

Investigating the Appropriateness and Relevance of Mobile Web Accessibility Guidelines

Raphael Clegg-Vinell
AbilityNet
Harella House
90-98 Goswell Road
London, EC1V 7DF, UK
+44 7531 5104320
raphael.clegg-
vinell@abilitynet.org.uk

Christopher Bailey
AbilityNet
Harella House
90-98 Goswell Road
London, EC1V 7DF, UK
+44 (0)207 490 8429
chris.bailey@abilitynet.org.uk

Voula Gkatzidou
Brunel University
St. Johns 201
Information Systems and Computing
Uxbridge, UB8 3PH, UK
+44 (0)1895 267237
voula.gkatzidou@brunel.ac.uk

ABSTRACT

The Web Accessibility Initiative (WAI) of the World Wide Web Consortium (W3C) develop and maintain guidelines for making the web more accessible to people with disabilities. WCAG 2.0 and the MWBP 1.0 are internationally regarded as the industry standard guidelines for web accessibility. Mobile testing sessions conducted by AbilityNet document issues raised by users in a report format, relating issues to guidelines wherever possible. This paper presents the results of a preliminary investigation that examines how effectively and easily these issues can be related by experts to the guidelines provided by WCAG 2.0 and MWBP 1.0.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces – *Evaluation/methodology*. K.3.2 [Computers and Education]: Computer and Information Science Education – *Curriculum*. K.4.2 [Computers and Society]: Social Issues – *Assistive Technologies for persons with disabilities*.

General Terms

Measurement, Human Factors, Verification.

Keywords

Mobile accessibility, mobile accessibility, guidelines.

1. INTRODUCTION

Designing and developing for mobile presents new challenges to web accessibility. The screen dimensions on mobile interfaces are much smaller than traditional desktop platforms and there has been an increasing shift towards gesture based interactions. The portability of these devices and associated connectivity challenges they pose mean designers and developers face fresh challenges to ensure content is delivered accessibly. Designing for effective use in the mobile context when websites and applications are used in a range of environmental conditions can be problematic.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

W4A '14, April 07 - 09 2014, Seoul, Republic of Korea
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-2651-3/14/04...\$15.00.
<http://dx.doi.org/10.1145/2596695.2596717>

The Web Content Accessibility Guidelines (WCAG 2.0) [1] and the Mobile Web Best Practices (MWBP 1.0) [2] are guidelines which provide documentation and support for ensuring content delivered on the web is accessible, particularly for those with disabilities. Despite potential limitations, guidelines are generally considered an important tool in raising awareness of accessibility, establishing its importance and guiding developers in developing sites which support accessibility [3]. Their primary importance in relation to this work is that guidelines provide a point of reference with which to design, develop and test a resource against. Unless there is a clear definition of what is to be tested, the testing process is rarely successful [4].

Previous studies have found that while there is a strong agreement between participants in the severity rating of problems encountered, there is no relationship between severity levels given in WCAG and the ratings given by the users. For this reason developers should obtain severity ratings from users or an expert rather than relying on those provided by guidelines [5]. Studies have also examined the relevance of MWBP 1.0 for users with visual impairments and have proposed a framework for filtering and selecting the relevant guidelines for different disability types [6].

This work aims to investigate the validity of MWBP 1.0 for a range of user groups by examining whether issues reported by users in live user testing sessions can be mapped to relevant WCAG 2.0 Success Criteria of MWBP 1.0 Statements. If the findings show problems clearly correspond to checkpoints, it proves the validity of the guidelines.

2. BACKGROUND

This paper reports on work that has been carried out at AbilityNet. As part of our consultancy services, regular mobile user testing of live applications is conducted to identify the strengths and weaknesses of a client's product or service and to highlight potential accessibility and usability issues. This follows the approach suggested by BS8878 [7].

In order to help clients gain an understanding of problems raised by testers and to support our reporting, accessibility issues are mapped wherever possible to checkpoints in WCAG and MWBP. In our experience, many clients will prioritize fixing issues unearthed from testing sessions which are reported alongside corresponding guidelines. Although this can be straightforward, it can also be time-consuming and complex when issues do not fall under a guideline, or require significant analysis and

interpretation from an experienced consultant in order to link them. There are clear benefits in reporting issues under WCAG and MWBP as it can help clients gain a better understanding on why a particular issue causes a barrier to accessibility.

Some requirements such as WCAG 2.0 Success Criterion 1.4.3 Contrast are clearly defined and unambiguous which makes reporting issues relating to this guideline uncomplicated. There are examples of other unequivocal guidelines under both WCAG and MWBP which not only leave little room for interpretation but also outline the same principles for both desktop and mobile contexts. WCAG 2.0 Success Criterion Page Titled “*Web pages have titles that describe topic or purpose*” is essentially the same as MWBP 1.0 5.4.1 [PAGE_TITLE] “*Provide a short but descriptive page title*”. Reporting is more complex when an issue does not clearly relate to WCAG or MWBP guidelines. In our experience, there are problems raised during testing sessions which do not fall under these. In other cases, there are issues which require expert interpretation by an experienced consultant for them to be associated to guidelines. Relating some issues reported by users to guidelines requires an in-depth knowledge of not only the guidelines but also all supporting documentation. In some cases, the interpretation can be quite abstract and is subject to a significant evaluator effect [8] as it can be based on the prior knowledge and experience of one consultant.

The overall aim of this work is to investigate the effectiveness of existing guidelines in supporting User Experience (UX) practitioners, designers and developers in ensuring accessibility. This paper presents a preliminary exploration of the extent to which issues we have documented from mobile testing reports can be clearly associated to checkpoints provided by WCAG 2.0 and MWBP 1.0. If the results of our testing sessions demonstrate that significant issues reported by users cannot be related to existing guidelines, there is an argument for refinement of the guidelines to ensure they reflect issues based on empirical evidence. Using the results of testing sessions conducted over a longitudinal period, we can propose additional MWBP statements that would address issues reported by real users, thus ensuring they have potential for greater validity and relevance. The ultimate aim of this work will be to develop a series of evidence-based practitioner guidelines (or heuristics) which can inform the development of accessible mobile applications and contribute to an effective user experience.

3. CASE STUDY

Five lab-based mobile user testing studies were conducted by four accessibility consultants from March to December 2013. They evaluated a variety of mobile websites and native applications across a range of platform. The testers had the following disabilities or impairments:

- Dyslexic – no AT (assistive technology) requirements
- Deaf – hearing aid and native British Sign Language (BSL) communicator.
- Motor impairment – no AT requirements
- Medium vision impairment – multi-gesture Zoom
- Severe vision impairment or blind - screen reader
- Silver surfer (aged 50+) – no AT

The length of the testing sessions was 60–90 minutes and these were conducted with 6–8 users across 1–2 days. The sessions were predominantly task-based and scenarios were provided. The tasks were actions or processes a user would typically perform on

the website or application. The number of tasks varied due to a number of factors, including time allocation, the complexity of the application, the user’s experience and their disability or impairment. Throughout testing, moderator intervention was kept to a minimum. Users were encouraged to think aloud, a method considered intrinsic to usability engineering practice and moderators only intervened if substantial time had elapsed without a user progressing. The moderator would also closely observe the users actions and regularly ask questions of the user to ensure a dialogue was constantly maintained.

The findings from these sessions were written up in reports which outlined the accessibility and usability issues. The severity of the issues were determined by a combination of three factors:

- The **frequency** with which the problem occurs: Is it common or rare?
- The **impact** of the problem if it occurs: Will it be easy or difficult for the users to overcome?
- The **persistence** of the problem: Is it a one-time problem that users can overcome once they know about it or will users repeatedly be bothered by the problem? [9]

The severity of an issue cannot be accurately rated without contextualising the circumstances under which the problem occurs. Contextual considerations include what the potential barrier is, the type of user affected and what their browsing capabilities are. No reliable conclusion can be derived regarding the impact that the problem will have on a task without this information [10]. For this reason, we rated each issue raised by a tester case-by-case, outlining the nature of the problem, how it occurred, the specific impact it had on the user and the barriers it may cause others.

The High and Medium Priority accessibility and usability issues were collated from AbilityNet’s mobile testing sessions, taken from reports spanning March–December 2013. For the purposes of this work we do not present the Low Priority issues. They are included in reports but usually ignored by clients, despite recommendations that all issues are addressed. This is can be for a number of reasons including financial restraints and time limits. The high and medium priority issues from the testing sessions were banded into three categories:

1. Issues which clearly correspond to guidelines under WCAG and or MWBP 1.0.
2. Issues which cannot be easily associated to guidelines under WCAG 2.0 or MWBP 1.0 without expert interpretation from an experienced consultant or developer.
3. Issues which do not fall under WCAG 2.0 or MWBP 1.0 but still cause a significant accessibility barrier.

The next step involved mapping (wherever possible) the accessibility issue reported to the closest guidelines under WCAG 2.0 and the W3 Mobile Web Best Practices (MWBP) 1.0. These results were then analysed to see if any trends emerged and how the issues reported relate to WCAG and MWBP.

4. PRELIMINARY RESULTS

The data was put into a table to collate the results. This shows the severity rating, the issue which occurred and any relatable guidelines from WCAG 2.0 and MWBP 1.0. Frequency refers to the number of studies in which a problem occurred. Instances of issues relating to the same guideline were reported under different severity ratings. Also, some of the issues reported just once by

users were excluded from the table. These were those considered to be quality assurance problems rather than accessibility concerns. For example, one unique issue raised concerned a financial app which displayed incorrect transaction dates.

Table 1. Mapping tester issues to guidelines

Severity Rating	Issue	Related Guideline(s)	Frequency
High Priority	Colour contrast	[ref: 1.4.3] [COLOR_CONTRAST]	3
High Priority	Bluetooth keyboard compatibility	[ref: 2.1.1]	2
High Priority	Elements not announced on focus	[ref: 3.2.1]	2
High Priority	Selectable elements undetected by VO	[ref: 2.1.1]	2
High Priority	Interactive elements poorly labeled	[ref: 2.4.6] [LINK_TARGET_ID]	2
High Priority	Unclear label instructions	[ref: 3.3.2]	2
High Priority	Icon purpose unclear	None	1
Medium Priority	Buttons poorly labeled	[ref: 2.4.6] [LINK_TARGET_ID]	6
Medium Priority	Finger pinch zoom	[CAPABILITIES]	4
Medium Priority	Text size or icons too small to read/identify	[ref: 1.4.4]	3
Medium Priority	Icon purpose unclear	None	2
Medium Priority	Confirmation messages gone before users could read them	None	1
Medium Priority	Target size of buttons too small	None	1

5. DISCUSSION

Analysing the data presented trends of certain issues which clearly correspond to guidelines under WCAG and or MWBP. Users frequently reported issues relating to contrast, most commonly due to text or icons which had an insufficient colour contrast with their backgrounds. These unambiguously correspond to WCAG 2.0 AA Success Criteria 1.4.3 Contrast (Minimum) and also MWBP 1.0 5.3.6 Color, so reporting issues of this nature was straightforward. This issue could be avoided relatively easily had this guideline been followed during design and development. However, there is discord between our own severity rating and that of WCAG. All contrast issues were reported by us as High Priority but contrast is AA under WCAG. Studies have found that

there is generally no relationship between the severity ratings given by testers and those given by guidelines. In the case of the colour contrast issues reported, the disharmony between the two rating scales has the potential to cause confusion to developers.

This paper highlights several weaknesses in the guidelines. A number of accessibility and usability issues were reported by testers which do not correspond to any of the guidelines under WCAG or MWBP and some of these issues caused significant barriers to users. One which was reported during three separate studies regarded icons which had an unclear purpose to users. These were reported under High and Medium Priority and therefore posed a considerable impact on the accessibility and usability of the apps. Icons are taking up unnecessary space in a relatively small mobile interface if their purpose is unclear. This particular issue is closely related to a usability heuristic – “Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility” [11].

In four different testing studies, multiple users tried using finger pinch zoom gestures to enlarge text or icons they struggled to read or identify. Their gestures either had no impact at all or caused instabilities with the app’s behaviour. These testers were not limited to a particular disability group and included dyslexic, visually impaired and silver surfers. Pinch based zoom gestures are common place with gestural based interfaces so their actions were instinctive. Users who could not enlarge elements on the screen often had great difficulties completing tasks and tended to hold their devices unnaturally close to their faces in order to read text or identify icons. Thus, the results from the testing sessions we conducted demonstrate a need for documentation which provides guidance on mobile or touch based interface zoom capabilities. MWBP 1.0 [CAPABILITIES] “Exploit device capabilities to provide an enhanced user experience” loosely corresponds but this requires expert interpretation because the checkpoint is abstract. WCAG 2.0 AA Success Criteria 1.4.4 Resize Text states “text can be resized without assistive technology up to 200 percent without loss of content or functionality”. This guideline does not provide information for ensuring content is scalable on mobile platforms and the techniques provided by WCAG to address issues of this nature are for desktop-based user agents. Significantly, the guideline says text should be resizable without assistive technology but with mobile devices, our testers always tried enlarging content using inbuilt zoom accessibility features. Furthermore, the guideline only refers to text and in tests, it was the size of icons users found too small.

Issues were raised during three testing studies regarding text size or icons too small to read or identify. In these cases, the problems were mapped to Success Criteria 1.4.4 Resize Text. This diagnosis is based on the expert knowledge that if the elements can be enlarged by users, the impact small icons or text size has on accessibility is diminished. However, our testing has proved that reporting issues of this nature under this checkpoint has flaws. In one testing study, a particular interactive icon was so small, and without sufficient contrast, that two users missed it completely when carrying out a task.

Neither Success Criteria 1.4.4 nor MWBP provide any documentation for minimum size requirements for text or icons. This is a complex issue due to the vast array of screen sizes on

web devices, heavily affecting the content space available. However, the Android and iOS frameworks do have suggested minimum sizes for text and target elements. App designers and developers should be aware of these but through our testing sessions, we have examples where they have been ignored or overlooked. Having similar guidelines under WCAG or MWBP has the potential to reinforce the importance of making content large enough for users to read or comprehend.

A high priority issue raised in two of our testing studies was selectable elements which were undetected by VoiceOver. These were items on screen, such as checkboxes, which required a user's interaction for task completion. Without an announcement by VoiceOver, these users were unaware of their existence and therefore could not complete certain tasks, such as a registration process. These problems were related to WCAG 2.0 A Success Criteria 2.1.1 Keyboard, which states "*All functionality of the content is operable through a keyboard interface without requiring specific timings for individual keystrokes*". The principle is fundamentally the same for mobile interfaces in that testing should be conducted to ensure content is operable using swipe or explore by touch gestures. These are the actions most frequently carried out by our VoiceOver and TalkBack users in order to navigate. However, Success Criteria 2.1.1 only contains guidelines for keyboard accessibility and nothing on checking operability with touch based gestures. There is also no documentation which covers this under MWBP. Clients might therefore be unclear as to why these issues have been mapped to this guideline and, more importantly, how to address them.

As well as contextual considerations discussed in Section 3.0, this work aims to raise awareness of the need for more investigation into issues which affect users in the mobile context, including the need for flexible severity scales.

6. CONCLUSION

Mapping results from our mobile results to guidelines has highlighted examples of issues which clearly relate to checkpoints in WCAG 2.0 and MWBP 1.0. However, the findings show that more frequently, issues do not easily relate to guidelines or require expert interpretation from an experienced consultant to do so. This is time-consuming for reporting problems and there is the potential for clients to be confused if they cannot fully comprehend how issues we document correspond to guidelines.

Guideline standards need to fully embrace a user-centred approach to ensure documentation addresses those issues which have the highest impact on accessibility. They should also provide greater clarification on how best to design and develop accessibly for mobile platforms. Although there is documentation in place, our testing highlights a need for changes, and in some cases additional checkpoints in order for all issues raised to clearly relate to existing guidelines. Our study therefore reinforces the need for a more user-centred approach, by detecting issues that are not covered by guidelines but are detected in a live context.

There is an increasing awareness of the need to support accessibility in the mobile context. As such, mobile guidelines need to combine elements of accessibility, usability and user experience in order to optimise their effectiveness is addressing issues which impact mobile web users. Combining these aspects

into the same documentation would be challenging but it could potentially reduce the time spent by consultants mapping issues to guidelines and the time developers spend addressing them. The work remains ongoing and we will continue to monitor how effectively issues detected by users in a live test environment can be mapped to WCAG 2.0 and MWBP 1.0.

Future work will focus on the development of our own evidence-based mobile guidelines or heuristics which can better form the development of accessible mobile applications.

7. REFERENCES

- [1] W3C. (2008a). Web Content Accessibility Guidelines (WCAG) 2.0. W3C Recommendation. Available Online: <http://www.w3.org/TR/WCAG20/>.
- [2] W3C. (2008a). Mobile Web Best Practices 1.0. W3C Recommendation. Available Online: <http://www.w3.org/TR/mobile-bp/>.
- [3] Kelly, B., Sloan, D., Brown, S., Seale, J., Petrie, H., Lauke, P., & Ball, S. (2007). Accessibility 2.0: People, policies and processes. *Proceedings of the 2007 International Cross-Disciplinary Conference on Web Accessibility (W4A)*, Banff, Canada. 138-147. Doi: 10.1145/1243441.1243471.
- [4] 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 (ICCHP)*, Linz, Austria. 156-163. Doi: 10.1007/11788713_24.
- [5] Harrison, C., & Petrie, H. (2006). Impact of usability and accessibility problems in e-commerce and e-government websites. *Proceedings of HCI 2006*, London, UK.
- Petrie, H., & Kheir, O. (2007). The relationship between accessibility and usability of websites. *Proceedings of the 2007 SIGCHI Conference on Human Factors in Computing Systems*, San Jose, California, USA. 397-406. Doi: 10.1145/1240624.1240688.
- [6] Carriço, L., Lopes, R., & Bandeira, R. 2011. Crosschecking the mobile web for people with visual impairments. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility (W4A '11)*. ACM, New York, NY, USA, Article 12, 4 pages. Doi: 10.1145/1969289.1969305.
- [7] Access 8878. Assuring Accessibility. Available Online: <https://www.access8878.co.uk/assuring-accessibility.aspx/>. [Accessed: 5th January 2014].
- [8] Hornbaek, K., & Frokjaer, E. (2008). A study of the evaluator effect in usability testing. *Human Computer Interaction*, 23(3), 251-277. Doi: 10.1080/07370020802278205.
- [9] Nielsen, J. (1995). Severity Ratings for Usability Problems. Available Online: <http://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/>.
- [10] Brajnik, G. (2008). A comparative test of web accessibility evaluation methods. *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, Halifax, Nova Scotia, Canada. 113-120. Doi: 10.1145/1414471.1414494.
- [11] Nielsen, J. (1995). 10 Usability Heuristics for User Interface Design. Available Online: <http://www.nngroup.com/articles/ten-usability-heuristics/>