Univ Access Inf Soc (2016) 15:63–76
DOI 10.1007/s10209-014-0378-8

LONG PAPER

Contrasting usability evaluation methods with blind users

Mei Miao · Hoai Anh Pham · Jens Friebe · Gerhard Weber

Published online: 27 August 2014

Abstract Usability tests are a part of user-centered design. Usability testing with disabled people is necessary, if they are among the potential users. Several researchers have already investigated usability methods with sighted people. However, research with blind users is insufficient, for example, due to different knowledge on the use of assistive technologies and the ability to analyze usability issues from inspection of non-visual output of assistive devices. From here, the authors aspire to extend theory and practice by investigating four usability methods involving the blind, visually impaired and sighted people. These usability methods comprise of local test, synchronous remote test, tactile paper prototyping and computer-based prototyping. In terms of effectiveness of evaluation and the experience of participants and the facilitator, local tests were compared with synchronous remote tests and tactile paper prototyping with computer-based prototyping. Through the comparison of local and synchronous remote tests, it has been found that the number of usability problems uncovered in different categories with both approaches was comparable. In terms of task completion time, there is a significant difference for blind participants, but not for the visually impaired and sighted. Most of the blind and visually impaired participants prefer the local test. As for the comparison of tactile paper prototyping and computer-based prototyping, it has been revealed that tactile paper prototyping provides a better overview of an application while the interaction with computer-based prototypes is closer to reality. Problems regarding the planning and conducting of these methods as they arise in particular with blind people were also discussed. Based on the authors' experiences, recommendations were provided for dealing with these problems from both the technical and the organization perspectives.

Keywords Blind people · Usability testing · Local test · Synchronous remote test · Tactile paper prototyping · Computer-based prototyping

1 Introduction

A user-centered design process such as the usability engineering life cycle [1] consists of several stages and various usability evaluation methods (UEM) may be applicable at different stages. To evaluate design concepts in the early stages, the most often used UEM is rapid prototyping in both paper and computer media. To evaluate prototypes or real products in the late stages, methods such as usability testing in laboratory (local test) or remote usability testing (remote test) are often applied.

Blind users are a particular user group. Usability testing with blind people is necessary if they are among the potential users. In the following, a scenario to illustrate why it is necessary to deal with this topic has been used. John is a designer who applied user-centered design process for developing a voice over IP program for the first time. With the help of a user requirements analysis, he was aware of the potential users and of their requirements on

M. Miao (✉) · H. A. Pham · J. Friebe · G. Weber
Department of Computer Science, Technische Universität Dresden, Nöthnitzer Straße 46, 01187 Dresden, Germany
e-mail: mei.miao@tu-dresden.de

H. A. Pham
e-mail: hoai_anh.pham@mailbox.tu-dresden.de

J. Friebe
e-mail: jens.friebe@gmx.net

G. Weber
e-mail: gerhard.weber@tu-dresden.de

the program. This program is to be used not only by sighted but also by blind people. Based on the requirements, he developed three design ideas, which should be evaluated by potential users. However, the challenge he faced is that he did not know which evaluation method he should use for testing with blind people and how could he show blind people his ideas. After 3 months, he finished the first version of the program and planned a usability test of the first version. He asked himself again: Which method is suited for testing of the first version with blind people?

In this paper, two critical questions are addressed:

- In the early stages of user-centered design, which method is better suited for blind users, tactile mock-ups (Tactile Paper Prototyping—TPP) or high-fidelity prototypes (Computer Based Prototyping—CBP)?
- In the later stages of user-centered design, which method is better for blind users, local testing or synchronous remote testing?

Several researchers have investigated usability methods with sighted people. However, research with blind users is insufficient. In two separate studies, the authors have investigated these four usability methods with blind people. In terms of effectiveness of evaluation and the experience of participants and the facilitator, TPP was compared with CBP in study 1 and local testing with synchronous remote testing in study 2.

In TPP, blind participants interact with tactile paper prototypes only, neglecting that in reality, they mostly interact with an interface supporting speech output as well as Braille display. They do not interact with tactile paper prototypes in the same manner as with the real applications. The different modes of perception could be a reason that TPP is not intuitive for evaluating design concepts for blind users. In CBP, blind participants interact with the prototypes in the same non-visual manner as with the final products. Thus, the differences between TPP and CBP have been investigated for testing design concepts in study 1. Study 2 focused on an issue that is challenging: the technical aspect in setting up synchronous remote tests with blind people.

2 Related work

2.1 Related works on paper prototyping and computer-based prototyping

Paper Prototyping is a widely used method to evaluate design concepts in the early stage of product design. Depending on modes of perception of the prototypes, there are visual, haptic and tactile paper prototyping [2]. Visual paper prototyping is the most widely used. The visual paper prototypes are usually created with pen and paper, which can only be used by sighted users but not by blind users. Haptic paper prototyping is a special form of haptic low-fidelity prototyping[1] [3]. To create haptic paper prototypes, cardboard is a common material. With such prototypes, haptic interaction of an application can be simulated and evaluated by users. However, both visual and haptic paper prototyping are not or infrequently applicable for blind users. Tactile paper prototyping allows the participation of blind users as the prototypes are printed by Braille printer. Miao et al. [2] outlined what should be taken into account while preparing and conducting tactile paper prototyping.

Computer-based prototyping is normally high-fidelity prototyping. It allows users to interact with prototypes as though it were a real product [4]. High-fidelity prototypes are created with programming tools such as Visual Studio. Henceforth, it takes more time and effort to create them than low-fidelity prototypes. To create high-fidelity prototypes for blind users, accessibility of the prototypes must be considered due to the fact that blind users interact with computers using assistive technologies such as screen readers utilizing synthetic speech and Braille displays.

ISO 9241-171 [5] and WCAG 2.0 [6] provide guidance and recommendations for designing more accessible software and Web content. There are also different comparisons of low-fidelity and high-fidelity prototyping with sighted people. Virzi et al. [7] compared the usability problems uncovered using low-fidelity (paper prototyping) and high-fidelity prototyping (computer-based prototyping). The same sets of usability problems were found in both conditions. They suggested that the use of low-fidelity prototypes can be effective throughout the whole product design process and not just at the initial stages. Miriam et al. [8] have tested the influence of fidelity (low or high) and media (paper or computer-based) on feedback from users. They compared user testing with low-fidelity and high-fidelity prototypes in both computer and paper media of Web sites. They found that usability testing results were independent of fidelity and medium of prototypes.

2.2 Related works on local and synchronous remote testing

Several UEMs have been developed to evaluate the usability of a product. Depending on the location where the usability tests will be conducted, there are local and remote usability tests. In a local test, participants are invited to come to the usability laboratory in an unfamiliar testing environment and perform some tasks. Remote testing allows participants and the facilitator to be in different locations. The participants perform some tasks in their own workplace or at home. Remote testing can be divided into synchronous and

[1] Fidelity means how similar are the produced prototypes to the final product.

asynchronous remote tests. In a synchronous remote testing, the communication between participants and the facilitator happens in real time, whereas in an asynchronous remote testing, an interview occurs after the test. Real-time communication requires a screen sharing/recording tool, an Internet connection and a live audio connection between participants and the facilitator. Consequently, it takes more effort to conduct synchronous remote testing.

Finding sufficient participants for performing tests locally can be difficult, in particular when it requires disabled participants to travel to a usability laboratory. Also, traveling can be very expensive to users [9]. As such, remote testing presents a sound alternative.

In recent years, different comparisons between local and remote tests have been carried out. Already in 1996, Hartson et al. [10] investigated local and synchronous remote tests. They found that there was no significant difference in terms of number of usability problems found and participant experience. Selvaraj [11] has compared local and synchronous remote tests in particular for usability novices and usability experienced participants. It was found that there was no significant difference in the number and the severity of critical incidents (CIs) [12] collected. Brush et al. [13] also compared local and synchronous remote tests and found no significant differences in terms of the number of usability problems found, their types and their severities. These studies were all carried out with sighted people. In 2006, Petrie et al. [9] investigated local and asynchronous remote tests with blind participants. A synchronous remote test was planned; however, due to technical reasons, an asynchronous remote test was actually undertaken. They found that the quantitative data collected with blind users in both tests were comparable. However, the local tests provided far richer data as the researchers were able to record problems that the participants may not have been aware of and could prompt the participants to explore and analyze these problems.

Noticeably, almost all of the studies mentioned above centered on the comparison with sighted people and did not address blind and visually impaired people.

Recognizing this, the procedural details for the comparison of tactile paper prototyping and computer-based prototyping with blind people, the comparison of local and synchronous remote tests with blind and visually impaired people and the results of the comparisons are reported.

3 Methods of data collection and criteria for comparisons

3.1 Methods of data collection

In order to collect the quantitative and qualitative data for the comparisons, diverse methods have been used in both studies. A camera which was positioned in the usability laboratory recorded the entire test sessions in both studies. All users were interviewed to collect the demographic information of participants before the tests and to evaluate the methods after the tests. The participants were asked to verbalize what they were doing, thinking and feeling as they were performing the tasks. With the help of *WebEx*™, it was possible to record screen activities. Additionally, the following data collection methods were used in study 2: (1) participants using synchronous remote testing were asked to use the microphone so that their screen reader outputs could also be recorded and (2) in order to analyze keyboard input data, a tool named *ShowKeyboard* was programmed. This tool was installed on the computers used in both local and synchronous remote tests.

Table 1 Criteria for comparison of the methods

Criterion 1: Effectiveness of the evaluations	
Quantitative data	*Qualitative data*
Task completion time	Severity of the problems found (study 2)
Number of usability problems found	Categories of the problems found (study 1)
Criterion 2: Perceptions and experiences of participants	
Quantitative data	*Qualitative data*
Difficulty of the tasks (study 1)	Positive and negative comments on the methods
Time for preparation, setup and dismantling of the tests (study 2)	Nonverbal reactions (if possible)
Number of positive and negative comments on the methods (study 2)	
Criterion 3: Perceptions and experiences of facilitator	
	Qualitative data
	Subjective opinions

3.2 Criteria for comparisons

For both studies, there were three main criteria (see Table 1) for comparison of the methods. The essential criterion was *the effectiveness of the evaluation*. This included the number of critical incidents (positive and negative comments of the test object) found, the severity of the problems found and the time all of the tasks required. The difficulty of the tasks and the category of the problems found were only used for the comparison of TPP and CBP. The other two criteria were *the perception and experience of both participants and the facilitator* with respect to the used methods.

For identifying usability problems, the definition of Virzi et al. [7] was used. They describe usability problems as following: (1) the user verbally indicated that something was unclear or confusing; (2) the user's utterances indicated a misconception regarding what was happening or what function a particular button may have had; or (3) the user's actions indicated an incorrect path. Based on [14–16], the usability problems found in study 1 were categorized in 4 categories: function, navigation, layout and wording. To the blind people, the term layout means just the arrangement of elements and their possibly spacing in Braille but not font size, color or typographic attributes which are common for sighted people. The usability problems are subdivided into minor, major and catastrophic problems. There are three factors, which influence the categorization, namely frequency, impact and persistence of the problems [17]. In study 2, the severity of the problems also depended on the user groups. For example, the same problem may be a minor issue for sighted people but a catastrophic problem for blind people.

4 Study 1: Comparison of tactile paper prototyping and computer-based prototyping

4.1 Approach of study 1

4.1.1 Participants

Five blind people took part in study 1. Visually impaired and sighted people were not recruited as control group because they are normally not familiar with Braille. In order to eliminate some influencing factors that could distort the test results, participants were required to have sufficient PC and Braille knowledge, have the same mother tongue, have no previous experiences with the test applications and have no other disability that may affect the task performance.

4.1.2 Materials

Prototypes were created based on two different applications so that the results would not be deemed application dependent. One application was a typical mobile and the other was a notebook system, which enabled users to create notebooks, add notes in each notebook and add keywords for each note. These two applications were selected because they had different complexity of user interfaces. The user interfaces of the mobile application were relatively simple, whereas the interfaces of the notebook system were relatively complex.

Both tactile paper prototypes and computer-based prototypes of the two applications were created. The two versions of the prototypes had the same functionalities. The tactile paper prototypes were first created using *Microsoft Office Visio 2007* and then printed by a Braille printer (*Tiger Professional ViewPlus*). In

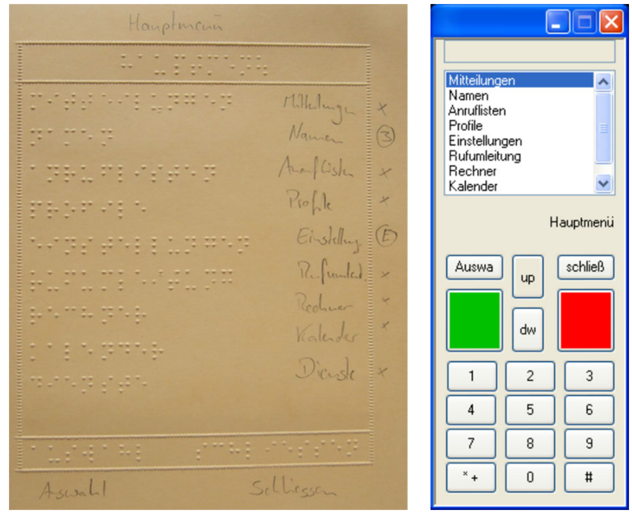

Fig. 1 Tactile paper prototype (*left*) and computer-based prototype (*right*) of the mobile application

addition, the tactile paper prototypes were labeled with pencil. Being sighted, it was helpful to understand what was expressed in Braille while conducting and analyzing the tests. The CBP version was created using *Microsoft Visual Basic 2008*. It was important for this study that the computer-based prototypes should be accessible for blind users. The control elements *AccessibleDescription*, *AccessibleName and AccessibleRole* in *Microsoft Visual Basic 2008* were used to provide screen reader information about elements. With respect to their accessibility, all of the computer-based prototypes created were tested and their accessibility ensured.

Figure 1 shows examples of the tactile paper and the computer-based prototypes of the mobile application.

In order to make the speech output of the mobile as realistic as possible, a screen reader for mobile devices was analyzed (*MobileSpeak 4* for Symbian S60). The speech output of the prototypes mirrored its output by controlling a screen reader for desktop (*JAWS 12*).

Figure 2 shows two examples of tactile paper prototypes of the notebook system. For the persistent part of the interface, a template was created (Fig. 2 left). In this way, it was not necessary to print a complete interface but only the parts that were changed if some elements were changed. Likewise, an example of computer-based prototype of the notebook system is shown in Fig. 3.

4.1.3 Procedure

Both TPP and CBP were conducted in the usability laboratory. As Fig. 4 shows, the desk equipped with a camera and several tactile paper prototypes was used for TPP and the one equipped with a computer was used for CBP.

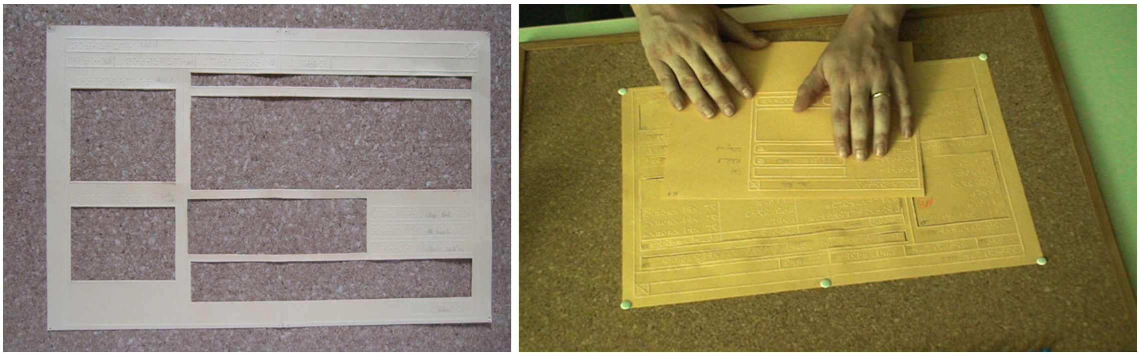

Fig. 2 Examples of tactile paper prototypes of the notebook system

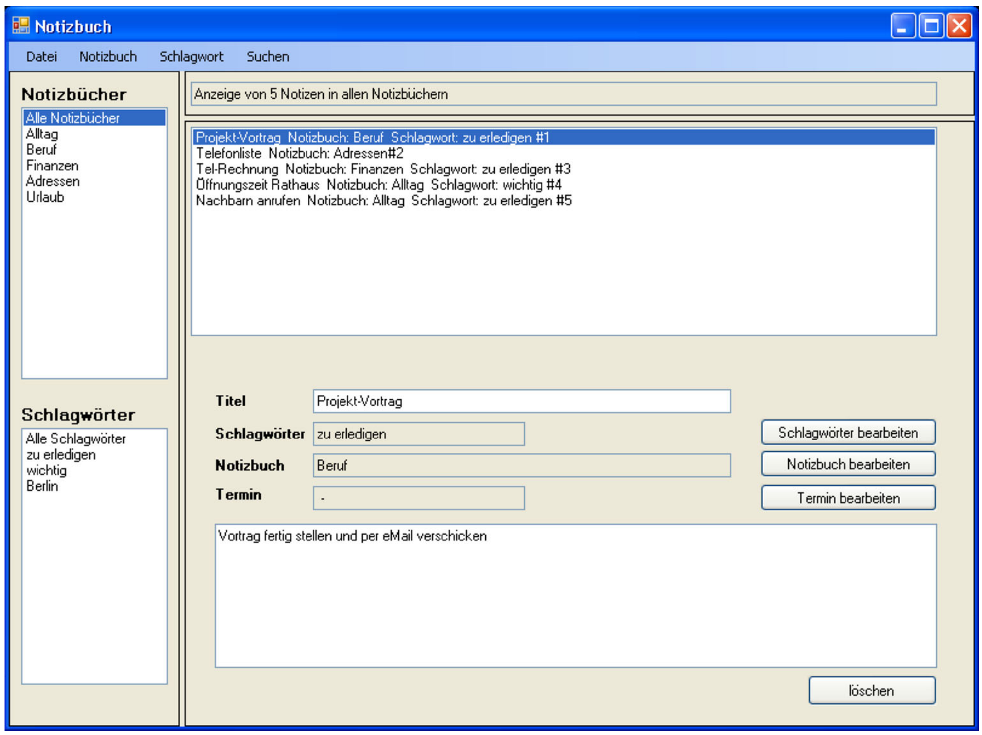

Fig. 3 Computer-based prototypes of the notebook system

To figure out which method blind people preferred, every participant (except participant 3, see Table 2) took part in two test sessions, each comprising of two usability tests.

At the beginning, the participants had few minutes to get an insight into the applications using prototypes and to feel comfortable. In the first session, two usability tests were ran. The first used one of the two methods (TPP or CBP) with one application (the mobile or the notebook application) and the second used the other method with the other application. After each test, participants had to complete a questionnaire. In the second session, they performed the same tasks with the applications though the methods were changed. At the end of the second session, they completed a questionnaire again to compare both methods. The order of the applications and methods used differed between the sessions by each participant to avoid possible order effects. With the prototypes of a mobile application, participants had to perform four tasks: (1) add a new contact; (2) edit an existing contact; (3) set an alarm; and (4) delete a contact. For the notebook application, there were five tasks: (1) create a new notebook; (2) find out the number of notes; (3) rename a keyword and create a new note; (4) create a new note and complete the date; and (5) delete a notebook.

4.2 Results of study 1

4.2.1 Effectiveness of the evaluation

4.2.1.1 Task completion time Figure 5 shows the time needed to complete all of the tasks. Regardless of the applications, participants using TPP needed considerably

Fig. 4 Test environment of TPP and CBP

Table 2 Procedure of study 1

Participant	1. Session				2. Session		
1	TPP(M)	Q	CBP(N)	Q	CBP(M)	TPP(N)	Q
2	CBP(M)	Q	TPP(N)	Q	TPP(M)	CBP(N)	Q
3	TPP(N)	Q	CBP(M)	Q	–	–	–
4	CBP(N)	Q	TPP(M)	Q	TPP(N)	CBP(M)	Q
5	CBP(M)	Q	TPP(N)	Q	TPP(M)	CBP(M)	Q

M mobile, *N* notebook and *Q* questionnaire

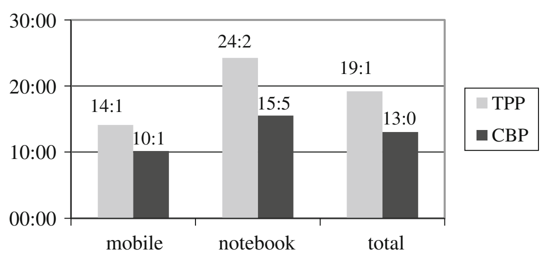

Fig. 5 Task completion time of TPP and CBP

more time than using CBP but not significantly (measured with the dependent *t* test). The time difference between the two methods for the mobile scenario can be traced back to the time-consuming reading of tactile paper prototypes. As the user interfaces of the notebook system were very complex, one participant encountered trouble to deal with the tactile paper prototypes. It is important to note that he did not have trouble with the computer-based prototypes of notebooks.

4.2.1.2 Number of usability problems found Considerably more problems were identified with CBP for the

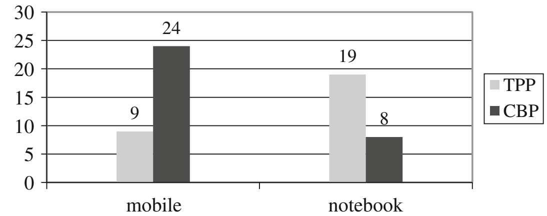

Fig. 6 Number of usability problems found with TPP and CBP

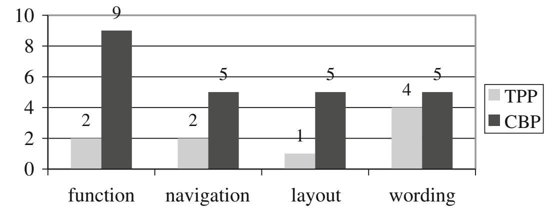

Fig. 7 Categories and number of usability problems found with TPP and CBP (mobile application)

mobile application, whereas with TPP, considerably more problems were identified for the notebook application (see Fig. 6).

The interfaces of the mobile application were relatively simple. As a result, participants felt very comfortable with the computer-based prototypes of the mobile device. They could get an overview of the interfaces in a rapid manner and navigated through the interfaces on their own. In contrast, the interfaces of the notebook were very complex. The participants could get a better overview of the notebook with the tactile paper prototypes and found more usability problems with the TPP method.

4.2.1.3 Categories of usability problems found The categories of usability problems found with both methods differed from each other in some degree. For the mobile application, more problems were found with CBP in each category (see Fig. 7). A considerably higher number of problems found with respect to function, layout and navigation.

For the notebook application, more problems were found with TPP in each category (see Fig. 8). Especially, six wording problems could be found with TPP, though none with CBP. Also, many more functional problems were found with TPP.

4.2.2 Perceptions and experiences of participants

4.2.2.1 Difficulty of the tasks After performing each of the tasks in session 1, participants were asked to rate the subjective difficulty of the task using a scale from 1 to 10

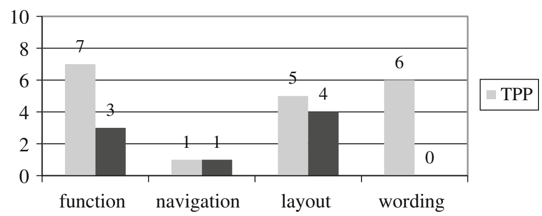

Fig. 8 Category and number of usability problems found with TPP and CBP (notebook application)

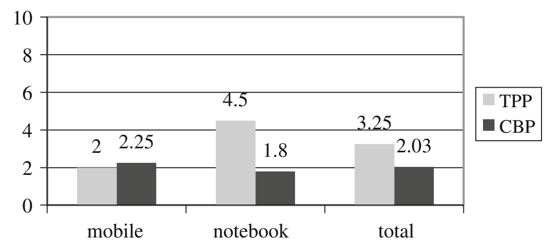

Fig. 9 Difficulty of tasks in session 1

(1 = very simple; 10 = very difficult). Figure 9 shows that in session 1, it is comparable for the mobile application using TPP and CBP while for the notebook interface, it is considered more difficult using TPP than CBP. For both, the difference is not statistically significant (measured with Wilcoxon rank-sum test).

In order to compare the two methods from the participants' point of view, they were asked to state which method they preferred for both applications.

4.2.2.2 Negative comments on the methods During the tests, participants did not criticize CBP at all. The drawbacks of TPP mentioned are summarized as follows:

- *Far from reality*. The interaction with tactile paper prototypes is too far from reality. It might be helpful for usability tests, but it does not match the interaction with a real application.
- *Search for something all the time*. For notebook, one participant indicated that it was nearly impossible for him to understand the structure because he had to search for the elements required all the time. He could not remember where changes occurred during the interaction. He said with CBP he only need to press the Tab key and did not care whether an element was on the left or right side. It is noted that this participant normally only uses speech output and not a Braille display.
- *Long response time*. In spite of a best possible preparation for TPP, the response time of the interactions was longer than using CBP.

4.2.2.3 Positive comments about these methods Four out of five participants indicated that the advantage of TPP is that they could get a better overview of the applications particularly for complex interfaces. Tactile paper prototypes permit them to understand the structure of the application and the context of individual element. One participant commented that this was especially helpful to congenitally blind people. In contrast to TPP, the interaction with computer-based prototypes is more intuitive because it is in the same manner as with the real application.

One participant suggested a combination of TPP and CBT for usability tests. This means that participants work with computer-based prototypes and they also get tactile paper prototypes of the main user interfaces.

4.2.3 Perceptions and experiences of facilitator

The development of the two versions of prototypes required different efforts. First, the prototypes must be accessible for blind users. Therefore, the development of the computer-based prototypes required good programming skills, and the development environment is also strongly constrained. Second, creating computer-based prototypes is more time-consuming than tactile paper prototypes. From the authors' own experience, creating computer-based prototypes takes almost twice as long as tactile paper prototypes for the same application.

4.3 Recommendations for practitioners

Both TPP and CBP have advantages and problems for usability testing with blind users in the early stages. Tactile paper prototypes provide a better overview of the application. However, blind users do not get this better overview while interacting with a real application. Consequently, it could be possible that the better overview has impaired the identification of usability problems concerning navigation and arrangement of elements. Communication with computer-based prototypes is closer to reality; however, there are two disruptive factors in CBP: accessibility issues of prototypes and screen reader. The prototypes have to be implemented with some programming language and should be accessible for blind people, thereby implementation mistakes could arise. Such mistakes interfere with the main aim of usability testing in the early stages of development, which is to identify problems of concept but not of implementation. Furthermore, to communicate with the computer-based prototypes, blind users have to use assistive technologies such as screen readers and Braille displays. How well does screen reader work and how good is the user's ability to use screen reader are also factors that have effects on usability testing.

This study could not identify which method is more effective in identifying usability problems. Both methods can be used for usability testing with blind users in the early stages. It is suggested that for applications with simple user interfaces, computer-based prototypes should be used. For applications with complex user interfaces, tactile paper prototypes are recommended under the condition that the prototypes are accessible for blind users.

5 Study 2: Comparison of local and synchronous remote testing with blind and visually impaired people

5.1 Approach of study 2

5.1.1 Participants

The main aim of study 2 is the comparison of synchronous remote and local tests with blind and visually impaired people. It is also reasonable to include sighted people as a control group as well. Twenty participants took part in the evaluation of whom eight were blind, six visually impaired (vision: 2, 2, 5, 5, 20 and 30 %) and six sighted (see Table 3). Half of each user group took part in the local test and the other half in the synchronous remote test. To figure out which method they prefer, nine of them participated in both approaches.

The participants were required to have sufficient PC knowledge, the same mother tongue, no previous experience with the test program and no disability that may affect the task performance. To get the participants' profiles, email or phone interviews were performed prior to the tests.

5.1.2 Test object and tasks

The test object was an Internet phone software. When choosing this test object, the accessibility of the object was taken into consideration. It needs not be absolutely accessible though blind people had to be able to work with it to some degree. The participants were asked to perform seven

Table 3 The makeup of participants of local and remote tests

	Local test			Remote test		
	m	*f*	Age ∅	*m*	*f*	Age ∅
Blind	3	1	45	3	1	44
Visually impaired	2	1	25	3	0	37
Sighted	2	1	25	2	1	29

m male, *f* female, *age* ∅ average age

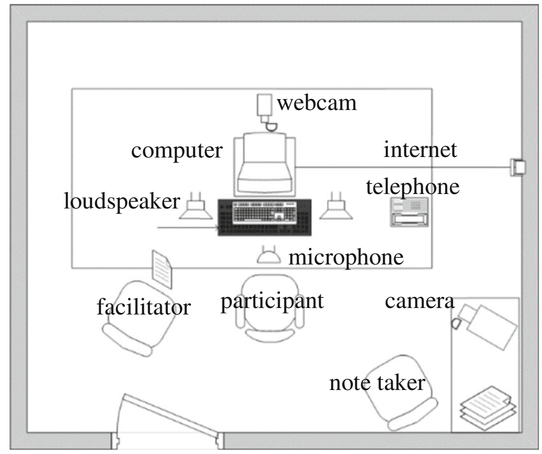

Fig. 10 Structure of the local test environment

tasks. These were as follows: (1) installing the test object; (2) logging in; (3) adding a contact; (4) calling the contact; (5) making a direct call; (6) modifying the contact; and (7) logging out. The participants performed these tasks in both local and remote tests.

5.1.3 Test environment

5.1.3.1 Local testing The local tests were conducted in the usability laboratory. Figure 10 shows the structure of the local test environment.

The equipment had the following characteristics: a computer (*Intel Core 2 Duo 1,66 Ghz, Windows XP Professional SP3, IE 8*), screen (26 inch), Braille display (*Braillex EL 80s*™), screen reader (*JAWS*™ 6.2) and screen magnifier (*Magnify*). *Camtasia Studio*™ (version 5) was used as screen recording software.

5.1.3.2 Synchronous remote testing The synchronous remote tests with blind and visually impaired people were a challenge for the researchers. This was due to the high technical requirements as illustrated and detailed as follows.

For the synchronous communication between participants and the facilitator, a screen sharing tool was necessary. This tool had to meet several conditions: It should have screen sharing, voice connection and it should be accessible and usable for blind and visually impaired people. Although there are many application sharing tools, most of them are not suitable for blind and visually impaired people. Nine screen sharing tools were tested, and *WebEx*™ from Cisco Systems was selected. It is accessible and can be used more easily by blind people than the others. In addition, it provides not only screen sharing but also screen recording. So in these tests, it was not necessary

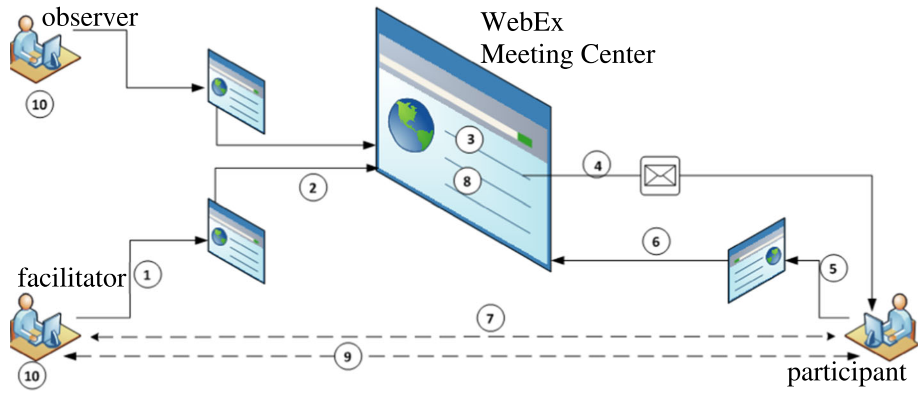

Fig. 11 Setup of a synchronous remote test via WebEx™

to use additional screen recording software. Figure 11 shows the setup process of synchronous remote test via *WebEx™*. There were ten steps to follow in order to complete it.

1. F (facilitator): log in, define topic and time
2. Conference is opened
3. F: distribute recording privilege
4. F: send invitation email
5. P (participant): click on the link in email and enter the welcome site
6. P: log in with name and email address
7. Communication between F and P via VoIP possible
8. F: allow P moderator privilege
9. P: can share screen with another member
10. Screen recording possible

The transmission of screen reader output from the participants to the facilitator was very important. On the one hand, the facilitator knew whether the participants were currently listening to the screen reader and therefore avoided unnecessary disturbances. On the other hand, he could follow the users' performance effectively. The *WebEx™* could not transmit screen reader output directly. It had to be made audible via loudspeaker and transmitted via microphone to the facilitator. Moreover, the transmission of screen reader output required that participants possess a microphone and a loudspeaker. A headset was not suited for the test.

Due to the visual disability, it is a challenge for many blind people to install software by themselves. For that purpose, a tool called *Autoconfigurator* was created to help with installation of *WebEx™*. This tool effectively automated several installation steps. Just in case something went wrong during the test, it was also possible to control the participants' computer remotely, if participants wanted the facilitator to intervene. This remote control was also provided by *WebEx™*.

Table 4 Task completion time of local and synchronous remote tests

	Local test/min (mean)	Syn. remote test/min (mean)	t	Critical value ($\alpha = 0.05$)	Sign. diff
Blind	34.38	23.58	3.59	2.45	Yes
Visually impaired	13.04	14.08	−0.14	2.78	No
Sighted	6.17	8.59	−1.58	2.78	No

5.2 Results of study 2

Altogether, 29 test sessions were conducted with 20 participants (14 local tests and 15 remote tests). As mentioned in Sect. 5.1.1, to figure out which methods they preferred, nine of them were confronted with both methods and performed the same tasks. This section presents the results of the comparison with respect to the three criteria (see Sect. 3.2).

5.2.1 Effectiveness of the evaluation

5.2.1.1 Task completion time The task completion time refers only to the time for performing all of the tasks exclusive of the time for setup and dismantling. For testing the statistical significance of the time, t test, which is a parametric statistical test, was used.

Table 4 shows that there is a significant difference for blind participants though not for visually impaired and sighted participants. It is particularly noticeable that blind participants in a local test needed about 10 min longer than in a synchronous remote test. This is essentially due to the blind participants' unfamiliarity with the equipment used for the local tests. For remote tests, they worked with their own familiar equipments, and therefore, they were fast.

5.2.1.2 Number and severity of usability problems found Figure 12 shows that the number of usability problems found in different categories with both methods

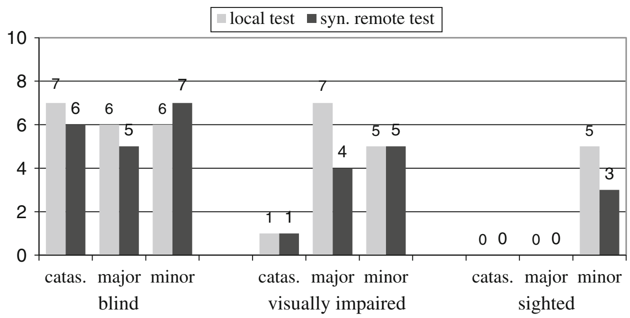

Fig. 12 Number and severity of problems found with local and remote test

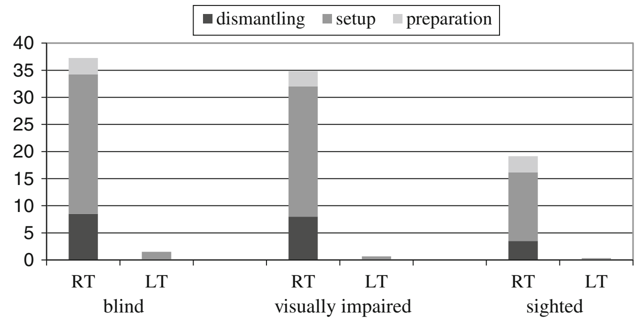

Fig. 13 The time required by participants for preparing, setting up and dismantling the tests (*RT* remote test, *LT* local test)

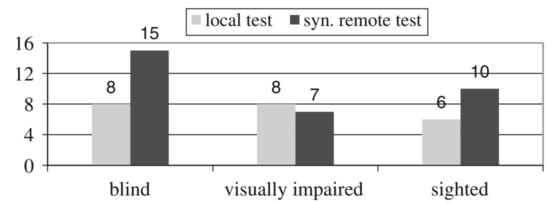

Fig. 14 Number of negative comments on local and synchronous remote tests

is comparable among blind, visually impaired and also sighted participants.

This confirms the results from [1, 4, 9]. For blind participants, several catastrophic problems arose during local tests. The reason for it is that the blind participants are not familiar with the assistive technologies in the laboratory. For example, in a local test, a blind participant had to work without Braille display because he could not deal with it at all. As a result, two tasks could not be solved successfully.

5.2.2 Perceptions and experiences of participants

To compare the two methods, participants were interviewed after the tests. The comments of the nine participants who participated in both approaches were analyzed separately.

5.2.2.1 Time for preparation, setup and dismantling of the tests

- *Time for test preparation:* In synchronous remote tests, participants had to complete tasks such as activating javascript, preparing the microphone and loudspeaker and installing the test object.
- *Time for setting up the test:* Set up the *WebEx*™.
- *Time for dismantling*: As in synchronous remote tests, participants had to uninstall the test object and other additional tools after the test.

Preparation and dismantling time were only required in synchronous remote tests. Participants who took part in the remote tests reported how much time they needed for the preparation and dismantling. The time for setup could be measured during the tests. As Fig. 13 shows, it is obvious that the time required in synchronous remote tests was considerably longer than in local tests by all user groups. Blind and visually impaired participants needed 30–35 min for preparation, setup and dismantling in synchronous remote tests while sighted participants only needed approximately half this time. It is noticeable that, even though the visually impaired participants have some sight, the time they needed to complete these tasks was not significantly less than the time needed by blind participants.

5.2.2.2 Negative comments on the methods Figure 14 shows that the blind participants had given the most (15) negative comments on synchronous remote tests. The frequently mentioned drawbacks of synchronous remote testing are summarized as follows:

- It took too much effort to prepare for the test (regarding preparation of hardware and software).
- The setup of *WebEx*™ was too complex. There were too many steps to do. As Fig. 13 shows, it took 25.75 min on average.
- The screen reader was too loud, which made it hard to concentrate on the tasks. As mentioned, the screen reader must be output via a loudspeaker so that it can be transmitted via microphone to the facilitator.
- Security and privacy concerns. Users felt insecure for three reasons. First, they had to install unknown software (*WebEx*™, *ShowKeyboard* and *test object*) on their own computers. Second, another person (the facilitator) observed their computers. Third, they felt uncomfortable that their computers were controlled remotely by the facilitator.

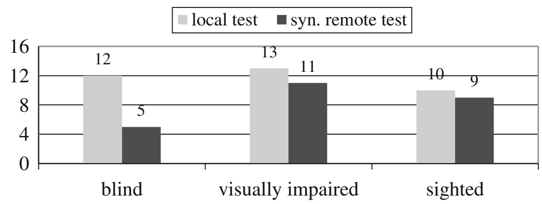

Fig. 15 Number of positive comments on local and synchronous remote tests

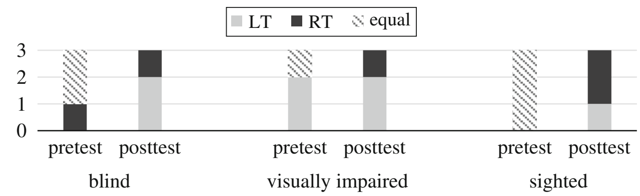

Fig. 16 Preference of the participants pre- and post-test

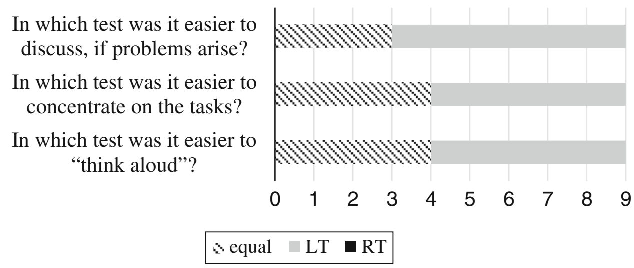

Fig. 17 Answers from the nine participants who took part in both approaches

Instructions made by the facilitator helped only partially when participants encountered problems. They had to setup all of the things on their own.

By contrast, blind participants in the local tests experienced other problems. They criticized local tests as follows: (1) unfamiliar screen reader and Braille display; (2) uncomfortable exam feeling; and (3) travel cost.

5.2.2.3 Positive comments on the methods Figure 15 shows that blind participants had given the fewest positive comments on synchronous remote tests.

They included: preference for their own equipment and remote control was helpful. The most frequent positive comments on local tests from blind participants were as follows: (1) no setup required; (2) face to face communication with facilitator made it easier; and (3) exam feeling was good, as performance was more serious in an exam situation. However, some blind participants were also uncomfortable with remote control and the exam feeling.

The visually impaired participants appreciated the big screen (26 inch) and the 80-cell Braille display (as commented by a severely visually impaired participant). Two visually impaired participants in synchronous remote tests felt that the *Autoconfigurator* (which automated several install steps) was helpful.

5.2.2.4 Comments from the nine participants Nine participants (three blind (b), three visually impaired (vi) and three sighted (s)) took part in both methods. In order to figure out which approach the nine participants preferred, they were interviewed after the second round of tests. Before participating in the first method, they were asked in which method they wanted to participate. The same question was asked following the second method. As Fig. 16 shows, after both methods, two of the three blind participants preferred the local test (LT). The same applied to visually impaired participants. However, two of the three sighted participants preferred the synchronous remote test (RT). The participants who opted for the synchronous remote test pointed out that they preferred it only if the setup of the test becomes easier.

Furthermore, participants answered other questions (see [1]) with respect to both methods (see Fig. 17). It is obvious that there was no answer in support of the synchronous remote test even though the effectiveness of both methods was comparable.

5.2.3 Perception and experience of facilitator

After 29 synchronous remote tests and local tests altogether, the facilitator gained some experiences with both methods. Concerning the practicability of these two methods by blind and visually impaired people, the facilitator compared the approaches as follows:

5.2.3.1 Advantages of the synchronous remote testing Compared with the local test, the synchronous remote test has two main advantages.

- *Realistic test environment.* Blind and visually impaired participants can use their own assistive technologies. Therefore, they can perform efficiently. For a local test, not all types of assistive technologies are available (e.g., different Braille display or screen reader). Moreover, in a local test, it is difficult to configure the assistive technologies for each participant as they wish.
- *Easier to recruit participants.* Neither participant nor facilitator is required to travel, which saves time and effort.

5.2.3.2 Problems with the synchronous remote testing There are several problems performing a synchronous remote test with blind people.

1. Insufficient observation.
   Brush et al. [1] found that it is as easy to observe issues in the remote test as in the local test with sighted participants. However, observing a test with blind and visually impaired participants is insufficient in several ways.

   - *No observation of Braille display.* Observation of haptic interaction on Braille display requires other methods than video. None of the synchronous systems inspected are capable of such a feature. Therefore, it is impossible to follow the Braille display. For example, it is unclear which routing keys are activated, and hence, errors cannot be identified. Furthermore, the facilitator is unable to know when the participants are reading on the Braille displays. In one test, the facilitator spoke to the participant while he was reading on Braille display. Thus, his concentration on tasks was disturbed.
   - *No observation of facial expression.* A video connection is not possible with blind participants.
   - *Insufficient observation of actual screen configurations.* This problem was caused due to the screen sharing program *WebEx*^{TM}. In one test, a visually impaired participant had the inverse color scheme on the screen, but this could not be correctly displayed for the facilitator.
   - *Delay of sound and picture transmissions.* Due to this issue, some user actions could not be recorded.
   - *No direct transmission of screen reader output.* Environmental sounds and think-aloud comments were also recorded.
   - *Problems resulting from configuration of the computer by participants.* In one test, a visually impaired participant had the mouse cursor so big on the screen so that it hid a part of the test program.

2. High technical requirements for participants.
   The basic requirements were that participants had to have a computer with appropriate software and hardware as well as fast Internet access. In addition to this, there were the requirements of the Web meeting tool (*WebEx*^{TM}). Furthermore, a loudspeaker (headset, if the participant did not need a screen reader) and microphone had to be available. To be able to participate in the test, one visually impaired participant had to buy a headset. Sometimes, participants had to possess administrator rights on their computers to install a program for the test. Another requirement was that all of the software and hardware should be fully functional. For example, a misconfigured sound system or a defective loudspeaker could lead to a considerable delay of the test.

3. High computer skills required by the participants.

Participants had to deal with many technical procedures on their own, for example, installation of some software, and dealing with any problems themselves.

4. High requirements of the facilitator.

Due to the insufficient observation during the test, sometimes the user actions could not be followed effectively. In any case, the job of the facilitator required high interpretation skills. The facilitator should be well grounded in computer configuration particularly with assistive technologies. In addition, the manner in which a facilitator dealt with blind participants was also highly important.

5. Security and privacy concerns of participants (see 5.2.2).
6. A high effort level for participants and facilitators due to the preparation, setup and dismantling of the test.
7. External interruptions experienced during tests by participants such as a call or a visit.

5.3 Recommendations for practitioners

Founded on the experiences of the participants and the facilitator, the following recommendations for synchronous remote testing particularly with blind and visually impaired people can be suggested. On the one hand, there are recommendations for technical improvements in setup especially on the screen sharing tool. On the other hand, suggestions for organizational improvements in the planning and preparation of a synchronous remote test were made.

5.3.1 Technical improvements

Several criticisms from participants were due to the screen sharing tool *WebEx*^{TM}. Furthermore, insufficient observation was one of the essential drawbacks of the synchronous remote test. For this reason, a new screen sharing tool should be developed. In the following, the proposed concept for such a tool is presented.

The tool has to meet the following requirements and some of them will be illustrated in detail.

1. *It should be executable on each operating system with minimum configuration.*
2. *Minimum effort should be required to install it.* The installation should be limited to receiving an email

with a download link from the facilitator, then downloading and starting the tool.

3. *It should be usable and accessible for blind and visually impaired people.* The tool should be as easy to use as possible. The interaction between participants and the tool should be limited to easy actions.
4. *The following functionalities should be available.*
   - *Internet speed test before transmitting the first data.* Depending on the Internet speed detected, it will be configured for optimal transmission.
   - *Remote control.* In the case that technical problems arise on participants' side, the facilitator can help them via remote control. However, this should not be carried out before participants allow it.
   - *Transmission and recording of the screen.*
   - *Transmission of voice and screen reader output from participants and recording all of the acoustic signals*. The acoustic signals are the voice of the participant and the facilitator as well as the screen reader output.
   - *Transmission and recording of Braille display, mouse and keyboard signals*.
5. *All of the data should be transmitted in real time and saved for analysis afterward.*

5.3.2 Organizational improvements

1. *Extensive understanding of participants.* A facilitator should be aware of which disability the participant has and which assistive technologies the participant uses. It is also important to understand how blind and visually impaired people deal with assistive technologies.
2. *Minimizing security concerns and gaining trust.* Trust is a precondition in order for someone to take part in a test. When recruiting participants, they need to know about the test objective and test procedure, particularly about the software that they have to install. They should be assured there would be support provided to uninstall the software and restore the initial condition of their computers.
3. *Pilot tests with blind or visually impaired people but not with sighted people.*
4. *Minimizing interference with privacy.* Before participants share their screen, one should give them enough time to close other opened applications such as emails. With this, the participants feel less observed.

6 Conclusion and outlook

This paper has presented two studies. In study 1, a comparison of tactile paper prototyping and computer-based prototyping with blind participants was made. Each method has advantages and problems. The main problem of tactile paper prototyping is that the interaction with tactile paper prototypes is very far from reality and its main advantage is that blind people could get a better overview of an application. In contrast, when using computer-based prototypes, the interaction is more intuitive for blind people. The creation of computer-based prototypes is, however, more time-consuming and required good programming skills. The developer should be aware of developing software. It is absolutely necessary that the prototypes are tested with screen reader according to accessibility. It is suggested that for applications with simple user interfaces, computer-based prototypes should be used. For applications with complex user interfaces, tactile paper prototypes are recommended.

In study 2, local and synchronous remote tests with blind and visually impaired participants were compared. Notably, the technical background plays a major role, and the results cannot easily be generalized. The numbers of positive and negative critical incidents found in different categories with both approaches are comparable. In terms of task completion time, there is a significant difference for blind participants though not for visually impaired and sighted people. Most of the blind and visually impaired participants prefer a local test. The main reason for this preference is that it takes too much effort to prepare, set up and dismantle the synchronous remote test. Therefore, recommendations have been provided with regard to technical and organizational improvements for conducting synchronous remote tests with blind users.

However, both methods have only been compared on a desktop Internet phone program. Web-based applications require no setup and sharing them between a facilitator and a user is technically less demanding. Acceptance by users may be much higher as the users' computers may be less affected. A proxy server may create the necessary communication and may shield the users from most analysis tools. Future work will show how assistive technologies can cope with Web 2.0 technologies while allowing inspection of the real interaction by the user. In this context, the minimum requirements will have to be identified and assistive technologies have to provide for better usability analysis by integrating user operations related to the use of assistive technologies. Various exploration activities degrade usability by introducing inefficiency, errors and making the user less satisfied.

Acknowledgments The authors would like to thank the participants and collaborators for spending their time on this study.

References

1. Mayhew, D.: The Usability Engineering Lifecycle—A practitioner's Handbook For User Interface Design. Morgan Kaufmann, San Francisco (1999). ISBN 1558605614
2. Miao, M., Köhlmann, W., Schiewe, M., Weber, G.: Tactile paper prototyping with blind subjects. In: Haptic and Audio Interaction Design (HAID), Dresden, pp. 81–90. Springer, Berlin, Heidelberg (2009)
3. Magnusson, C., Brewster, S. (eds.): Guidelines for Haptic Lo-Fi Prototyping. Workshop, NordiCHI 2008, Lund, Sweden, 19 Now. (2008)
4. Rudd, J., Stern, K., Isensee, S.: Low vs. high-fidelity prototyping debate. In: Interactions 3, Nr. 1, pp. 76–85 (1996)
5. ISO 9241-171: Ergonomics of Human-System Interaction—Part 171: Guidance on Software Accessibility (2009)
6. Web Content Accessibility Guidelines (WCAG) 2.0, W3C Recommendation 11 December 2008. http://www.w3.org/TR/WCAG20/
7. Virzi, R.A., Sokolov, J.L., Karis, D.: Usability problem identification using both low- and high-fidelity prototypes. In: CHI'96: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 236–243. New York, USA (1996)
8. Miriam, W., James, A.L.: High-fidelity or low-fidelity, paper or computer? Choosing attributes when testing web prototypes. In: Proceedings of the Human Factors and Ergonomics Society, pp. 661–665 (2002)
9. Petrie, H., Hamilton, F., King, N. et al.: Remote usability evaluation with disabled people. In: Proceedings of CHI, pp. 1133–1141. ACM Press (2006)
10. Hartson, H.R., Castillo J.C., Kelso, J., Kamler, J., Neale, W.C.: Remote evaluation: the network as an extension of the usability laboratory. In: Proceedings of CHI Conference on Human Factors in Computing Systems ACM, pp. 228–235. New York (1996)
11. Selvaraj, P.: Comparative Study of Synchronous Remote and Traditional In-Lab Usability Evaluation Methods. Master of Science Thesis. The Virginia Polytechnic Institute and State University (2004)
12. Hartson, H.R., Castillo, J.C.: Remote evaluation for post-deployment usability improvement. Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 22–29. L'Aquila, Italy (1998)
13. Brush, A.J.B., Ames, M., Davis, J.: A comparison of synchronous remote and local usability studies for an expert interface. Proceeding of CHI, pp. 1179–1182. Austria, Vienna (2004)
14. Johnson, J.: Don'ts and Do's for Software Developers and Web Designers, 1st edn. Morgan Kaufmann (2000)
15. Nielsen, J.: Usability Engineering. Morgen Kaufmann (1993)
16. Sefelin, R.: Comparison of Paper- and Computer Based Low-Fidelity Prototypes/Center for Usability Research & Engineering (CURE). Research report (2002)
17. Nielsen, J., Loranger, H.: Prioritizing Web Usability. New Riders Press, Berkeley (2006)