The Relationship between Accessibility and Usability of Websites

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

Omar Kheir
HCI Group, Department of Computer Science
University of York,
York YO10 5DD, United Kingdom
Kheir@cs.york.ac.uk

ABSTRACT

Accessibility and usability are well established concepts for user interfaces and websites. Usability is precisely defined, but there are different approaches to accessibility. In addition, different possible relationships could exist between problems encountered by disabled and non-disabled users, yet little empirical data have been gathered on this question. Guidelines for accessibility and usability of websites provide ratings of the importance of problems for users, yet little empirical data have been gathered to validate these ratings. A study investigated the accessibility of two websites with 6 disabled (blind) and 6 non-disabled (sighted) people. Problems encountered by the two groups comprised two intersecting sets, with approximately 15% overlap. For one of the two websites, blind people rated problems significantly more severely than sighted people. There was high agreement between participants as to the severity of problems, and agreement between participants and researchers. However, there was no significant agreement between either participants or researchers and the importance/priority ratings provided by accessibility and usability guidelines. Practical and theoretical implications of these results are discussed.

Author Keywords

Accessibility, usability, guidelines, user testing, severity ratings.

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

Both accessibility and usability are now well-established concepts used in relation to user interfaces and more recently to websites. For usability, there is a precise and widely accepted definition now provided by ISO 9241: “the extent to which a product [or website] can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [8]. For accessibility, the situation is less clear [9]. The Web Accessibility Initiative (WAI), founded by the World Wide Web Consortium (W3C) in 1997 to promote the accessibility of the Web, gives a widely accepted general definition of Web accessibility as “people with disabilities can use the Web … more specifically [they] can perceive, understand, navigate, and interact with the Web” [7]. This might be termed the “usability for people with disabilities” or “usable accessibility” [1, 17] definition of Web accessibility, as it appears to be promoting a user-based definition similar to that provided by ISO 9241. However, rather than defining more precise user-based criteria, WAI has promoted conformance to the Web Content Accessibility Guidelines (WCAG) [3] as the criteria for achieving and measuring accessibility (when referring to WCAG, this paper will mean WCAG version 1, as version 2 is currently still in draft). This might be termed the “technical accessibility” definition of Web accessibility, as it relies largely on meeting technical criteria in the underlying Web code [2, 15].

The relationship between the usable accessibility definition and the technical accessibility definition is unclear. Little empirical data have been gathered to show that websites that achieve higher conformance to WCAG are also more usable by people with disabilities and what the criteria for usability for people with disabilities should be. For example, the study of 1000 websites conducted for the Disability Rights Commission [5] found no significant relationship between conformance with WCAG and a number of measures of user performance and satisfaction for five different categories of disabled people. We believe that the ultimate criteria for accessibility should be user-based and we can adapt the ISO 9241 definition for this purpose: the extent to which a product/website can be used by specified users with specified disabilities to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. In the context of the Web, one of the interesting questions is the contribution of technical accessibility to achieving these criteria.

Usability can also be defined as the lack of usability problems in using a product or website; this is important, as in measuring usability, one can either measure effectiveness, efficiency and so on, or one can measure the problems that a user encounters or might encounter (for example usability inspection methods tend to concentrate on identifying usability problems [4]). Similarly, accessibility can be defined as the lack of accessibility problems. But this is not the same as saying that usability problems are only encountered by people without disabilities and accessibility problems are only encountered by users with disabilities. The relationship between accessibility and usability and accessibility and usability problems are rarely explicitly analysed, either in the context of the Web or other computer-based systems.

Thatcher *et al* [17] propose that accessibility is a subset of usability, suggesting that accessibility problems are particular types of usability problems. However they also state that *usability* problems affect all users equally, regardless of ability or disability, whereas *accessibility* problems hinder access for people with disabilities and put people with disabilities at a disadvantage relative to people without disabilities. These latter statements suggest a more complex relationship between accessibility and usability than the former being a subset of the latter.

Shneiderman [13, 14] proposes "universal usability" as a term to encompass both accessibility and usability, which he defines as "having more than 90% of households as successful users of information and communication technologies at least once a week" (p85). Shneiderman [13] notes that "access is not sufficient to ensure successful usage", suggesting that accessibility is a first but not sufficient step towards universal usability, but does not analyse the relationship between the two concepts further.

If we consider the problems that disabled and non-disabled people respectively encounter in using a website, we can propose that a number of possible relationships might hold between these two sets of problems.

Firstly, the problems might be two distinct, non-intersecting sets, meaning that there are no problems that disabled people encounter that are also encountered by non-disabled people and vice versa (so accessibility problems would be those encountered by disabled people and usability problems those encountered by non-disabled people). In fact, this is the way accessibility and usability are usually dealt with in the development of most websites. The processes for conceptualizing, assessing and removing problems encountered by each group of users are completely distinct, most likely dealt with by different individuals within an organization, at different times in the development process. The idea of dealing with the two types of problems in a unified process, either via the use of guidelines or user testing, is rarely encountered in web or interface development practice.

Secondly, as noted above, Thatcher *et al* [17] propose that accessibility problems (which we take them to mean problems encountered by disabled people relevant to their disability and assistive technologies) might be a subset of usability problems. This definition is attractive in that accessibility can be dealt with as part of the usability evaluations process. But it also suggests the possibility that some problems that we typically think of as accessibility problems also affect non-disabled users. For example, providing an informative set of headings can make a webpage much more usable for blind people using screenreading technologies, but it is also very helpful for non-disabled people as well. However, some problems appear to only affect people with specific disabilities. For example, having a "submit" button with green text on a red background will not pose any problems for people with full color vision, but will be a catastrophic problem for people with red-green color vision deficiency. So not all accessibility problems affect non-disabled users, and are therefore not within the scope of usability problems.

Thirdly, Shneiderman's concept of universal usability might be thought of as expanding the scope of what we traditionally think of as usability to include disabled users, so that usability problems become a subset of accessibility problems. This can account for the color vision problem discussed above, as some accessibility problems are beyond the scope of usability. But this formulation suggests that all usability problems are within the scope of accessibility, meaning that people with disabilities encounter all the same problems that people without disabilities encounter.

Finally, we believe that accessibility and usability problems can be seen as two overlapping sets, which would include three categories:

- Problems that only affect disabled people; these can be termed "pure accessibility" problems;
- Problems that only affect non-disabled people; these can be termed "pure usability" problems;
- Problems that affect both disabled and non-disabled people; these can be termed "universal usability" problems.

It should also be noted that one could expand this analysis, taking users with each specific disability separately, as the problems encountered by the different disability groups can have a range of relationships with each other. Nonetheless, all these possible basic relationships between the problem sets highlight useful aspects of the situation. However, we lack empirical data on the *actual* breakdown of problems into these sets on websites. This paper sets out to investigate these relationships.

In addition, there is the question of whether a particular problem affects disabled and non-disabled people equally. As noted above, Thatcher *et al* [17] define usability problems as those which affect disabled and non-disabled people equally, whereas accessibility problems hinder access to a website for disabled people. However, in our recent research [e.g. 5, 6, 12] we have often noticed that disabled and non-disabled people often encounter the same problems, but are affected by them differently, which leads to a somewhat different analysis: some problems appear negligible or minor to non-disabled people, but pose major barriers for disabled people, or certain specific groups of disabled people. Thus, problems encountered by non-disabled people (usability problems) appear be *amplified* or intensified for people with disabilities. This is a particularly interesting effect. It suggests that usability problems could be detected more easily by conducting evaluations with disabled people rather than with the non-disabled people currently used in usability evaluations. However, this anecdotal evidence needs a sound empirical basis, so this relationship will also be investigated in this paper.

Guidelines for both accessibility and usability attempt to quantify the importance of particular problems. For example, the usability guidelines originally developed by the US National Cancer Institute and then extended by the Department of Health and Human Sciences (henceworth, HHS guidelines) [10] provide two five point ratings for each guideline: the "relative importance" of the guideline, and the "strength of evidence" used in making that judgement. The WCAG divides the checkpoints (sub-sections of the more general guidelines) into three groups:

Priority 1: a Web content developer **must** satisfy this checkpoint. Otherwise, one or more groups will find it impossible to access information in the document. Satisfying this checkpoint is a basic requirement for some groups to be able to use Web documents.

Priority 2: A Web content developer **should** satisfy this checkpoint. Otherwise, one or more groups will find it difficult to access information in the document. Satisfying this checkpoint will remove significant barrier to accessing Web documents.

Priority 3: A Web content developer **may** address this checkpoint. Otherwise, one or more groups will find it somewhat difficult to access information in the document. Satisfying this checkpoint will improve access to Web documents.

The WCAG documentation states that the priority levels were assigned by the Working Group (the consultation group for the development of the guidelines), but does not elaborate on the process or evidence used to define the priorities. Harrison and Petrie [6] investigated the relationship between the WCAG priority levels for a set of problems encountered by disabled and non-disabled Web users and the users' own ratings of the severity of the problems and an expert's rating of the problems. Although there was a significant correlation between the users' and the expert's ratings, there was no significant correlation between either of these two sets of ratings and the WCAG priority levels. However, this was a small study with only two blind and two dyslexic participants and one expert. Further evidence on the relationship between the importance ratings provided by the guidelines and user experience is particularly interesting and important because of the practical and legal relevance of these ratings. With both the accessibility and the usability guidelines, developers may use these ratings to prioritize their work in improving a website. In the case of the accessibility guidelines, legislation and directives in a number of countries requires websites to meet guidelines with Priority 1 and 2 ratings.

The above discussion a number of issues about the relationship between accessibility and usability of websites and the importance ratings of accessibility and usability problems provided by guidelines. This study will explore these issues with a user-based study of two websites. Given the extensive data collection required, it was decided to concentrate on the experience of blind people interacting with the Web using screenreaders, as they encounter the most difficulties in using the Web [5], and compare their experience with that of a matched group of non-disabled people.

The following research questions were investigated:

1. What is the nature of the relationship of the problems encountered by non-disabled (sighted) people and disabled (in this case blind screenreader users) people?
2. If the same problems are encountered by both blind and sighted people, are they more severe for the blind people than the sighted people?

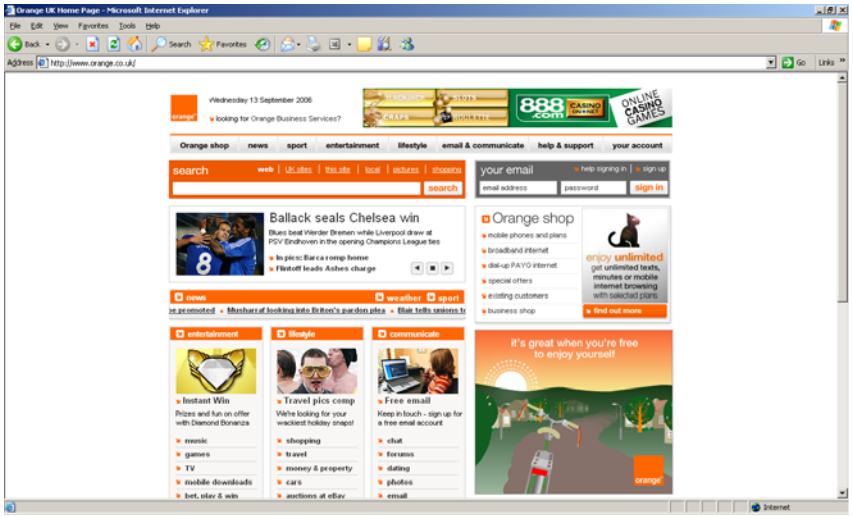
In order to answer this second research question, the relationship between different measures of severity of accessibility and usability problems needs to be further investigated, an area also of interest in itself. So the third research question was:

3. What is the relationship between users' ratings of the severity of problems, expert's ratings and the ratings provided by accessibility and usability guidelines?

METHOD

Participants

Six sighted and six blind participants undertook the evaluation study. The six sighted participants were 5 men and 1 woman with a median age of 30. The six blind participants were 4 men and 2 women with a median age of 40. The two groups of participants were matched as far as possible on age, gender and most importantly, general computer and Internet experience and expertise. Participants were asked to rate their computer experience and expertise on 5 point scales. They were also asked to estimate how many hours a week they spend using the Web.



Orange UK Home Page - Microsoft Internet Explorer

Address http://www.orange.co.uk/

Wednesday 13 September 2006

looking for Orange Business Services?

Ballack seals Chelsea win

Blues beat Werder Bremen while Liverpool draw at PSV Eindhoven in the opening Champions League ties

In pics: Barca romp home

Flintoff leads Ashes charge

...e promoted • Musharraf looking into Briton's pardon plea • Blair tells unions t...

Instant Win

Prizes and fun on offer with Diamond Bonanza

Travel pics comp

We're looking for your wackiest holiday snaps!

Free email

Keep in touch - sign up for a free email account

enjoy unlimited

get unlimited texts, minutes or mobile internet browsing with selected plans

it's great when you're free to enjoy yourself

Figure 1: Orange website

Table 1 shows the ranges and median values on each of these variables. The blind participants were all experienced users of a screen reader, either JAWS (5 participants, a range of versions were used from 5.0 to 7.1) or Window-Eyes (1 participant).

Table 1: Computer and Internet experience and expertise of participants

	Blind participants	Sighted participants
Computer Experience	Range: 3 – 5 Median: 5	Range: 3 – 5 Median: 5
Computer Expertise	Range: 3 – 5 Median: 5	Range: 2 – 5 Median: 4
Hours/week using WWW	Range: 2 – 5 Median: 4	Range: 2 – 5 Median: 4.5

N.B. For Computer Experience 1 = none at all; 2 = a little; 3 = reasonable amount; 4 = quite a lot; 5 = great deal.
For Computer Expertise: 1 = not at all expert; 2 = not very expert; 3 = reasonably expert; 4 = quite expert; 5 = very expert.
For Hours/week using WWW: 1 = less than 1 hour; 2 = 1 – 5; 3 = 6 – 10; 4 = 11 – 20; 5 = more than 20 hours.

Websites evaluated

Two mobile telephone company websites were evaluated, Orange (www.orange.co.uk) and T-Mobile (www.t-mobile.co.uk). Initially, we planned to evaluate three such websites, but a pilot study showed that the individual evaluation sessions would be too long and be too tiring, particularly for the blind participants. This class of website was chosen as they have a complex range of functionality and interaction styles, but allow participants to do the same range of tasks on each website. The two particular sites were chosen as they represent two very well known mobile companies in the U.K., but their web solutions are quite different. None of the participants were familiar with the websites, they had not used either of the websites frequently or visited them recently. Figures 1 and 2 show screenshots of the home pages of the two websites at the time of the evaluation.

Tasks undertaken

Each participant was asked to attempt seven tasks with each website, the tasks being the same for each site. It was thought that repeating the same tasks on both websites was not a problem, as the two websites structured their information and functionality quite differently. Nonetheless, the order in which participants evaluated the sites was counter-balanced within each user group, to minimize order effects. The tasks were organized into a scenario of choosing a new mobile phone, exploring different options of phones and payment plans, finding the closest shop to go and try the phone, and finding information about video call coverage and the use of the phone abroad. The order of tasks was not counterbalanced, as the sequence of tasks formed a meaningful scenario and began simply and then progressed to more complex tasks.

Procedure

Before commencing the evaluation, participants were briefed about the study and procedures to be used and their written consent was obtained. They were assured that the evaluation was of the websites, and not their ability to use the Web. With their permission, all evaluation session were recorded using Morae [13] for later viewing and analysis.

Participants were given the tasks one at a time and told that they could ask to be reminded of the task at any point (some tasks required detailed information concerning phone features and payment plans). They were asked to

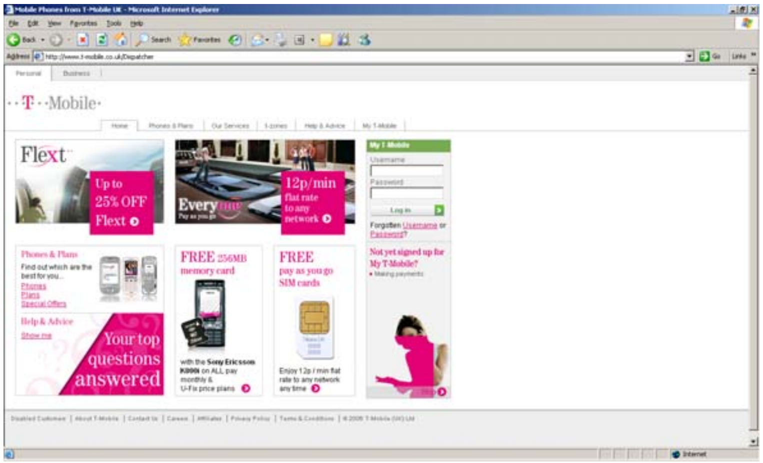

Mobile Phones from T-Mobile UK - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.t-mobile.co.uk/Dispatcher
Go Links
Personal Business
T··Mobile·
Home Phones & Plans Our Services t-zones Help & Advice My T-Mobile
Flext
Up to 25% OFF Flext
Everyone Pay as you go
12p/min flat rate to any network
My T-Mobile
Username
Password
Log in
Forgotten Username or Password?
Not yet signed up for My T-Mobile?
Making payments
Phones & Plans
Find out which are the best for you...
Phones
Plans
Special Offers
FREE 256MB memory card
with the Sony Ericsson K800i on ALL pay monthly & U-Fix price plans
FREE pay as you go SIM cards
Enjoy 12p / min flat rate to any network any time
Help & Advice
Show me
Your top questions answered
Disabled Customers | About T-Mobile | Contact Us | Careers | Affiliates | Privacy Policy | Terms & Conditions | © 2006 T-Mobile (UK) Ltd
Internet

Figure 2: T-Mobile website

"think aloud" as they did the tasks [18], particularly to articulate whenever there was a problem with the website.

Every time that occurred, the researcher conducting the evaluation (both authors conducted evaluations) asked the participant to briefly pause in doing the task, and to rate the severity of the problem on a four point scale, taken from Nielsen's heuristic evaluation method [11]: Cosmetic, Minor, Major or Catastrophic problem.

The researcher conducting the evaluation also rated the severity of the problem using the same four point scale. The procedure was that the researcher rated the problem first (but did not tell the participant the rating) and noted it on a coding sheet and then asked the participant for their rating, which was also noted. Occasionally the participant spontaneously gave their rating first, then the research tried not to be affected by the participant's rating. This did not happen often enough to affect the independence of the two sets of ratings. Due to technical difficulties, Morae recordings for one sighted and one blind participant were incomplete; where this affected results, this is indicated.

RESULTS

Overview of participant experience with the websites

Table 2, below, shows a summary of the problem results for the Orange website. The difference between "distinct problems" and "problem instances" is that any particular problem may have been encountered by more than one participant or by the same participant on more than one occasion. Table 3 shows a summary of the performance of blind and sighted participants with the Orange website. On average, sighted participants visited more distinct pages on the site than blind participants (28 vs 23.8) and also made more page visits (which includes returns to the same page (82.6 vs 64.4). This reflects the different strategies typically used by sighted and blind participants. The sighted participants tended to spend only a short time on a page and then move on if they could not find the information they were seeking, although they might well return to a page. Blind participants tended to spend longer on each page looking for the appropriate information and were less likely to return to a page. Sighted participants also encountered far fewer problems than blind participants. However, the mean severity of the problems was very similar for blind and sighted participants, both as rated by the participants themselves and the researchers.

Table 2: Problem statistics for the Orange website

	Sighted Ps	Blind Ps
Number of distinct problems	54	113
Number of problem instances	90	168
Number of pages visited by participants	51	50
Number of distinct problems per page	1.06	2.26
Number of problem instances per page	1.76	3.36

Tables 4 and 5 show the same information for the T-Mobile website.

There was a significant difference between the two groups in their success rate (F = 772.65, df = 1,10 p = 0.001), with sighted participants having a significantly higher success rate of 70.2% (on average 4.9 tasks successfully completed out of 7) compared to blind participants success rate of 50.7% (on average 3.6 tasks successfully completed out of 7). There was also a significant difference between the two websites (F = 23.28, df = 1, 10 p 0.001), with T-Mobile having a significantly higher success rate of 74.6% (on average 5.2 tasks successfully completed out of 7) compared to Orange's success rate of 46.4% (on average 3.3 tasks successfully completed out of 7). There was no significant interaction between the website and group factors. Figure 3 shows this result graphically.

Table 3: Participants' performance with the Orange website

	Sighted Ps	Blind Ps
Success rate (number of tasks successfully completed, out of total of 7)	57%	36%
Mean number of distinct pages visited	28.0	23.8
Mean number of page visits	82.6	64.4
Mean number of problem instances	15.0	28.0
Mean severity of problems as rated by participants (and standard deviation)	2.8 (0.79)	2.6 (0.68)
Mean severity of problems as rated by researchers (and standard deviation)	2.7 (0.61)	2.8 (0.66)

Table 4: Problem statistics for the T-Mobile website

	Sighted Ps (N=6)	Blind Ps (N= 5)
Number of distinct problems	62	83
Number of problem instances	102	120
Number of pages visited	50	30
Number of distinct problems per page	1.24	2.77
Number of problem instances per page	2.04	4.00

Relationship between severity ratings by participants, researchers and guidelines

To investigate the relationships between the different measures of the severity ratings of problems encountered by participants, correlations were calculated between the severity ratings provided by the relevant Guidelines and the severity ratings provided by the participants and the researchers. Before these correlations could be calculated, a number of preliminary investigations had to be undertaken, which are described first.

Table 5: Participants' performance with the T-Mobile website

	Sighted Ps (N=6)	Blind Ps (N= 5)
Success rate	83%	66%
Mean number of distinct pages visited	24.0	19.6
Mean number of page visits	58.5	38.8
Mean number of problem instances per participant	17.0	24.0
Mean severity of problems (as rated by participants)	2.4 (0.61)	2.5 (0.78)
Mean severity of problems (as rated by researchers)	2.7 (0.68)	2.6 (0.69)

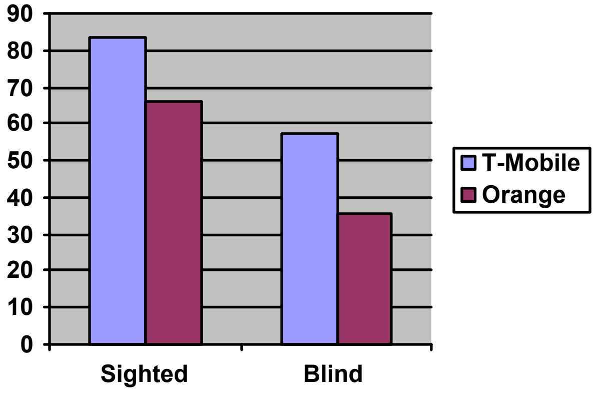


Figure 3: Success rate (%) for sighted and blind participants on the two websites

For the ratings given by the participants, for any problem in a particular condition there might be between one and six ratings, depending on the number of participants who encountered the problem. The key variable is how much agreement there is between the participants when more than one participant encountered the same problem. Figure 4 shows the levels of agreement for the 32 cases (out of a total of 312 problems) when 3 or more participants encountered the same problem. The ratings agreements were scored in the following way:

- "Total Agreement" (Total A in Figure 4) - all participants gave the same rating of severity;

- "1 difference" (1 Diff) - participants only differed by a maximum of 1 level of rating; for example 2 participants gave a rating of 2 (minor problem) and 1 gave a rating of 3 (major problem);
- "2 difference" (2 Diff) - participants differed by up to 2 levels of ratings; for example 1 participant may have given a rating of 1 (cosmetic problem), 1 participant a rating of 2 (minor problem) and 1 participant a rating of 3 (major problem);
- "3 difference" (3 Diff), is the maximum possible disagreement between participants, and means that the participants differed by up to 3 levels of ratings, for example 1 participant may have given a rating of 1 (cosmetic problem), 1 participant a rating of 2 (minor problem), 1 participant a rating of 3 (major problem) and 1 participant a rating of 4 (catastrophic problem).

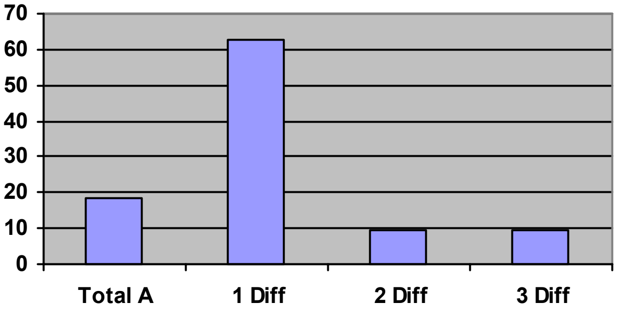


Figure 4: % Agreement between participants in severity ratings

Figure 4 shows that the majority of ratings agreements were of the "1 Difference" category (62.5%), and when the "Total agreement" and "1 Difference" agreements are combined, this accounted for over 80% of all problems encountered by 3 or more participants, a very high level of agreement was achieved. Therefore it was decided to take the mean value for the ratings of the participants who encountered a particular problem and consider that to be the measure of participant severity rating.

Finally, for both the HHS and WCAG Guidelines, for a particular problem, a number of the guidelines might have been considered relevant to any particular problem. Again, a mean was taken of the severity ratings (Relative Importance Level in the case of the HHS Guidelines, Priority Level in the case of WCAG) of the different guidelines considered relevant as the measure of severity as indicated by the Guidelines.

Table 6 shows that for the HHS Guidelines, there were no significant correlations between the severity ratings provided by the guidelines and those given by either the sighted participants or the researchers. There were significant correlations in the ratings provided by the guidelines and the ratings given by the blind participants on both websites, but in both cases these correlations were in the opposite direction to the predicted – negative correlations, which indicate that for problems with higher ratings on the HHS Guidelines, blind participants tended to give them lower ratings and vice versa. Nor were there any significant correlations between the priority levels given by WCAG and the ratings given by the participants and the researchers.

Table 6: Correlations between severity ratings of problems by participants, researchers and guidelines

	Participants and Researchers	**Participants and Guidelines**	**Researchers and Guidelines**
Orange – Sighted Ps HHS Guidelines	$r = 0.518$ $p < 0.000$	$r = 0.177$ n.s.	$r = 0.190$ n.s.
Orange – Blind Ps HHS Guidelines	$r = 0.709$ $p < 0.000$	$r = - 0.244$ $p < 0.04$	$r = - 0.179$ n.s.
Orange – Blind Ps WCAG Guidelines	$r = 0.709$ $p < 0.000$	$r = - 0.190$[1] n.s.	$r = -0.226$ n.s.
T-Mobile – Sighted Ps HHS Guidelines	$r = 0.441$ $p < 0.003$	$r = 0.012$ n.s.	$r = - 0.127$ n.s.
T-Mobile – Blind Ps HHS Guidelines	$r = 0.772$ $p < 0.000$	$r = - 0.273$ $p = 0.05$	$r = - 0.081$ n.s.
T-Mobile – Blind Ps WCAG Guidelines	$r = 0.717$ $p < 0.000$	$r = -0.095$ n.s.	$r = -0.121$ n.s.

[1] Priority levels in WCAG use lower numbers to indicate greater priority and higher numbers to indicate lower priority, whereas the ratings of severity given by participants and researchers use the reverse system. So for the correlations between participant/researcher ratings and WCAG priority levels, a significant negative correlation indicates agreement between the two sets of ratings.

Relationship between problems encountered by different participant groups

To investigate the nature of the relationship between problems encountered by blind and sighted participants, all problems were tabulated for whether they were encountered by sighted participants only, by blind participants only or by both blind and sighted participants. For these analyses, only pages on the websites that had been visited by at least three blind and three sighted people were included, to create equity in the potential for detecting problems.

For the Orange website, there were 106 distinct problems encountered by both groups. Table 7 shows the breakdown of these problems between the three categories. 17% of problems were encountered by sighted participants only. 66% of problems were encountered by blind participants only. 17% of problems were encountered by both blind and sighted participants.

Table 7: Breakdown of problems encountered by blind and sighted participants (Orange website)

Encountered by:	**Number**	**%**
Blind participants only	70	66.0
Sighted participants only	18	17.0
Both blind and sighted participants	18	17.0
Total	106	

To investigate whether the problems encountered by sighted and blind participants were more severe for the blind participants than for the sighted participants on the Orange website, the ratings of the severity of the problems encountered by both participant groups as given by both the participants themselves, the researchers and the guidelines were analysed. The mean rating of severity given by the blind participants was 2.60 (s = 0.543) and the mean rating by the sighted participants was 2.61 (s = 0.57). This difference was not significant (t = 0.07, df = 15, n.s.). Table 8 shows the mean ratings for the researchers, again this difference was not significant.

Table 8: Mean ratings (and standard deviations) of severity of problems encountered by both blind and sighted participants (Orange website)

	Mean rating problems encountered by blind Ps	**Mean rating problems encountered by sighted Ps**	**t-test**

Participants	2.60[1]	2.61	t =0.07
			df = 15
			n.s.

Researchers	2.97
	(0.56)
		2.78
		(0.60)
			t= 1.10
			df= 19
			n.s.

Guidelines	-0.119
	(0.79)
		-0.07
		(0.94)
			t =0.24
			df =33
			n.s.

[1] As the HHS guidelines rate importance on a five point scale and the WCAG guidelines rate priority on a three point scale, z-scores were taken of all ratings to allow comparison.

For the T-Mobile website, there were 113 distinct problems encountered by both groups. Table 9 shows the breakdown of these problems between the three categories. Just over 30% of problems were encountered by sighted participants only, so were clearly usability problems. Just over half (65, 57.5%) of problems were encountered by blind participants only. Just over 10% of problems (12, 10.6%) were encountered by both blind and sighted participants.

Table 9: Breakdown of problems encountered by blind and sighted participants (T-Mobile website)

Encountered by:	**Number**	**%**
Blind participants only	65	57.5
Sighted participants only	36	31.9
Both blind and sighted participants	12	10.6
Total	113	

To investigate whether the problems encountered by sighted and blind participants were more severe for the blind participants than for sighted participants on the T-Mobile website, the ratings of the severity of the problems encountered by both participant groups as given by both the participants themselves, the researchers and the guidelines were analysed. The mean rating of severity given by the blind participants was 3.08 (s = 0.524) and the mean rating by the sighted participants was 2.25 (s = 0.665). This difference was significant (t = -2.99, df = 5, p < 0.03), with blind participants giving more severe ratings of their problems when compared to their sighted peers. Table 10 shows the mean ratings for the researchers, and both researchers individually, none of these differences were significant.

Table 10: Mean ratings (and standard deviations) of severity of problems encountered by both blind and sighted participants (T-Mobile website)

	Mean rating problems encountered by blind participants	**Mean rating problems encountered by sighted participants**	**t-test**

Participants	3.08
	(0.52)
		2.25
		(0.67)
			t =2.99
			df = 5
			p<0.03

Mean of both Rs	2.95
	(0.42)
		2.68
		(0.51)
			t =1.49
			df = 11
			n.s.

Guidelines (HHS/WCAG)	0.034
	(1.06)
		0.04
		(0.98)
			t =0.16
			df = 37
			n.s.

DISCUSSION AND CONCLUSION

This study investigated three research questions concerning the accessibility and usability of websites, using two mobile phone company websites as the target domain, and blind and sighted people as the target user groups.

Nature of the relationship between problems encountered by blind and sighted participants

For both websites, the problems encountered by the blind and sighted participants constituted intersecting sets, with some problems only encountered by blind participants (over half the total number of problems encountered), some problems encountered by only sighted people and some problems, although not a large percentage, encountered by both groups. These results were not simply a consequence of sampling, that is the fact that a particular problem happened to be encountered by a blind person or persons rather than a sighted person or persons. It was clear that some of the problems resulted specifically from the fact that the blind people used screenreaders which results in problems not encountered when using the sites visually. For example, although both websites provided clear headings within pages to divide text into identifiable chunks, neither site indicated these headings in the html code, so they could not be detected by the blind participants. Conversely, some of the problems encountered by sighted users related to the visual presentation of the information and thus did not affect blind participants. For example, on the T-Mobile website, information about price plans was set out in a table, but the headings for the 12 month and 18 month plans were poorly grouped visually, so some sighted participants were unsure whether the information was a heading for one section of the table, or a footnote to the previous section. This problem did not affect blind participants, as the information was before the price group, so they correctly assumed it related to the prices that followed.

Thus accessibility problems were not a complete sub-set of usability problems, as suggested by Thatcher *et al* [17] nor were usability problems a complete sub-set of accessibility problems, as might be inferred from Shneiderman [13, 14]. However a more detailed analysis of the nature of the problems encountered by each group and the problems shared by both groups, will be undertaken in a future paper.

The degree of overlap between the problems encountered by the two groups was not large, certainly not as great as we had predicted. However, it should be remembered that the disabled group studied here were blind people using screenreaders, whose interaction with the Web is conceptually most different from that of non-disabled, sighted people. So the fact that there is about 14% overlap in problems even in this most extreme case shows that there is communality in accessibility and usability that is perhaps being neglected by researchers and the web industry. If this study were repeated with other disabled user groups, one might expect more overlap with problems encountered by non-disabled users.

Severity of problems encountered by both blind and sighted people

For those problems encountered by both blind and sighted participants, it was hypothesized that they would be more severe for blind people than for sighted people and this would be reflected in the ratings given by participants, researchers and the guidelines. Only limited support was found for this hypothesis, with a significant difference between participants' ratings on only one of the two websites, and no differences in the ratings given by researchers and the guidelines. However, it is the participants' ratings that are of most interest here. Unfortunately, in spite of a large number of user problems studied, the number of problems encountered by both blind and sighted participants was quite small, so the test of this research question is not highly robust. Further evidence to support this hypothesis is needed. But if it can be found, it would be particularly useful in accessibility and usability evaluations, as it would show that studying the accessibility problems on a website highlight and amplify the usability problems. This is both useful practically and of theoretical interest in understanding the relationship between accessibility and usability.

Relationship between the different measures of the importance or severity of problems

Perhaps the most interesting findings from this study concern the relationships between the different measures of importance or severity of the problems encountered by participants. Firstly, the fact that there is strong agreement between participants in the severity of a particular problem is encouraging. Although the participants for this study were chosen to be relatively homogenous in their computing and internet expertise, they had different strategies for interacting with the sites and different knowledge that they brought to the tasks. Nonetheless they broadly agreed on which problems were important and which problems were not.

However, the most important finding on this research question is the lack of any significant relationship between the ratings given by the participants (and the researchers) of the importance of the problems they encountered and the ratings provided by the HHS and WCAG guidelines of the severity of problems. This study confirms the preliminary findings of Harrison and Petrie [6] who failed to find a relationship between participants' ratings, one expert's ratings and the guidelines ratings, but the current study is a considerable larger study, both in terms of numbers of participants and numbers of user problems. In the case of the HHS guidelines, this is particularly interesting, as these guidelines and their importance ratings were derived from a careful study of a considerable literature on usability studies, so one might have hoped that the ratings would be shown to be valid in the context of a particular evaluation. In the case of the WCAG guidelines, the problem is that there has been little detailed research into the ways screenreader users interact with the Web and the problems they encounter, so the empirical basis for the Priority Levels in WCAG was actually quite weak. Nonetheless, it should not be thought that this study is suggesting that providing importance ratings on guidelines is not possible, rather that considerably more research on the relationship between users' experiences of problems and the ratings to be given to those problems is required.

ACKNOWLEDGMENTS

We would like to thank all the participants who took part in the study for their patience and enthusiasm. We would also like to thank Fraser Hamilton, Chandra Harrison and Christopher Power for both practical assistance and useful discussions in the preparation of this paper. We would like to thank the Darzentas and Scott-Surridge clans for ensuring the paper actually got written and submitted and the anonymous reviewers for very helpful comments.

REFERENCES

1. DiBlas, N., Paolini, P. and Speroni, M. (2004). Usable accessibility to the Web for blind users. In C. Stary and C. Stephanidis (Eds.), *User-centered Interaction Paradigms for Universal Access in the Information Society. Lecture Notes in Computer Science, No 3196.* Berlin: Springer.
2. Brajnik, G. (2006). Web accessibility testing: when the method is the culprit*. Proceedings of the 10th International Conference on Computers Helping People with Special Needs. Lecture Notes in Computer Science,* No 4061. Berlin: Springer.
3. Chisholm, W., Vanderheiden, G., and Jacobs, I. (1999). Web content accessibility guidelines 1.0. www.w3.org/TR/WCAG10
4. Cockton, G. and Woolrych, A., (2001). Understanding Inspection Methods: Lessons from an Assessment of Heuristic Evaluation. In A. Blandford, J. Vanderdonckt and P.D. Gray (Eds.) *People and Computers XV*, Springer-Verlag, 171-192, 2001
5. Disability Rights Commission. (2004). The Web: access and inclusion for disabled people. London: The Stationery Office.
6. Harrison, C. and Petrie, H. (2006). Impact of usability and accessibility problems in e-commerce and e-government websites. In *Proceedings of HCI 2006, Volume 1*. London: British Computer Society.
7. Henry, S.L. (2006). Introduction to Web accessibility. www.w3.org/WAI/intro/accessibility.php
8. International Standards Organization. (1992 - 2000). *Standard 9241: Ergonomic requirements for office work with visual display terminals*. www.iso.org .
9. Iwarsson, S. and Stahl, A. (2003). Accessibility, usability and universal design – positioning and definition of concepts describing person-environment relationships. *Disability and Rehabilitation*, 25(2), 57 – 66.
10. Koyani, S., Bailey, R., Nall, J., Allison, S., Mulligan, C., Bailey, K. and Tolson, M. (2004). *Research-based Web design and usability guidelines.* www.usability.gov/guidelines/guidelines_book.pdf
11. Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen and R.L. Mack (Eds.), *Usability inspection methods*. New York: John Wiley and Sons.
12. Petrie, H., King, N. and Hamilton, F. (2005). *Accessibility of museum, library and archive websites: the MLA audit*. http://www.mla.gov.uk/webdav/harmonise?Page/@id=73&Document/@id=23090&Section%5B@stateId_eq_left_hand_root%5D/@id=4302
13. Shneiderman, B. (2000). Universal usability. *Communications of the ACM*, 43(5), 85 – 91.
14. Shneiderman, B. (2003). Promoting universal usability with multi-layer interface design. *Proceedings of the 2003 Conference on Universal Usability (CUU 2003).* New York: ACM Press.
15. Sloan, D. (2006). Two cultures? The disconnect between the web standards movement and research-based web design guidelines for older people. *Gerontechnology*, 5(2), 106 – 112.
16. Techsmith. http://www.techsmith.com/morae.asp
17. Thatcher, J., Waddell, C.D., Henry, S.L., Swierenga, S., Urban, M.D., Burks, M., Regan, B. and Bohman, P. (2003). Constructing accessible web sites. San Francisco: glasshaus.
18. van Someren, M. W., Barnard, Y. F. and Sandberg, J. A. C., (1994). The think aloud method: A practical guide to modeling cognitive processes. San Diego, CA: Academic Press