CrossMark

LONG PAPER

Evaluation of e-commerce websites accessibility and usability: an e-commerce platform analysis with the inclusion of blind users

Ramiro Gonçalves[1] · Tânia Rocha[1] · José Martins[1] · Frederico Branco[1] ·
Manuel Au-Yong-Oliveira[2]

Abstract Considering the importance associated with e-commerce website accessibility and usability, a study on one of the most relevant Portuguese e-commerce websites has been performed using both automatic and manual assessment procedures. In an initial stage, we evaluated the chosen website with a Web accessibility and usability automatic tool called *SortSite*; after that, we performed a manual evaluation to verify each previously detected error and present possible solutions to overcome those faults. In a third phase, three usability specialists have been used to perform a heuristic evaluation of the chosen website. Finally, user tests with blind people were carried out in order to fully assess the compliance with accessibility and usability guidelines and standards. The results showed that the platform had a good score regarding the automatic evaluation; however, when the heuristic and manual evaluations were performed, some accessibility and usability problems were discovered. Moreover, the user test results showed bad marks regarding efficiency, effectiveness, and satisfaction by the group of participants. As a conclusion, we highlight user interaction problems and propose seven recommendations focused on enhancing accessibility and usability of not only the evaluated e-commerce website, but also of other similar ones.

Keywords Web accessibility · Web usability ·
E-commerce · Heuristics · User tests · Blind people

✉ Ramiro Gonçalves
ramiro@utad.pt

Tânia Rocha
trocha@utad.pt

José Martins
jmartins@utad.pt

Frederico Branco
fbranco@utad.pt

Manuel Au-Yong-Oliveira
mao@ua.pt

[1] INESC TEC and University of Trás-os-Montes and Alto Douro, Quinta de Prados, 5000-801 Vila Real, Portugal

[2] GOVCOPP, Department of Economics, Management, Industrial Engineering and Tourism, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

1 Introduction

Currently, there is a real concern to develop accessible and usable tools and environments as a result of the increasing importance of digital environments and the universality of the Internet. However, these environments are not fully accessed, because the tools and environments are created almost without concerns for accessibility or usability [1–3].

We believe this type of research is significant as it identifies and reveals accessibility and usability problems in digital environments, allowing to raise awareness to these issues. Specifically, our study aims at identifying accessibility and usability flaws in a major Portuguese e-commerce platform. Due to the fact that we recognize its relevance, we find of the most importance that this platform should be accessible and useful to all consumers, with or without disabilities. Ultimately, we intend to increase consciousness regarding accessibility and usability issues, working to improve access to a more inclusive Digital World.

Therefore, the two concepts, accessibility and usability, are of major importance in today's information society in order to improve people's quality of life [4].

🖄 Springer

Accessibility can be defined as "*the ability of a product, service, environment or equipment to be used by a large range of people with very different capabilities*" [5]. Specifically, Web accessibility is defined as easy access to Information and Communication Technology (ICT) products and services, involving three factors: users (independently of the abilities/capacities of the users, the products and services must be accessible); situation (regardless of the software and hardware being used, the systems must be accessible); and environments (irrespective of the place where the user is situated, the systems must be accessible) [3, 6].

There are several entities that provide guidelines and rules for Web accessibility, such as the World Wide Web Consortium (W3C), the International Organization for Standardization (ISO) and the US Federal Government through Section 508 [7]. Despite this diversity of concerned organizations, the most relevant to the area is the World Wide Web Consortium (W3C). At the end of the 1990s, W3C presented the Web Content Accessibility Guidelines (WCAG), currently recognized as the international standard for Web accessibility. This standard is already in its second version (WCAG 2.0) and provides guidelines divided into four principles: "Perceivable (information and user interface components must be presentable to users in ways they can perceive); Operable (user interface components and navigation must be operable); Understandable (information and the operation of the user interface must be understandable); and Robust (content must be robust enough that it can be interpreted reliably by a wide variety of user agents, including assistive technologies)" [8]. WCAG 2.0 are assessed through compliance levels: "Level A (the minimum level of conformance, the Web page satisfies all the Level A Success Criteria, or a conforming alternate version is provided); Level AA (the Web page satisfies all the Level A and Level AA Success Criteria, or a Level AA conforming alternate version is provided); and Level AAA (the Web page satisfies all the Level A, Level AA and Level AAA Success Criteria, or a Level AAA conforming alternate version is provided)" [8].

Drawing on existing literature, one can perceive that the methods and techniques most often used for assessing accessibility compliance are automatic tools or manual evaluation performed by experts using guidelines [3]. Automatic tools are effective in the identification of accessibility errors; however, they do not have the same ability to assess the accessibility of a website that a human user has. This can lead to some accessibility errors that tend to exist over time, even when the tool indicates that all is correct. For this reason, automatic tools should not be a substitute for manual evaluation but a complement of it.

Regarding usability, this concept can be defined as "*the level of effectiveness, efficiency and satisfaction a certain group of users, of a product or environment, accomplishes for a specific purpose in a certain context of use*" (ISO 1997). Usability evaluation depends on the registering and analysis of these three variables: "*effectiveness is the accuracy and completeness with which specified users can achieve specified goals in particular environments; efficiency, the resources expended in relation to the accuracy and completeness of goals achieved; and satisfaction, the comfort and acceptability of the work system to its users and other people affected by its use*" [5]. The most used methods of usability assessment are heuristic evaluation, where reviews and/or tests are carried out by experts in usability and user tests, where one uses usability evaluation techniques involving a group of potential users of the system as a representative group sample [9–12]. User tests tend to be more accurate and have more objective results than the heuristic evaluation [13].

With the above in mind, and following the indications of Albert, Tullis [14], and Richey, Klein [15] concerning the proper focus and format to define research questions, the ones inherent to the present project are the following:

- Can we achieve accessibility by following the guidelines provided by W3C?
- Is it possible to develop accessible e-commerce websites for blind users by following the W3C guidelines?
- Is the chosen case study e-commerce website accessible and useful for blind users?

The above-mentioned research questions will allow, from our perspective, to highlight the value and achievements of our project and consequently to establish a basis for further research.

In this assessment context, a case study divided into three evaluation stages is presented. Firstly, an accessibility evaluation was performed using an automated tool, *SortSite* [16], following which the WCAG 2.0 were checked manually by an expert. Subsequently, a heuristic evaluation was executed by three usability experts. Finally, a user test with blind users was carried out. For the above, the main purposes of the assessment were defined: to verify compliance with the WCAG 2.0 standard (levels A, AA, and AAA); to evaluate the system functionality by verifying compliance with user requirements and performance; and to identify problems related to the functionality or usability of the system using blind users.

The paper is divided as follows: first, related studies are presented; second, the materials and methods used are described; the following sections are divided concerning the evaluation phase (automatic and manual evaluation, heuristic evaluation, and the user tests); then, a section of

discussion is presented; and finally, conclusions and future work are presented in the last section.

2 Related studies

Blind users often experience frustration during their Web interaction. Lazar et al. [17] reported five such frustrations: confusion caused between the page layout and screen reader feedback; conflict between screen reader and application; forms that had been poorly designed and/or unlabeled; no addition of the alternative text in pictures; and problems encountered with ambiguous hyperlinks, inaccessible PDF, and screen reader malfunctions.

W3C guidelines for accessibility predict these frustrations and provide guidance to developers to create accessible websites by following the Web Content Accessibility Guidelines 2.0 (WCAG 2.0). There are few professionals that follow these guidelines, but even when the guidelines are implemented, people with disabilities encounter problems. Some authors affirm that WCAG specify a "technical accessibility as opposed to usable accessibility, not ensuring an effective user experience" proposed by [18].

Specifically, in the work of Power et al. [19], it was found that "50.4% of the problems encountered by users were covered by the WCAG 2.0." The study also indicated that "16.7% of websites implemented techniques recommended in WCAG 2.0 but the techniques did not solve the problems." Thus, the use of the W3C–WCAG guidelines is not enough to guarantee accessibility [19]. Furthermore, as automatic assessment tools merely check guidelines through the HTML tags [20], they also were not an effective tool to perform a complete accessibility and usability assessment on its own.

It is considered that all assessment techniques have strengths and weaknesses; however, if several combined methods for finding accessibility and usability problems affecting blind users are used, we can identify a maximum set of problems, as described by Mankoff et al. [21].

Another important aspect that regard ensuring accessibility and usability in Web interaction by blind users is the appropriate functioning of the screen reader software. Screen readers are tools that "read" Web page text aloud and are considered a useful technology devised for blind users because of its usage in their Web interaction. The software can interpret HTML code and read it aloud (with a synthesized voice). The interaction can be made by way of using a normal keyboard or, alternatively, by using a Braille keyboard. There are specific guidelines in the W3C–WCAG 2.0 regarding this technology; however, it only guarantees "technical readability", i.e., if the screen readers function. The guidelines do not identify if a website is "accessible" by blind users, i.e., if blind users can effectively access it [18].

With this in mind, we present a complete accessibility and usability assessment study of an e-commerce platform, using different assessment techniques and methods, allowing to identify interaction problems, i.e., design and screen reader problems, presented in the next section.

3 Materials and methods

The accessibility assessment was based on the verification of the conformance levels of WCAG 2.0, through automatic tools and manual evaluation.

An automatic evaluation system has the ability to conduct tests on an html page/website, from a URL/URI supplied by the user or database, by checking the guidelines of a standard (i.e., WCAG 1.0, WCAG 2.0, Section 508) in terms of usability and accessibility, among others, and producing reports that identify the deviations found and suggesting ways to correct them. Some more sophisticated software products provide dashboards with indicators to rank a website in terms of a guideline [22]. After this automatic evaluation, we proceeded with a manual evaluation where three accessibility experts confirmed all errors and warnings provided by the automatic tool, and confirmed false positives.

The accessibility evaluation was followed by the usability evaluation, where three experts were asked to make their assessment based on heuristics. The chosen heuristics are based on the work of [23–25].

As previously stated, some assessment issues remain, that can only be answered when the real users experience the interface. To that end, we performed user tests.

Regarding the research method, we also chose one that involves the target audience (blind people) in a real context (interacting with the interface). Thus, the case study method was used as the intention was to investigate a contemporary phenomenon in a real situation. This is a qualitative research method. The overall research approach was, however, mixed as it involved various research efforts and perspectives and made possible the generalization of the facts observed in a particular case, transferring them to general cases.

We also herein describe the data collection methods used, which are directly related to the research methods adopted: logbooks (very common in the case study method), document analysis, questionnaires, interviews, and direct observation. Specifically, as concerns observation, we used the think-aloud technique, and the feedback from the participants was recorded through written notes; with regard to questioning, the technique used was to conduct post-test interviews and questionnaires.

Two questionnaires were used. The first questionnaire, regarding the usage of ICT and user satisfaction, aimed to characterize the target audience, specifically with regard to their IT knowledge—we took into account the recommendations of [23, 26]. The ICT questionnaire had two dimensions: the first on computer literacy (if users had a computer, Internet access, and how often they used these technologies and for what purpose); and the second, on the usage of ICT, in what context they used these technologies. The second questionnaire was used to assess the degree of user satisfaction and was adapted from the SUMI (Software Usability Measurement Inventory) questionnaire. This questionnaire had 25 questions, divided into five sections: 1–9 concerned the structure of the site; 10–16 were related to aspects linked to navigation; 17–20 asked questions about the interface; 21–24 referred to the contents and the last question, 25 to the overall assessment of the Website.

In the usability evaluation, we assessed the efficiency, effectiveness, and satisfaction of blind users. The quantitative variables registered were time to complete the task, number of attempts to complete the task successfully, number of dropouts. The qualitative variables included the difficulties encountered, comments and concerns raised by the user.

It was decided to make individual evaluation sessions, only with an observer and a volunteer, so that the participants feel relaxed and comfortable, simulating a normal interaction with the Web. The intention was for the users to understand that it was the website that was under assessment and not the individual.

These evaluation sessions took place in different locations in the city of Porto—at the Rodrigues de Freitas School, at the Faculty of Arts, University of Porto (FLUP), and also in some houses of the participants who kindly welcomed us. At the end of the sessions, we analyzed the results and conclusions were formulated.

4 Assessment stages

The e-commerce platform chosen to be tested was one of the first chains of hypermarkets in Portugal, a reference in the food retail sector in the country, not only dedicated to the marketing and distribution of food products but also providing a wide range of products and services in areas such as health, home, toys, animals and others. This online platform allows for free access and purchase of all products and services, also offering delivery at home, one feature considered by the blind consumers as an asset. Therefore, we believe that it is necessary to perform a complete accessibility and usability assessment to identify possible flaws preventing blind users from enjoying a fully inclusive access.

In the following, we present three assessment stages made with the objective of performing a full accessibility and usability study of a Portuguese e-commerce website. First, we performed an automatic and manual accessibility evaluation; then we executed a heuristic evaluation with experts; and finally, we carried out user tests.

4.1 Accessibility: automatic and manual evaluation

In this first evaluation phase, we used two accessibility evaluation methods, i.e., automatic and manual evaluation. In the first approach, we used an automatic tool, *SortSite,* and following this report, accessibility experts executed a manual assessment to detect error and/or non-compliances of the success criteria that the automatic tool did not detect.

4.1.1 Methods

The accessibility assessment was based on the verification of the conformance levels of the W3C accessibility guidelines, through automatic tools and manual evaluation. WCAG 2.0, are the guidelines most often used, as there is a large number of automated tools that facilitate their evaluation. In this regard, first we used an automatic tool, *SortSite*. Despite its strength in identifying accessibility errors, it does not have the same ability to assess the accessibility of a website as a real user. Next, we presented the guidelines to a specialist, who manually verified each guideline and logged all failures observed. Two problems were verified concerning the automatic tool's results: false positives (when something is detected as correct, but is actually an error) and false negatives (when an error is detected and then it is confirmed that the error does not exist).

The automatic evaluation was performed on all Web pages of the platform. On the other hand, the manual evaluation was only performed on Web pages where the end user would have some interaction (when performing user tests). Still, we ascertained the entire platform to see if there were large disparities between the overall evaluation of the site and the Web pages in question.

4.1.2 Results

First, a summary of the description of the errors identified by the automatic tool, *SortSite,* is presented.

In a first analysis, it was found that there are seven Level A, one Level AA, and two Level AAA errors. Specifically, we present the errors through the description of non-compliance with the success criteria. Regarding the first principle of WCAG 2.0, Perceivable, two Level A and one level AA cases of non-compliance were identified. These were as follows:

- Guideline 1.1 (Text alternatives).

  - Non-text Content (Level A): The links, images, and complementary text descriptions had the same description as the ALT text and this caused the screen reader to read the same text twice. This failure was present on pages where a list of products exists. Each product has an associated name and image, which are links with exactly the same text alternative.

- Guideline 1.3 (Adaptable).

  - 1.3.1 Info and Relationships (Level A): It is necessary that the headers of rows and columns of tables have to be identified using TH elements while checking the layout of tables with role = 'presentation', as the data tables allow screen reader users to understand the relationships between columns and rows. If these elements are not properly identified, it is impossible for the user to perceive its content. This failure was identified on the icons related to social networks, throughout the content of the pages that were developed from tables, and also in specific product pages where product information was presented (these pages are also formatted as tables) and when we clicked on the icon of "more information" of the product.

- Guideline 1.4 (Distinguishable).

  - 1.4.3 Contrast Minimum (Level AA): The colors of the foreground and background must have sufficient contrast. This failure is found mainly in the LOGIN specific location. Some users could find it difficult to correctly display content, and as such, both the color and background text must be taken into account while finding the optimal contrast between them, to make them more noticeable.

Regarding the second principle of WCAG 2.0, Operable, four Level A and three Level AAA cases of non-compliance were identified. These were as follows:

- Guideline 2.1 (Keyboard Accessible).

  - 2.1.1 Keyboard (Level A): All ONCLICK events must have an equivalent ONKEYPRESS event, that is, when users are unable to use the mouse, they should be able to select using the keyboard as an alternative. Thus, there must be a keyboard event equivalent to the mouse to help these users. This failure occurred on the HELP hyperlink, specifically on the homepage of the website; it also occurred when we clicked on the FULL HELP button;

  - 2.1.3 Keyboard (No exceptions) (Level AAA): As seen in item 2.1.1, also here, in 2.1.3., we verify the

same situation. The failure occurs with the HELP link and also with the FULL HELP button.

- Guideline 2.2 (Enough Time).

  - 2.2.2 Pause, Stop, Hide (Level A): The user of assistive technologies has no way to pause the content, making the interaction with the website confusing. This failure was found in the presentation on the banner animation, in the right upper corner of the HOME page. The animation duration is longer than 5 s and cannot be turned off or stopped.

- Guideline 2.4 (Navigable).

  - 2.4.1 Bypass Blocks (Level A): The failure occurs in the iframe (Goggle Tag Manager), having no title assigned. As there is no TITLE attribute associated with the iframe on the page, the screen reader does not identify the iframe. This default covers the entire website;

  - 2.4.4 Link Purpose (In Context) (Level A): The failure in the success criterion occurs when displaying the icons of social networks, as well as with the option choice of language (bottom left). These identified links only contain non-text content, i.e., the links were composed by images, and these cannot be identified with an accessible name in ALT attribute. To overcome this breach, each tag "A" must contain text or an IMG with the ALT attribute, in which case this does not apply;

  - 2.4.9 Link Purpose (Link Only) (Level AAA): If on a page there are multiple links using the same text on the link, all those links should point to the same destination. All links with the same text should clearly point to the same destination so as to not mislead users or shuffle their browsing. This problem occurred with the links that say "Campaigns"—one on the top menu bar and another in the "Sitemap", which had a subsection campaigns;

  - 2.4.10 Section Headings (Level AAA): Each content section should begin with a heading tag (H1, H2, H3), but the pages did not have header tags, which made it very difficult for screen readers to realize the proposed structure. The non-compliance of this item occurred at the beginning of every page.

Regarding the third principle of WCAG 2.0, Understandable, one Level AAA non-compliance case was identified as follows:

- Guideline 3.2 (Predictable).

  - 3.2.5 Change on Request (Level AAA): For this failure not to occur, it is necessary to avoid opening

windows without prompting the user or opening new windows or opening a window using a link with target = "_ blank" (the user must be in control and aware of what is happening). This non-compliance was found on the client card icon and on all icons of social networks and it could also be found at the bottom of the link "Site Map", where all the links in this section opened in a new tab with the target = "_ blank".

Within the forth principle of WCAG 2.0, Robust, one Level A and one Level AA cases of non-compliance were identified. These were as follows:

- Guideline 4.1 (Compatible).

  - 4.1.1 Parsing (Level A): This error was only visible in the page code. Easily identified with the automatic tool, being that the page displayed mark-up errors, causing the screen readers to not function efficiently. Also, the failure was verified in the "notLoggedPopup" ID since the pages had duplicate IDs, causing malfunctions to screen readers, and these IDs were used to control the labels and titles of the tables;

  - Name, Role, Value (Level AA): this failure occurred in multiple situations: when the icons of social networks appeared, as well as with the language of choice option; it occurred in the area of the Newsletter where the form did not have the associated label element, i.e., the assistive technologies could not describe the purpose of the form control; it could also be found in the filters (price, brand, category, etc.) available on the right side of the website—in this case, there was a label element though it was not attached to the checks that rank it. Furthermore, in the available filters, there was another failure—this time the LABEL elements are blank, which does not allow assistive technologies to properly inform users; a failure occurred in the product listing as well, when selecting the amount one wants to buy, and also in the filters available on the right side of the screen, where there is a form; also, the error appeared in the area of the Newsletter, where there was a form with no LABEL elements associated.

4.2 Usability: heuristic evaluation

In the first stage of usability evaluation, we performed a heuristic evaluation, where two types of data were collected: quantitative, where we recorded the number of non-compliances for each of the heuristics and identified their locations (URLs) and the severity of the errors found; and

qualitative, involving the comments made by the evaluators. This review aimed to analyze ten typed pages (templates), representative of information architecture, content, structure, presentation, and interaction available on the e-commerce platform, which allowed users to select products and checkout with the 'home delivery' option.

4.2.1 Participants

Three usability experts were invited to perform the heuristic evaluation: two multimedia and computer engineering professors of the University of Trás-os-Montes and Alto Douro, and one researcher in the area of human–computer interaction of INESC TEC. They were invited by e-mail and made the heuristic evaluation by filling out an evaluation table previously created, which contained the heuristics, the classification of errors found, and descriptions of problems and their respective locations (URLs).

4.2.2 Methods

For this evaluation, we created a table with heuristics based on the 10 heuristics of Jakob Nielsen and Rolf Molich and five of the six heuristics compiled by Figueiredo [24]. These heuristics are also supported by Mark Pearrow and Steve Krug [17, 18]. Thus, the final table contains 15 heuristics and respective sub-heuristics, in a total of 68 items. To sort the heuristic problems encountered, the experts classified the defaults following a scale proposed by [27], which combines the frequency with which the error occurs, the impact of the error, and the persistence of the error:

0   I do not agree that this is a usability problem;
1   Only a cosmetic problem: correct only if there is time;
2   Minor usability problem: it must be corrected, but can be saved for last;
3   Major usability problem: must be fixed and must be a priority;
4   Usability catastrophe: It is imperative to resolve this issue.

The analysis focused on the main and only version of the e-commerce platform.

4.2.3 Results

Of the 15 heuristics inspected 9 presented problems. Figure 1 shows, for each heuristic, the proportion of the respective sub-heuristics problems.

Each heuristic was subdivided into sub-heuristics, for a total of 68. Of these, 16 (24%) had some type of problem and 52 (about 76%) were not mentioned as having any kind of problem. Specifically, the heuristics and respective sub-
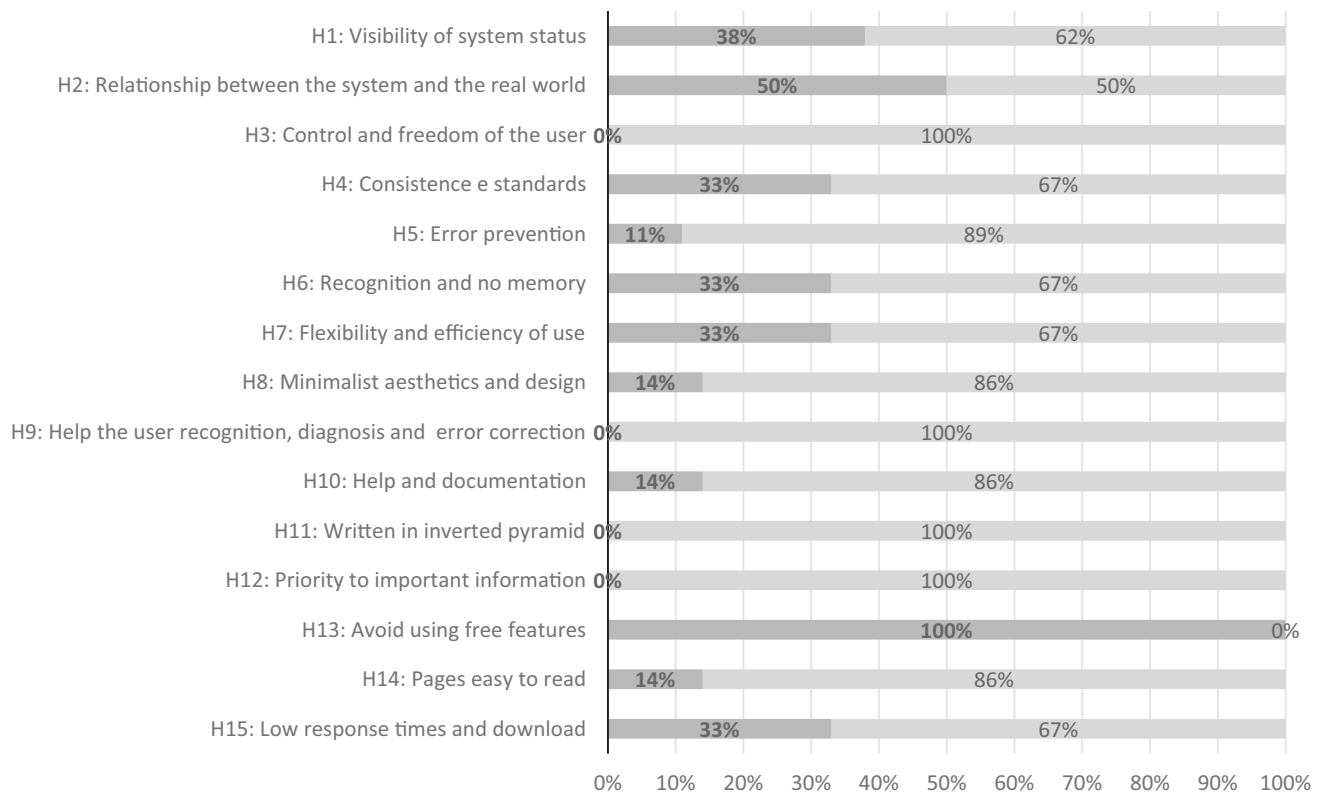
Fig. 1 Percentage sub-heuristics with problems regarding their respective heuristic

Table 1 Grade 1 problems detected by the evaluators during the heuristic evaluation

Heuristic description	Sub-heuristic description	Problems found
H4: Consistence and standards	4.6: Hyperlinks are noticeable and uniform, distinguishing the active, the visited and unvisited	There are no distinctions between hyperlinks
H10: Help and documentation	10.2: The help option is easy to find	The help hyperlink is not easy to find
H13: Avoid using free features	13.1: Existence of purely decorative or useless elements	On the homepage, there are several elements that have no utility and are merely decorative
H15: Low response time and download	15.2: The site loads quickly	The site is not very fast on loading

heuristics were classified according to the severity of the problem encountered. Of the 68 sub-heuristics evaluated, 52 were classified as severity 0, i.e., had no problems. The remaining 16 sub-heuristics were classified with degrees of severity between 1 and 4. Respectively, four sub-heuristics problems were obtained with grade 1, four with grade 2, three with grade 3, and four with grade 4.

Next, the problems encountered were classified in tables. Table 1 lists the problems classified as grade 1 (cosmetic problem).

Concerning sub-heuristic 4.6, experts confirm that there was no distinction between the various states of the hyperlinks, making them barely perceptible regarding the state in which they were (among active, visited and

unvisited). In sub-heuristic 10.2 (help is easy to find), experts considered it was not easy to find the help option, because it was hidden in the start menu or in a feature at the bottom right of the screen. Regarding 13.1 (existence of purely decorative or useless elements), it was found that in the homepage there were several elements that did not bring any benefit to the interaction with an e-commerce platform. Finally, in 15.2 (the site loads quickly), experts stated that the loading of the initial page, the product category page, and help page were not very fast.

The problems identified by the experts and classified as grade 2 (minor problem usability) are listed in Table 2.

With regard to grade 2, experts began by describing a problem in sub-heuristic 1.8; that is, in a particular

Table 2 Grade 2 problems detected by the evaluators during the heuristic evaluation

Heuristic description	Sub-heuristic description	Problems found
H1: Visibility of system status	1.8: I always know where I am	In at least one page, the location of the user is blank, which does not happen in other pages
H2: Relationship between the system and the real world	2.1: The pages use common words and words that are familiar to you ("speak the user language")	Some words may be strange to the user
H5: Error prevention	5.6: The images are of good quality and take some time to load	The images are too small
H7: Flexibility and efficiency of use	7.2: Existence of common command aliases	There is no mechanism for abbreviations

Table 3 Grade 3 problems detected by the evaluators during the heuristic evaluation

Heuristic description	Sub-heuristic description	Problems found
H4: Consistence and standards	4.2: Uses the same terms throughout the interface	The terms used are not the same across the interface
H6: Recognition and no memory	6.1: User should not be required to memorize the options available to complete a task	The hyperlinks of the checkout process are not objective
H13: Avoid using free features	13.2: User attention deviation from the main goal	There are several elements that divert the user's attention

hyperlink, they found a problem of users' non-location, and this position is blank, which does not occur in other pages of the platform. As for 2.1, experts said there were some pages where the words were not so common or familiar, and they were very technical and difficult to understand in an e-commerce context. In sub-heuristic 5.6, images were too small to be considered to have a good visualization. Finally, in 7.2, they found no abbreviation mechanism for frequent commands used to enhance efficiency in the interaction.

The problems identified by the evaluators and classified as grade 3 (major usability problem) are listed in Table 3.

With regard to grade 3 (major usability problem), the experts began by describing that the titles of the checkout process were not consistent, and this could lead users to feel "lost" and confused during the interaction. Specifically, the section name changed from Product listing (untitled) to SUMMARY, on the second step of checkout. In sub-heuristic 6.1, it appeared that the links of the checkout process were not objective; it is believed that if we spend time without visiting the site, we will encounter interaction problems on our first-time visit. Concluding, in 13.2, there are several goals that only serve to direct the user to pre-planned sales mechanisms defined by the company that owns the site.

The problems identified by the experts and classified as grade 4 (serious usability problem) are listed in Table 4.

In relation to grade 4 (usability catastrophe), the experts began by describing problems with the feedback that the site provided; it gave no feedback regarding the users' action or concerning information to proceed. Explicitly, in

sub-heuristic 1.1, it appeared that the platform did not return any information while navigating along the pages, nor when the user performed an action. Moreover, in 1.2, experts determined that the platform did not inform the user about what they were doing and the decisions they were taking. Also, in 2.2, experts found problems in the information provided, since the products should be arranged in alphabetical order for an easier search. Finally, in 8.7, experts affirmed that there was not enough contrast in the text, making it too difficult to read.

4.3 Usability: user tests

Our research path is based on the premise that an evaluation of accessibility and usability is not complete if it does not search for real users to interact with the website. Thus, we chose to conduct review sessions with blind users.

4.3.1 Participants

Specifically, twenty blind people participated in this study, five women and fifteen men, aged between 18 and 57. All participants were daily and experienced users of computers and the Internet. Regarding screen readers' usage, twelve participants used, preferably, a JAWS screen reader (version 14 four users, version 15 three users and version 16 five users). Two participants used NDVA, and three participants used the Apple Voice Over, as shown in Fig. 2.

As for the browser, users of JAWS often use Internet Explorer. However, depending on the site, of these fifteen four also use Mozilla Firefox, one used Microsoft Edge,

Table 4 Grade 4 problems detected by the evaluators during the heuristic evaluation

Heuristic description	Sub-heuristic description	Problems found
H1: Visibility of system status	1.1: The site provides feedback	The site does not provide feedback regarding the actions.
H1: Visibility of system status	1.2: User is always informed of what is happening.	The site does not inform the user of his own actions
H2: Relationship between the system and the real world	2.2: Information is available in a logical and orderly manner	The products are not listed in a logical way
H8: Minimalist esthetics and design	8.7: Existence of contrast between the text color and background color	There is not enough contrast in the text detail of the products

two used Google Chrome, and four Safari. Among the Voice Over screen reader users, four used Safari and, depending on the sites, two also used Google Chrome. Sixteen participants used technologies to work and four to perform schoolwork. The most commonly used programs for these functions were Microsoft Office (Word, Excel, and Access), Adobe Reader, Movie Maker, Windows Media Player, Sound Forge, and Sony Vegas. To communicate, all twenty participants often used e-mail services, such as Microsoft Outlook, fourteen used Skype, eleven used Facebook, and seven used WhatsApp. Seven participants often used the computer and Internet to play games such as Playroom, Travian, Whoever million, Papa Sangre, and thirteen participants did not play with these technologies.

4.3.2 Methods

It was necessary to select an investigation method that met the users' segment (defined target audience) in a real context (interacting with technology). Thus, we used the case study approach; it was intended that we investigate a contemporary phenomenon in a real situation, which is a method of qualitative research; the overall research

Screen Reader' usage by brand

- 3; 18%
- 2; 12%
- 12; 70%

■ JAWS
■ NDVA
■ Apple Voice Over

Fig. 2 Screen reader usage by brand

approach was, however, mixed as it involved various research efforts and perspectives and required the generalization of the facts, transferring, for general cases, the facts observed in a specific case [28].

Various data collection methods were used, directly related to the research methods adopted: logbooks (very common in the case study approach), document analysis, questionnaires, interviews, and direct observation. Specifically, as concerns observation, we used think-aloud techniques and the feedback registration given by participants was made through written notes; with regard to the questioning, the technique used was to conduct post-test interviews and questionnaires. In particular, we used two questionnaires: the use of ICT and satisfaction. The first aimed to characterize the target audience, specifically with regard to their IT knowledge; in its creation, we took into account the recommendations of [23, 26]. This questionnaire comprised two sections. The first was related to computer literacy, i.e., if users have a computer, use the Internet, how often they use them and the purpose of use. The second section was related to the actual use of ICT. The user satisfaction questionnaire was used to assess the degree of user satisfaction, with a questionnaire adapted from the SUMI questionnaire (Software Usability Measurement Inventory). The questionnaire was divided into five parts: items 1–9, concerning the structure of the site; items 10–16, related to aspects of navigation; from 17 to 20, linked to the interface; from 21 to 24, referring to the contents, and question 25 bringing together the overall assessment of the website.

For user evaluation, the efficiency, effectiveness, and satisfaction variables of blind users were registered. Specifically, the registered variables, in quantitative terms, were the following: rate of success of tasks, time to conclude the task, the number of attempts to complete the task successfully, and number of dropouts. Furthermore, as qualitative measures, we registered the difficulties encountered, as well as comments and concerns raised by the users.

We chose to make individual evaluation sessions, just with an observer and a volunteer user, so that the
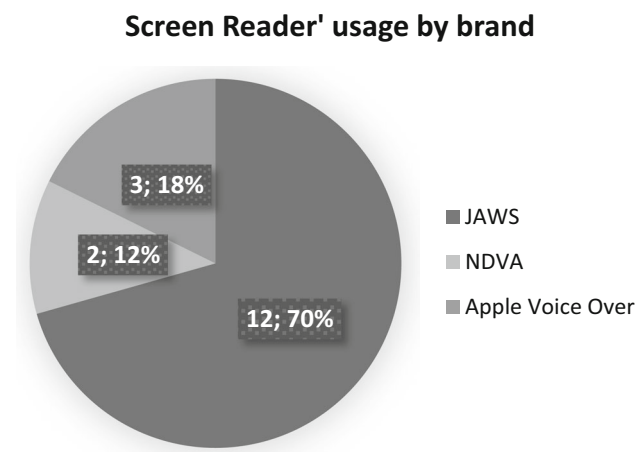
participant felt relaxed and comfortable, simulating as much as possible a normal interaction with the Web. The intention was for the user to understand that it was the website and not the individual that was being assessed.

These assessment sessions were held in different locations in the city of Porto, such as at the Rodrigues de Freitas School, at the Faculty of Arts, University of Porto (FLUP), and also at the houses of some of the participants, who kindly welcomed us. At the end of the sessions, we analyzed the results and conclusions were formulated.

4.3.3 Experimental design

For the evaluation, we specifically chose Web pages from the platform that allowed to replicate an online purchase with a focus on home delivery.

At this point, we specified the tasks proposed to blind users. We had in mind the following list of tasks: searching for specific information, requesting details on the information searched for; conducting a total count of areas of interest on the website; "yes/no" tasks, where it was proposed that the user look for an existing information on the website content and search for information on a table provided on the website.

In this context, we defined six tasks to be performed:

1. Website Identification (assessment goal: verify if the page titles are consistent, direct, and noticeable);
2. Select a specific product (Skimmed Milk UHF Mimosa) and add to shopping cart (assessment goal: check easiness of selection process/add to cart and understanding of button titles);
3. Identify the section where social networks are located and name them (assessment goal: check for correct labeling graphics and an easy identification of the location);
4. Select a second product, chosen by the participant, from a specific section, from the "Groceries", and add it to the shopping cart (assessment goal: to check the easiness of navigation by sections);
5. Identify the price of the second product (assessment goal: to see if users can use the reverse option easily);
6. Finalize the purchase (assessment goal: to evaluate the checkout process—Buy, Select Delivery, Payment, and Confirmation).

4.3.4 Procedure

We started the evaluation activity with the reading and subsequent signing of the consent form for the use of the data to be obtained. After this, the questionnaire regarding the ICT usage was filled out. Next, the participants were asked to maximize the browser window (already on the

platform's home page with a login to start the activity). The participant started the interaction using the screen reader and browser that he/she used every day. At the beginning of the activity, we invited the participants to freely explore the platform. The observer presented the tasks, one at a time, to the participants. At the end of each task, we explained the aim of the task and what it was intended to measure. During the execution of the activity, the participant could request explanations from the observer in order to carry out the task. The assessment technique adopted for the participants to orally express their opinions and feelings was thinking-aloud. This technique encourages participants to verbalize their thoughts during the execution of tasks, allowing the viewer to interrupt the task to answer questions. The evaluation sessions had a maximum duration of 1 h to 1 h and a half.

4.3.5 Results and discussion

The results are presented with regard to the usability evaluation, according to the effectiveness (completing a task in terms of quality of the result), efficiency (resources required to achieve effectiveness: experienced difficulties) and satisfaction (comfort felt by the user to use the interface).

The user test results were registered using an observation grid where, for each task, the difficulties and user comments were noted, as well as the time to conclude the task and the number of attempts made to complete the task and still further the number of dropouts.

Firstly, the effectiveness results are presented. In Table 5, we present the analysis of successful task completion or non-completion. The scattering coefficient of the success variable is also presented, which indicates that, except for tasks 3 and 6, all the other tasks were concluded with success.

Respectively, regarding Task 1 (identify the Web platform), Task 2 (select the product, Skimmed Milk UHF Mimosa and add to shopping cart), Task 4 (select a second product of the grocery section and add it to the shopping

Table 5 Analysis of successful completion of the task

Task ID	Number of participants that concluded the task successfully	Number of participants that did not conclude the task successfully	Scattering coefficient (success) (%)
Task 1	20	0	100
Task 2	20	0	100
Task 3	19	1	90
Task 4	20	0	100
Task 5	20	0	100
Task 6	7	13	30

cart), and Task 5 (identify the price of the second product chosen), all participants completed the tasks successfully.

For Task 3 (identify the section and label of the Social Networks), in the section identification, only one participant did not complete the task because he/she was not a social networks user so he/she did not recognize the hyperlink named "Share" needed to identify the section location. For the second part of the task (namely social networks), only one participant successfully completed the task because he/she opened the three links presented, leading to external websites. These social network hyperlinks had no labeling associated, and this issue caused problems in the screen reader performance because it could "read" the icons.

Finally, for Task 6 (finalize the purchase), only seven participants completed this task successfully. Thirteen participants did not complete the task successfully because the checkout process proved to be very complex. This is noticeable because the buttons and necessary links to perform the task were not objective and therefore did not allow direct identification of the function they were representing, according to participants' feedback.

In addition, the results concerning efficiency are presented—starting with the conclusion time of the six tasks, presented in Table 6.

The mean value was obtained by summing the conclusion times of the participants by task and dividing them by the total number of participants. The values are displayed in seconds (values rounded to the unit).

Knowing that the standard deviation measures the dispersion, indicating the correctness of a data set averages function, and the higher the standard deviation, the higher the dispersion relative to the average is. Thus, in Tasks 2 and 6, dispersion was higher compared to the other tasks.

Users performed Task 1, on average, in 6 s. The fastest participant took 1 s identifying the Web platform and the slowest took 20 s. It was noticed that this variation depended on the user experience in relation to its own screen reader.

For Task 2, users needed, on average, 3 min (180 s). The fastest user took 1 min and 7 s and the slowest 6 min and 13 s. Here, it was found that those who used the search

box to select the product (UHT Mimosa) were faster than those who tried to go the grocery section.

For Task 3, users needed, on average, 2 min and thirty-one seconds (151 s). The fastest participant took twenty-seconds and the slowest 6 min and 17 s. This variation was caused due to fast or slow perception of the word "Share" (often, the hyperlinks of social networks are identified by the keyword "SHARE"). So, if users identified the keyword, they quickly found the hyperlinks, which, according to their feedback, "Usually indicates Social Networks". However, if users did not do this assimilation, they took a longer time or even did not find the hyperlinks.

For Task 4, users needed 3 min and 9 s (189 s). The fastest user took 50 s and the slowest 7 min and 6 s. Also, in this task we found that the fastest users used the search box, despite being asked to use the "Groceries" section. This happened because they had their screen reader customized with shortcut keys and they were experiencing the search option or the list of links option, and these options decreased the time to complete the task. Those who chose to identify the "Groceries" section were those who needed the longest time to complete the task because the site did not display headers, and without this, header users had to listen link by link until they got to the section "Groceries".

For Task 5, users needed one minute and fifty-one seconds (111 s). The fastest user took 3 s and the slowest, 9 min and 13 s. The fastest participants quickly used the go back option and listened to the price information. Others had problems with site navigation and often lost themselves. Some of them chose to write in the search field the product selected previously and listen to the price; others tried to go to the shopping cart section, and those who could not find the section of the cart chose to try to finalize the purchase to verify if they could listen to the details of the products they had added to the cart.

Finally, for Task 6, users needed 10 min and twenty-four seconds (624 s), on average. The fastest user (who was the only one to successfully complete the task) needed 5 min and 19 s. It was found that this participant had already used this platform previously. The slowest took 17 min and 10 s.

In Table 7, we present the number of attempts made by the participants for the successful completion of the six tasks and the number of dropouts per task.

Regarding the number of attempts to complete the task, emphasis must be given to Task 1, which the participants completed on their first attempt. Up to ten attempts were needed in order to complete task 2. For task 3, twelve attempts were needed, for Task 4, fifteen, for Task 5, eleven, and for Task 6, thirty-five attempts. The last task was considered the most complex of the assessment activity.

Table 6 Conclusion time per task

Task ID	Mean (s)	SD (s)
Task 1	6	6
Task 2	180	87
Task 3	151	101
Task 4	189	110
Task 5	112	169
Task 6	624	289

In relation to participants' dropouts, in Tasks 1, 2, 4, and 5 there were no dropouts. However, in Task 3, one participant gave up while in Task 6, thirteen did so. This last task had the highest dropout rate because it was found that many difficulties in understanding the information necessary to perform the task existed, and much more complexity was found in the interaction of the screen reader and the interface.

Next, a description of the difficulties and comments made by users at the time of interaction is presented. In the first task (Task 1), no one showed difficulties in identifying the platform. Participants considered this task very easy. As soon as they maximized the browser window, the screen reader automatically read the page and informed the user what the page they were on was. Thus, we considered that the page title was objective and properly identified.

In the second task (Task 2), globally, users used the search box to search the requested product (UHT Mimosa) and did not interact directly with the section's hyperlinks. The main difficulty that users mentioned was that the platform did not have headings defined throughout the different pages. This fact was mentioned as having "greatly delayed" the interaction, because the users had to listen element by element until they reached the aim, the "Add" and "Remove" buttons. These specific buttons were considered very confusing because participants claimed that hearing the name of the buttons reminded them of the shopping cart despite the product quantity window being editable. Consequently, the "Cart" button was not very direct either, as they were waiting to hear "Add to Cart". They also mentioned that they would like to receive feedback once products were added to the shopping cart. Often they commented that they did not know whether the product was added to the cart or not. Another comment made regarding this task was that the section "Shopping Cart" was easy to identify though difficult to see how it worked. Also, some users commented that if the product name was not spelled correctly in the search box, it was more difficult to find the product, and there was no feedback regarding the misspelling.

In the third task (Task 3), most of the participants did not experience difficulties in the identification of the location of social networks. Only participants who did not use social networks on a daily basis had more difficulty, since they would have to recognize the word "Share" to find the section of the icons. Here, the icon titles (identified as graphics) were not the most direct and clearly described, as they should be. Still, they could count three social network icons displayed on the footer of the platform. The greatest difficulty was to name these icons, as they appeared to be unlabeled. However, some participants found the option of sharing on social networks to be unattractive and confirmed that they would not use it even if the icons were correctly labeled.

In Task 4, participants used their own screen reader search option to find the grocery section. Also, they tried to use the hyperlink listings. However, they claimed this was a difficult task because they could not find the section. In a second attempt, some participants experienced the search box option and others navigated element by element on the platform until they found the section, and then chose the product. With regard to selecting the product and adding it to the shopping cart, they showed no difficulty. Again, they stressed the need for feedback on the state of the product (if it was added to the cart or not). They commented that an interesting option would be that, after adding the product to the cart, this product should be inaccessible, since it would be already in the cart. Again, they highlighted that the absence of headers did not allow for an effective navigation.

Regarding the fifth task (Task 5), some participants showed many difficulties in using the "Back" button, and even felt "lost" on the platform. Another difficulty encountered was to identify the "CART" section to check the products purchased and then identify the price of the product. Participants showed a lot of frustration to complete the task. One solution presented was to try to finalize the purchase and to hear the description of the products and consequent price information, though this process was time-consuming and also very complex, and the participants gave up and had to try again using the search box

Table 7 Number of attempts made to successfully complete the tasks versus the number of dropouts per task

Task ID	Number of attempts made to successfully complete the tasks	Number of dropouts	Mean	Standard deviation	Scattering coefficient (%)
Task 1	1	0	0.5	0.5	100
Task 2	10	0	5	5	100
Task 3	12	1	6.5	5.5	85
Task 4	15	0	7.5	7.5	100
Task 5	11	0	5.5	5.5	100
Task 6	35	13	24	11	46

option, writing the name of the product so that they could then hear the price. One of the comments most often made regarding this task was related to text illegibility, as in the following example: "The euro symbol appears before the price, so the screen reader reads: EUR 1 point 99 per unit; and it should not be like this, we do not speak that way, it should be: 1 point 99 Euros per unit."

Finally, in Task 6, all participants found the buying process to be very difficult and time-consuming, firstly because of the complex nomenclature and also because of the information structure of the whole process. They did not think that the hyperlinks to the completion of the process were clear or direct, and the titles did not have a description that directly identified the function. First of all, they needed to select the "Buy" button, which all found relatively easy. However, they again had problems with site navigation because of the non-definition of the headers, making it harder to complete the task and taking much more time (they had to run the platform, element by element) to find the "Buy" button. Then, they had to choose the "SELECT DELIVERY" option. They commented that they would like the hyperlinks to be called "SELECT DELIVERY DATE". The procedure requirement is that the home delivery option must be chosen. Here, the first interaction problem appeared, as selecting the hyperlink "SELECT DELIVERY", opened a secondary window with a table. In this table, the greatest difficulty was to understand the information presented. Users expected to select a delivery day and hour; however, they needed to select the price of delivery, instead. Another difficulty was related to screen readers. In older versions, one cannot use the selection from the screen reader, so they could not select any row/column of the table. What happened was that users had to activate the mouse click simulation option on the screen reader, to be able to select the desired cell. Many claimed that the interaction was too complex because there was a great cognitive effort for the interpretation of the table, and then a major technical effort to resolve the issue of the mouse click simulation. In this step, the Voice Over screen reader seemed to facilitate this action. Another comment was that, after selecting the delivery price, there was no feedback on whether this selection was well executed or not, leaving the user again with many concerns about what happened. Then, they needed to choose the payment option. In this step, users had two options: to select the "PAYMENT" hyperlink, they clicked at the bottom of the page, after the description of the products; alternatively they could select the same "Payment" link in the "Cart" section. However, the "CART" section name changed to "SUMMARY", which was difficult for users to understand, once they were accustomed to look for the "CART" section. Also, users commented that the counterintuitive link should be "ADVANCE TO PAYMENT",

since they still had not yet finished the payment process (they still had to choose the way of payment). Next, users had to select the payment method. However, after selecting the link "PAYMENT", they commented they felt "lost" because the screen reader began to read the page from the login and password area, letting them know what they had to do to complete the operation. Here many participants asked, "What now? What is intended for me to do to complete the purchase?" The observer had to explain that they had to select the form of payment. After confirming the payment mode, the site redirected to a new page, "Checkout". Here, the page remained on the CONFIRMATION step. However, the user had to choose the "Check out" hyperlink. Again, users commented that they felt a little lost because it was not clear (they had no feedback) if they had already completed the process or had to select any additional link. After selecting the "CHECKOUT" link, the same happened again; the screen reader started reading the first page elements, login and password, and the user had to navigate element by element to hear the successful order submission message, making the interaction very complex and tedious. At this point, users frequently asked if they had finished the task, indicating again that the feedback page was not enough and did not permit clear identification of the point at which they were interacting.

Other comments were related to the discounts available. Users found that the discount information was only in the description of each product, when they added some product to the shopping cart. The platform gave no feedback on the discount elsewhere. Users only got this information after adding the product in the product listing, in the confirmation of purchase or again on the product detail page.

Regarding satisfaction, the results of the questionnaire adapted from the SUMI questionnaire are presented in Fig. 3.

Analyzing the answers of the SUMI questionnaire concerning the site structure (items 1 to 9), users did not find it easy to use (70%), found it to be too slow (45%), would not recommend it to other colleagues (65%), and considered the use of the platform a waste of time (60%). Also, the need to perform many steps to do an operation was stressed (80%). All participants (100%) stated that the site did not always do what was expected, 45% felt it was not easy to change from one part of the platform to another, and it was very difficult to learn to use all of the platform's potential. 80% of the users did not find it easy to understand all options available on each page.

With regard to issues related to navigation (items 10 to 16), 45% of the users found it easy to learn how to navigate the site, and with the exception of one user, no one was amused to navigate it. Regarding the forward and backward options, users agreed it was easy to use in a proportion of
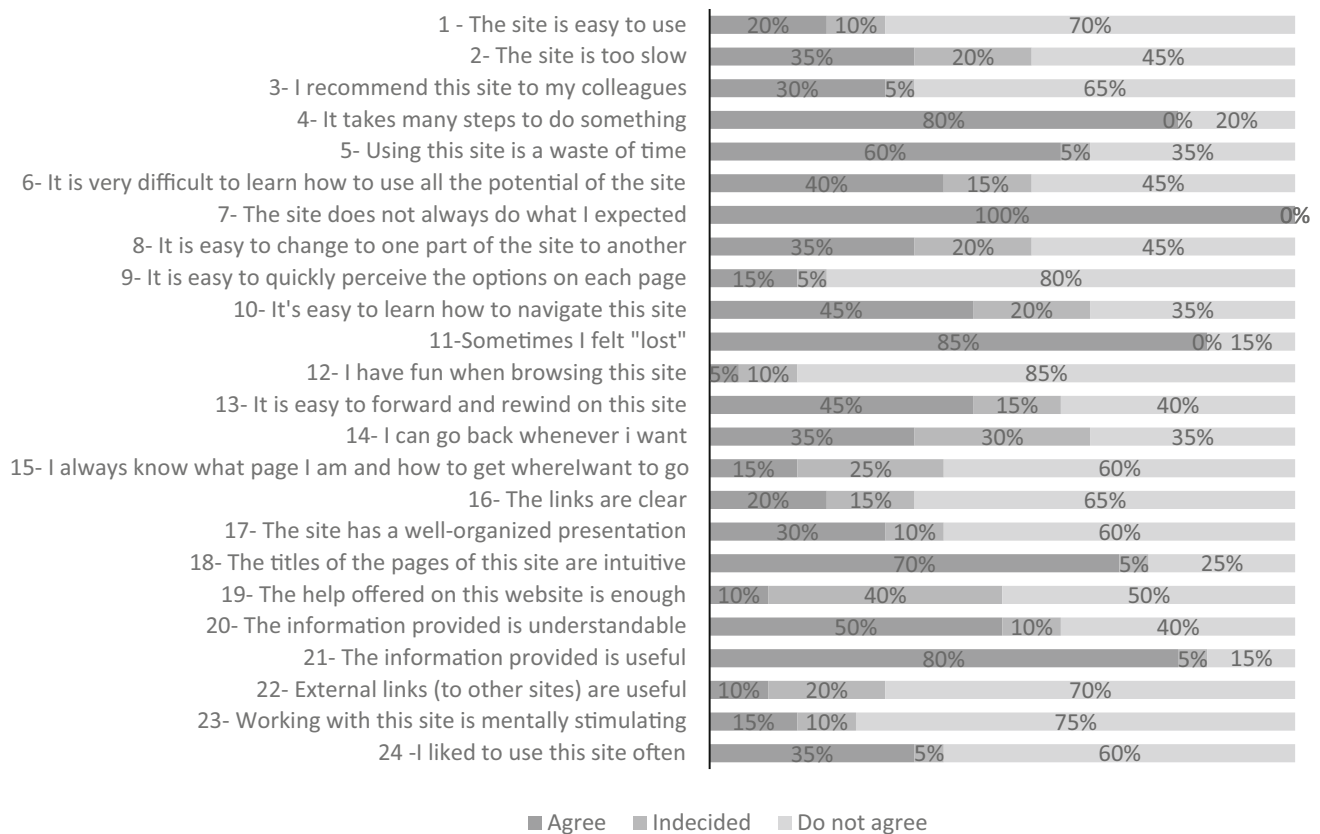
Fig. 3 Results of the satisfaction questionnaire

45%, while 65% of users found that the links were not clear and that they did not always know where they were (they felt lost) or how to get where they intended to go (65%). Seventeen users (85%), "felt lost" sometimes, and three users were undecided on this issue.

Moreover, 65% of the users found that the links were not clear, as they did not always know where they were and how to get where they intended to go. Seventeen users (85%) said they sometimes "felt lost", and two users were undecided on this issue.

As for the interface (items 17–20), 60% of the users considered that the site was not well organized, however presented intuitive titles, and 50% thought that the information provided on the site was understandable, though stated that the aid provided was not enough.

Regarding the content (items 21 to 24), 80% of the users considered the information provided useful. With regard to external links, 14 users considered them useless (as they were not labeled correctly). In general, 75% were not excited to work with this site and 60% said they would not want to use this site frequently.

To conclude, the overall rating of the perceived quality of the site by participants in this study is presented next (Fig. 4).
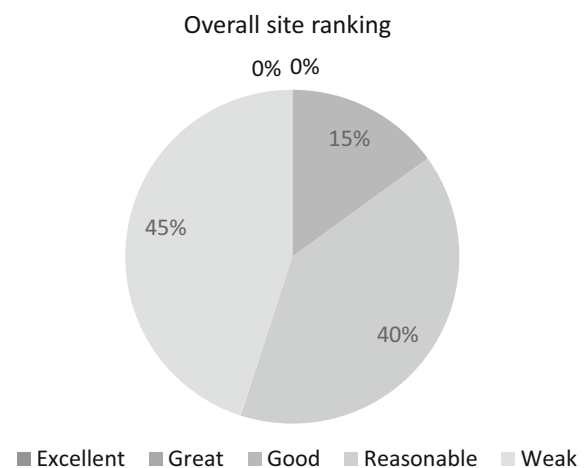

Fig. 4 Overall site rating

The overall assessment of the site, performed by users, considering all the parameters discussed above, ranged between reasonable and weak. Comments made by the participants when filling out this questionnaire were also registered below, because it is important to have a perception about the overall assessment and the weak results it generated:

"I did not know if it was the Internet that was slow or the platform"; "They could simplify some steps"; "Working with this site was not exciting, it was tiring"; "I had fun and it did not make me give up"; "Navigation is not very intuitive, especially the part concerning the checkout"; "I will not recommend this site to anyone, if it does not improve"; "The hyperlinks that were well identified were clear, the problem is the hyperlinks that were not well identified"; "It was not easy to navigate from one part of the site to another because the site did not have headers"; "Working with this site was not a very positive experience"; "I had a hard time finding things"; "With time to learn I think I could use all the potential of the site, but it is a lot of work"; "Working with this site is stultifying"; "The information provided was useful, but it was not very accessible"; "Sometimes I felt frustrated"; "I have not seen any help option"; "I was unable to identify external links"; "I did not understand why the page expired when I tried to go back"; "The website did not do what I wanted, as in the case of pop-up windows, I did not realize that they were there"; "To realize where I am is very difficult"; "Working with this site is not stimulating, I wanted to give up in the middle of the process"; "The contents organization was a little confusing"; "I would use this site often if the checkout process became more accessible"; "The site has been better."

5 Final considerations

To perform complete accessibility and usability assessment sessions is considered quite enriching because it allows for the extending of knowledge to the real state of the platform and the real needs of the participants.

Initially, it was possible to understand the problems detected by the automatic assessment tool, *SortSite*. The method used involved the assessment of specific pages with which the end user interacts to make an online purchase. Assessing only those pages did not change the overall assessment of the site, because all pages generally behave in the same way. Thus, with this analysis we realized that all scanned pages contain non-conformities with the WCAG 2.0 guidelines, and the most affected compliance level is the one with the highest priority. This situation allowed to observe that the existing accessibility errors posed a serious barrier making user interaction more difficult. More detailed observations of this issue revealed some of the detected errors, such as "ONCLICK event" without a "ONKEYPRESS" associated event; several tags "a" did not contain text or images using the corresponding "alt" attribute; forms without any label associated; failure to identify row and column in the data tables, using the

"TH" element and checking the layout of tables with "role = 'presentation'"; as well as a lack of "TITLE" attributes in the tables on pages; duplicate IDs on pages that cause problems for screen readers and pages with mark-up errors.

In a second assessment, the heuristic evaluation helped to identify the most problematic usability issues of the platform. The results show that from 68 sub-heuristics, 16 (24%) had some type of problem. Briefly, these 16 sub-heuristics were classified in different degrees of severity between 1 and 4, four sub-heuristics problems with grade 1, four with grade 2, three with grade 3, and four with grade 4.

Specifically, for grade 1 (a cosmetic problem that should only be addressed if time permits), the hyperlinks may not be noticeable since there is no distinction between active, visited and unvisited; help is not easy to find, but as the telephone contact number is available, this can be an option to address questions or clarify issues. Evaluators found many purely esthetic/decorative elements that do not facilitate the interaction and may even make it more complex. Also, it was verified that the site was not very fast. In the case of grade 2 (minor problem usability whose resolution should not be a priority), there were problems with the system status visibility, as at least one link in the user's location was blank; also, some strange words not used much in the context of e-commerce were found. They claim that the site displays small images that can be difficult to view for people with visual impairments. It also appears that there are no abbreviations for frequently used commands that could facilitate and reduce the time of interaction. Regarding grade 3 (higher usability problem whose resolution should be a priority), it turns out that there is no consistency of terms across the interface. There were some hyperlinks that require great cognitive effort, as is the case of checkout options. Various elements that divert the user's attention could also be found. Finally, for grade 4 (usability catastrophe whose resolution should be an imperative), evaluators claimed that the platform did not provide proper feedback and users were not always informed of what was happening. The information is not available logically, as is the case for product listing, and there is not enough contrast in the description text of the product and the background, and this is problematic for people with visual disabilities and dyslexic users.

With the feedback provided by the blind users, we created a list of recommendations (and also a comparison with relevant published guidelines and recommendations) that could, in the future, be implemented, to increase the level of accessibility and usability of this e-commerce platform. These recommendations are summarized as follows.

- "To reorganize the site with headers": in addition to getting a clearer organization of content this would also help to reduce the time that the blind users spend on finding products and services.

This user recommendation follows the navigable principle (guideline 2.4) provided by the W3C–WCAG (2.0) that indicates that developers must "provide ways to help users navigate, find content, and determine where they are." For blind people, it is very important that the website is developed following the guidelines that describe the good functioning of the screen reader functions, using headers. It is one the most important features for an effective interaction by blind people.

- "To provide clear identification of the graphic elements, whether they are a graphic chart, a button or a form": all elements must have mandatory labels, which should be as clear and objective as the context they represent.

Users must be sure of the elements' function to have an effective interaction. The W3C–WCAG 2.0 presents the first principle, to be perceivable. Specifically, guideline 1.1 defines the importance to "provide text alternatives for any non-text content so that it can be changed into other forms people need, such as large print, braille, speech, symbols or simpler language." Users underlined the importance of functional elements (whether text or not) having a clear description (label), title and an alternative text.

- "To provide more feedback in specific and crucial situations, such as when selecting important links"; for example, when adding a product to the shopping cart, when the product has a discount or when users need to increase or decrease the quantity of goods.

This user recommendation meets the second W3C–WCAG 2.0 principle, to be navigable. Particularly, guideline 2.4 indicates that developers must "provide ways to help users navigate, find content, and determine where they are." Noticing audio feedback of their actions is very important for blind users.

- "The search algorithm seemed to be optimized in relation to the store policies (merchandising policies) and not according to the clearest keyword to search for, and find, products"; for example, when users searched for carrots, the first product that appeared was a moisturizing cream.

This user recommendation matches the third W3C–WCAG 2.0 principle, to be understandable, which indicates that the "information and the operation of the user interface must be understandable." Specifically, we underline that it is crucial that the correct keywords (text) provide the

correct product or service searched for. This feature must be re-thought to adjust to the users' needs, particularly for blind people, because performing searches is one of the most used functions.

- "Crucial functions of e-commerce platforms, such as the CART section, must be in a clearly identified section to facilitate its interpretation and proper functioning": many users had difficulties to recognize this important section and perform tasks in it, such as: adding or deleting products, returning to the option, searching for quantities and prices.

The recommendation is specific for e-commerce platforms but can also be upgraded for crucial sections and elements of different types of platforms and is directly related to how well-structured platforms must be, enhancing the effectiveness and efficiency of users' interaction. In this context, we believe that this section must be reorganized, creating a list of products added with clear information about quantity and price.

- "Overall, the platform needs to be more accessible and usable, for any browser and screen reader (all versions)": the interaction of blind users depends directly on their screen reader, the browser used and their experience in the Web environment. So, interaction must be simpler and clearer to be accessible to users with less experience.

Globally, this recommendation follows the fourth W3C–WCAG 2.0 principle, to be Operable, that is developing websites in order to "maximize compatibility with current and future user agents, including assistive technologies" (W3C–WCAG 2.0, guideline 4.1); it is crucial when we comprehend how blind people interact with websites, using different technologies which at the same time must be compatible.

The users' recommendations are directly related to the W3C–WCAG 2.0 guidelines. We believe that, during the development of the e-commerce platform, developers needed to follow W3C–WCAG 2.0 and this effort helped provide a truly accessible Web experience for blind people. Nevertheless, as it is, the website is not accessible for blind people.

Most of the items presented herein, as referred to in the work of Lazar et al. [17], are impairments to user interaction, such as inappropriate form and graphic labels and confusing page layout, and are relatively simple to solve if developers focus on this effort. Regarding screen readers, developers must take into consideration that different users use different versions. The same happens with browsers.

One of the key observations made is that users had a great cognitive load because of the complexity of

interaction with the platform. They must set up strategies when working with tables, as for the selection of delivery or when they need to navigate without access to the headers (for example, they usually said that they need to always start at the bottom of the page or mount the table in their head to understand). Notice that while a user without visual impairments needs less than 20 s to execute the purchase process, a blind user, due to the complexity of the inter-action and the cognitive effort required to recognize the links, needs on 10 min and twenty-four seconds (as measured in the experiment described in this study). The participants in this study sustain that the blind users lost, on average, 30.4% more time due to situations that limited their user interaction.

Users were able to perform the five tasks proposed without major difficulties. However, the last task (Task 6), the most important one, i.e., to pay and choose where the products were delivered, was found to be very hard to fulfill, and the dropout rate was very high. Blind participants were likely to repeatedly perform tasks until they successfully concluded them, using different strategies (changing browsers or screen reader version); however, Task 6 was impossible to perform by a blind user as it could not be accessed with the screen reader or any other assistive solution.

Thus, overall user satisfaction was negative, classified between reasonable and weak, with the exception of three users (out of a total of twenty) who found the overall qualitative ranking of the site to be good. However, these three users identified the need to expend effort to make the site more accessible for the blind.

References

1. Gonçalves, R., Martins, J., Pereira, J., Santos, V., Cota, M.P.: Can I access my school website? auditing accessibility of the portuguese teaching institutions websites. J. UCS **19**(18), 2639–2655 (2013)
2. Martins, J., Gonçalves, R., Branco, F., Pereira, J., Peixoto, C., Rocha, T.: How Ill Is Online Health Care? An Overview on the Iberia Peninsula Health Care Institutions Websites Accessibility Levels. In: New Advances in Information Systems and Technologies. Springer, pp. 391–400 (2016)
3. Rocha, T.: Interaction metaphor for Access to Digital Information an autonomous form for People with Intellectual Disabilities. Doctoral Thesis. University of Trás-os-Montes and Alto Douro, (2014)
4. Tânia, R., Diana, C., Ramiro, G., José, M., Frederico, B., Maximino, B.: Usability Evaluation of the Touch Screen and Mouse as Input Devices by People with Intellectual Disabilities (2015)
5. ISO (2008) ISO 9241-171—Ergonomics of human-system interaction—guidance on software accessibility. http://www.iso.org/obp/ui/#iso:std:iso:9241:-171:ed-1:v1:en. Accessed 20 Jan 2016
6. Baptista, A., Martins, J., Gonçalves, R., Branco, F., Rocha, T.: Web accessibility challenges and perspectives: A systematic literature review. In: 11th Iberian Conference on Information Systems and Technologies (CISTI), 2016, AISTI, pp. 1–6 (2016)
7. Rocha T, Martins J, Gonçalves R, Branco F: Usability evaluation of an entertainment platform by people with intellectual disabilities. In: 11th Iberian Conference on Information Systems and Technologies (CISTI), 2016. AISTI, pp 1–5 (2016)
8. W3C-WAI (2008) Web Content Accessibility Guidelines 2.0. http://www.w3.org/TR/2008/REC-WCAG20-20081211/. Accessed 28-01-2011
9. Steed, A., Tromp, J.: Experiences with the Evaluation of CVE Applications. In: Proceedings of collaborative virtual environments, 1998. pp 123–127 (1998)
10. Karoulis, A., Sylaiou, S., White, M.: Usability evaluation of a virtual museum interface. Informatica **17**(3), 363–380 (2006)
11. Hix, D., Hartson, H.R.: Developing User Interfaces: Ensuring Usability Through Product & Process. Wiley, Hoboken (1993)
12. Kreps D (2008) How the web continues to fail people with disabilities
13. Law, L-C, Hvannberg, E.T.: Complementarity and convergence of heuristic evaluation and usability test: a case study of universal brokerage platform. In: Proceedings of the second Nordic conference on Human-computer interaction, 2002. ACM, pp 71–80 (2002)
14. Albert, W., Tullis, T.: Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Newnes, London (2013)
15. Richey, R., Klein, J.: Design and Development Research: Methods, Strategies, and Issues. Routledge, Abingdon (2014)
16. PowerMapper (2016) SortSite Online. http://www.powermapper.com/products/sortsite/. 2016
17. Lazar, J., Allen, A., Kleinman, J., Malarkey, C.: What frustrates screen reader users on the web: a study of 100 blind users. Int. J. Human-computer Interact. **22**(3), 247–269 (2007)
18. Di Blas, N., Paolini, P., Speroni, M.: Usable accessibility to the Web for blind users. In: Proceedings of 8th ERCIM Workshop: User Interfaces for All, Vienna, 2004. Citeseer (2004)
19. Power, C., Freire, A., Petrie, H., Swallow, D.: Guidelines are only half of the story: accessibility problems encountered by blind users on the web. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2012. ACM, pp 433–442 (2012)
20. Takagi, H., Asakawa, C., Fukuda, K., Maeda, J.: Accessibility designer: visualizing usability for the blind. In: ACM SIGACCESS Accessibility and Computing, 2004. vol 77–78. ACM, pp 177–184 (2004)
21. Mankoff, J., Fait, H., Tran, T.: Is your web page accessible?: a comparative study of methods for assessing web page accessibility for the blind. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2005. ACM, pp 41–50 (2005)
22. Gonçalves, R., Martins, J., Branco, F., Pereira, J., Peixoto, C., Rocha, T.: AccessWeb Barometer: A Web Accessibility Evaluation and Analysis Platform. In: INTERNET 2015, The Seventh International Conference on Evolving Internet, 2015. IARIA, pp 18–25
23. Nielsen, J.: Heuristic evaluation. Usability inspection methods **17**(1), 25–62 (1994)
24. Figueiredo, B.: Web design: estrutura, concepção e produção de sites web. FCA—Editora de Informática, Lisboa (2004)
25. Krug, S.: Não me faça pensar!: uma abordagem de bom senso à usabilidade na web. Alta Books, Rio de Janeiro (2008)
26. Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T.: Human-Computer Interaction. Addison-Wesley Longman Ltd., London (1994)
27. Nielsen, J.: Severity ratings for usability problems. Papers and Essays 54 (1995)
28. Yin, R.K.: Case Study Research: Design and Methods. SAGE Publications, London (2014)