentation of technical diagrams for blind people. *New Review of Hypermedia and Multimedia*, 10(2), 1 - 23.

11 Ivory, M. Y. and Hearst, M. A. (2001). The state of the art in automated usability evaluation of user interfaces. *ACM Computing Surveys*, 33(4), 470 – 516.

12 King, A. et al.. (2004). Presenting UML software engineering diagrams to blind people. *Proceedings of 9th International Conference on Computers Helping People with Special Needs, (Lecture Notes in Computer Science LNCS 3118).* Berlin: Springer Verlag.

13 Law, C., Barnicle, K. and Henry, S.L. (2000). Usability screening techniques: evaluating for a wider range of environments, circumstances and abilities. *Proc. UPA 2000*. Available at: http://trace.wisc.edu/docs/usability_screen_techs_UPA2000/usability_screening_techs.htm

14 Löwgren, J. and Nordquist, T. (1992). Knowledge-based evaluation as design support for graphical user interfaces. *Proc. CHI '92*. New York: ACM Press.

15 Mosier, J. N. and Smith, S. L. (1986). Application of guidelines for designing user interface software. *Behavior and Information Technology*, 5(1), 39-46.

16 Newell, A.F. and Gregor, P. (2000). User sensitive inclusive design. In J. Scholtz and J. Thomas (Eds.), *CUU 2000: First ACM Conference on Universal Usability*. New York: ACM Press.

17 Petrie, H. (1997). User-centred design and evaluation of adaptive and assistive technology for disabled and elderly users. *Informationstechnik und Technische Informatik, 39*(2), 7 - 12.

18 Petrie, H. (2001). Acessibility and usability requirements for ICTs for disabled and elderly people. In J.G. Abascal and C. Nicolle (Eds.), *Inclusive guidelines for human computer interaction.* London: Taylor and Francis. 0-748409-48-3.

19 Petrie, H., and Hamilton, F. (2004). The Disability Rights Commission Formal Investigation into Web Site Accessibility. In Dearden, A. and Watts, L. (Eds.), *Proceedings HCI 2004: Design For Life (Leeds, UK)*, pp. 157-158.

20 Petrie, H., O'Neill, A-M. and Colwell, C. (2002). Computer access by visually impaired people. In A. Kent and J.G. Williams (Eds.), *Encyclopedia of Microcomputers Volume 28*. New York: Marcel Dekker.

21 Petrie, H., Weber, G., and Fisher, W. (2005). Personalisation, interaction and navigation in rich multimedia documents for print-disabled users. *IBM Systems Journal*, 44(3).

22 Poulson, D., Ashby, M. and Richardson, S. (1996). *USERfit, A practical handbook on user-centred design for Assistive Technology*, Brussels: ESC-EC-EAEC.

23 Smith, S.K. and Mosier, J.N. (1986). Guidelines for designing user interface software. Available at: http://hcibib.org/sam/

24 Techsmith. www.techsmith.com/products/morae

25 Vanderheiden, G. (2000). Fundamental principles and priority setting for universal usability. In J. Scholtz and J. Thomas (Eds.), *CUU 2000: First ACM Conference on Universal Usability*. New York: ACM Press.

26 Vanderheiden, G. and Tobias, J. (1998). Barriers, incentives and facilitators for adoption of universal design practices by consumer product manufacturers.

27 World Wide Web Consortium. Authoring Tool Accessibility Guidelines. Available at: http://www.w3.org/WAI/intro/atag.php

28 World Wide Web Consortium. User Agent Accessibility Guidelines. Available at: http://www.w3.org/WAI/intro/uaag.php

29 World Wide Web Consortium. Web Content Accessibility Guidelines. Available at: http://www.w3.org/WAI/intro/wcag.php