

Usability Evaluation of a Web System for Spatially Oriented Audio Descriptions of Images Addressed to Visually Impaired People

José Monserrat Neto, André P. Freire, Sabrina S. Souto, and Ramon S. Abílio

Universidade Federal de Lavras (UFLA) - Departamento de Ciência da Computação
Caixa Postal 3037 – 37200-000 – Lavras – MG – Brasil
{monserrat, apfreire}@dcc.ufla.br, sabrina@comp.ufla.br,
ramon@posgrad.ufla.br

Abstract. This paper describes a web system designed to provide spatially oriented audio descriptions of an image for visually impaired users. The system uses a hardware-independent platform of the technique of multimodal presentation of images. Visually impaired users interact with an image displayed on the screen while moving the cursor – with a mouse or a tablet (pen or finger touch) – and listening to the audio description of previously marked areas within the image. The paper also describes the usability evaluation performed with five participants and its main results. Generally, the five participants accomplished the usability test tasks and could better understand the image displayed. The paper also describes the main findings and discusses some implications for design, suggesting some improvements.

Keywords: Spatial orientation, audio descriptions.

1 Introduction

There are 285 million visually impaired people worldwide of which 39 million are blind [1]. In Brazil, there are 6.5 million visually impaired, of which 582 thousand are blind [2]. In the last 20 years screen reader software have been increasingly improved and become more accessible, rendering broader educational and social inclusion.

However, visually impaired people face serious problems with images when using screen readers. Generally, they cannot “read” digital images such as gif, jpg or png files of a web page, as often it lacks alternative text. And even when this is available, it is usually insufficient for visually impaired users to comprehend the image [3].

Therefore, we need an alternative that allows visually impaired people interact with an image and understand what is represented in it. This paper presents a proposal of software to address this issue, along with this alternative and, in order to test it, also a task-based usability test with five participants that aimed at verifying the usefulness and validity of the implemented solution.

Initially in Section 2, the educational, scientific and technological context of the image information problem for visually impaired people is briefly discussed. It is then

presented a simpler and cheaper solution for the technique of multimodal presentation of images, implemented with web technologies. Using such solution, in Section 3 a prototype of a web system called AudioImagem for spatially oriented audio descriptions of images addressed to visually impaired people is described. Next, in Section 4 the usability evaluation of the prototype in order to verify its usefulness and validity is presented, seeking to make a proof-of-concept of the technology implemented; and in Section 5 results of evaluation are presented. At last, in Section 6 the findings and implications for design are presented, and in Section 7 the conclusion and future works.

2 Educational, Scientific and Technological Context

For didactic-pedagogic purposes, a student is regarded as visually impaired when he or she cannot grasp regular classes using traditional materials and teaching methods which require visual skills.

To overcome this situation, many visually impaired people use *screen reader* software that works through electronic voice. Through keyboard (or eventually with a mouse) users can select a text of a file, screen element, or web page, hence resulting in an electronic voice audio that reads the text by means of the computer loud speaker, or an ear headset. Visually impaired users can listen to the texts of files, screen selected items and web pages. There are several screen readers on the market, such as JAWS, Virtual Vision, DOSVOX, Orca, WindowEyes, among others.

However, the presence of inaccessible images still emerges as a great challenge for screen readers. Any digital image such as jpg, gif or png file does not keep in itself information in text format about the image, to be read by a screen reader. Usually, images such as photos, drawings, charts, diagrams and so on are in the middle of texts about any subject matter, and their information are often essential for a clear understanding of those texts. In education in general but particularly in higher education, that fact is sheer apparent. Any student learns the several concepts of sciences with numerous figures and diagrams such as those of an eukaryotic cell, a block on inclined plane, a helium atom and so on. As screen reader is unable to “read images” visually impaired people are usually excluded from full understanding of usual digital didactic texts.

Fortunately in web pages it is possible to write an 'alt text' in place of an image in case this cannot be loaded or accessed for any technical reason, or cannot be seen in the case of visually impaired users.

There are two problems with using alternative text in web pages. Firstly, alternative text are seldom properly placed and written to describe the image. Strictly few web pages follow the recommendations of the Web Content Accessibility Guidelines 2.0 [4], which propose to always place alternative text for visually impaired users.

Secondly, even when there is such a text description of the image, it is usually insufficient for the visually impaired person to understand the image [3]. It is difficult to write an image description that can cover various types of information that a visually impaired user could ever possibly need to understand a certain image in a particular

context of a web page. Besides, the text description is insufficient to make users understanding clearly the spatial arrangement of elements in the image.

An alternative to face the problem of image understanding by visually impaired people is the use of technology with tactile interaction [5]. The basic idea is to develop a screen-like device capable of creating shapes in high relief to be felt by the user's fingers with visual disability. For instance, in the contest winner work of the Mobile Design Competition 2012, organized by the LG Company, proposed a device that has a "membrane of a touch sensitive polymer that changes constantly to give user a tactile feedback according to what is displayed on the screen" [6]. Another similar alternative, though older, is to use micro-pins on the screen device surface which dynamically creates the relief of a figure and can be felt by a visually impaired user's touch. Such alternatives are very promising, however involve developing specific hardware, which ends rendering these solutions more expensive.

Power & Jürgensen [5] analyzed several options available to render information available to people with visual disabilities, and conducted an extensive review of presentation tools and techniques of textual documents and graphics for visually impaired people, using audio and tactile modalities.

One trend is the combination of techniques based on touching with audio feedback in tactile interaction with a special haptic device developed for such, being this technique called "multimodal presentation of charts" [5, p107]. The technology presented in this paper (AudioImagem) follows that technique. In this case, unlike the special screen devices with pins or sensitive polymer membranes, the technique simply uses the positional information of the finger or a pen on a tablet to activate the audio feedback. Many devices were created with this technique [7, 8, 9, 10].

In common, all these devices are based on specific hardware system, that is, each system has its own tablet device. For instance, Touch Graphics Company [11] sells a product called "Talking Tactile Tablet 2" which allows a visually impaired user to interact by touch with an image displayed on the tablet and listen audibly the information about the image displayed.

Again, such tablets have specific hardware and are generally more expensive. Furthermore, the material created to be accessible by people with visual disabilities can only be created and accessed through the system itself, generating dependence on the company that produces the system, and making it difficult to share the material created via internet, for instance.

A different solution would be a combination of technique of the audio feedback with that of the positional interaction on the image by touch of the visually impaired user. And fortunately, unlike the alternatives above, this solution does not need to employ any specific hardware, making it a platform independent alternative. User interaction on the image may be accomplished with a mouse, tablet (that one used for drawing), hand held tablet for general use, such as Ipad and Android tablets, touch-screen monitor, and any interface that is able to control the cursor movement (the little arrow) on the screen monitor.

Thus, a simple and cheap but no less promising solution is to implement the multimodal presentation technique with web technologies, generating web pages with images that contain audio information about themselves which may be accessible to impaired visual people in a broader way through the internet using a regular web browser.

3 Description of AudioImagem

Based on the works of Freitas [12] and De Sousa [13], the Polaris Company [14] has developed a software prototype of a web technology – called AudioImagem – through which a sighted person is able to delimit areas within an image and associate audio descriptions to them. The system can then yield spatial references of the marked areas on the image and describe them audibly whenever a visually impaired user “walks over” the image using a mouse or a tablet.

There are two modes for the visually impaired user to interact with the image and listen to its audio descriptions: the static mode and the navigational mode. The first mode does not depend on the cursor position, that is, does not depend on the interaction of visually impaired users with the image.

There are two static audio descriptions: the short and the long one, which are activated by the keys 'C' and 'L' in the keyboard, respectively. The short description is envisaged to describe briefly the image; and the long description to describe it with more details. The short one is also envisaged to describe the kind of image, like photo, drawing, diagram, graphic, table, chart, formula, etc. The long description may describe the image elements with more details, specially their disposition on the image providing the first spatial orientations to visually impaired users before navigating on the image.

The second mode of the user interaction with the image is the navigational one. Visually impaired users may interact with the image using a mouse or tablet (pen or finger touch), so that they may “walk over” the image. When the cursor is over an image area – which was previously marked and described – the user immediately listens to its audio description. So the user may listen to all demarcated and described areas of the image, when “walking over” and “exploring” the image.

However this way of the navigational mode over the image may be not enough for the user to be spatially oriented on the image in that she or he may leave the image inadvertently and stay out of it. That is, the cursor may go out or be out of the figure and the user may not find it and come back to the image easily. To face this problem the navigational mode also adopted two solutions: (a) at the moment the cursor goes out of the image, it provides an audio message describing to which side it left the image (1 to left, 2 to above, 3 to right or 4 to below); and (b) it provides the 'N' key that activates an audio message of the cursor position both outside the image and within it, as shown in Figure 1.

When the 'N' key is pressed in the computer keyboard and the cursor is out of the image the system yields an audio description that says in which of eight positions the cursor is placed at the moment (1 above left, 2 above, 3 above right, 4 left, 5 right, 6 below left, 7 below, 8 below right).

And when the key 'N' is pressed and the cursor is within the image the system provides an audio description that speaks in which of nine positions the cursor is placed within the image (1 above left, 2 center above, 3 above right, 4 center left, 5 center, 6 center right, 7 below left, 8 center below, 9 below right).

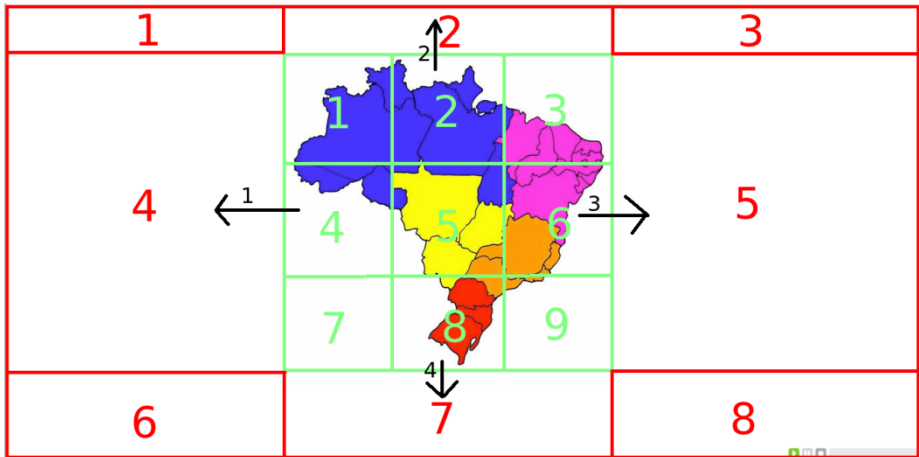


Fig. 1. Audio feedback of the cursor position in the navigational mode

To sum up, the navigational mode depends on the cursor position and yields audibly both its position on the screen – out and within the image – and the description of a marked area of the image. It also yields to which side the cursor moved when leaving the image.

Yet, it is not well known what the actual usefulness of such technology would be for visually impaired users. Among other questions, we investigate whether such a model of interaction between the visually impaired user and the image is enough for her or him to understand the image. Thus, it is necessary to carry out a proof-of-concept test of the “spatial audio description” technology by means of a usability test of the prototype developed.

4 Usability Test

The usability test comprised a task-based user evaluation by participants and employed a think-aloud protocol while the users performed their tasks. Audio and video recording were taken.

The test set consisted of a desktop computer with: a) a tablet controlled by pen and finger touch; b) two webcams, one addressed to the participant’s face and another to the keyboard and tablet; c) a microphone embedded in one of the webcams, that was over the keyboard and tablet; d) a loud speaker set; e) Linux operation system; f) Guvview to capture video in two windows in the screen, for each webcam (user face and keyboard-tablet) to be shown on the screen during test; and (e) SimpleScreenRecorder for recording the screen and the audio. The Figure 2 exhibits a test set photo.

Five visually impaired persons have participated in the usability evaluation which was composed of twelve tasks. Each task was elaborated as a web page presenting the image information audibly described and a specific question associated to the image, which could only be answered by the user by “wandering” on the image and listening to its spatial audio descriptions.

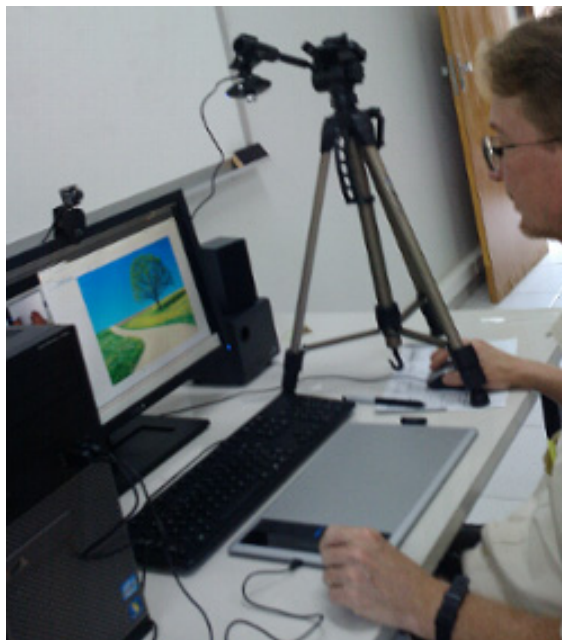


Fig. 2. Photo of the test set showing the first task

The log of the tasks was accomplished through three tools: a) a screen recorder that recorded cursor movements, made by the user on the system during the tasks; b) a voice recorder which recorded all user comments and the system audio descriptions; and c) two webcams which recorded user movements over the keyboard and tablet and the user face during task performance.

In each task the system presented two descriptions of the respective image, a short and a long one, explaining statically the image for the visually impaired user. She or he could always listen to an audible help about the navigation commands on the page and on the image. Besides the short and long descriptions, the system provided also the audio description of the task to be accomplished by the user. So, the keys 'C', 'L' and 'T', when pressed, activated the three audio descriptions, respectively short, long and task one. For instance, in the second task (Figure 3) its image was presented audibly with the following descriptions:

- Short description: “Photo of an autumn landscape”.
- Long description: “This is a photo of a bleak autumn day landscape. There is a street crossing the photo on the left and a leafy tree with orange leaves on the right side”.
- Task: “Observe the photo and describe as you understand it”.

Various types of image were selected for each task such as photo, plan drawing, graph, flowchart, diagram and table.



Fig. 3. Landscape photo used in the second task of the usability test

Briefly, tasks 1 and 2 displayed landscape photos, task 3 a drawing of the world globe, task 4 a geographical map of Brazil, task 5 an ecosystem flowchart, task 6 a trigonometric table, task 7 an algorithm flowchart, task 8 a drawing of the kinds of cow meat, task 9 a phase graph diagram of chemical substance, task 10 a plan drawing of chessboard with all pieces, task 11 a diagram of Daniel chemical cell, and task 12 a figure of the periodic chemical table.

The objective of performing several tasks with so many images was to test different types of images usually employed in the sciences that may be potentially useful in education and described audibly, listened to and comprehended by visually impaired students.

Figure 4 shows the graph image used in the ninth task, that was to find out the triple and critical points.

Each task was designed to be accomplished in between 5 and 10 minutes. The user could choose to employ the tablet with pen or finger touch that was configured to function in absolute mode. After some time becoming familiarized with the system, the five users attempted the tasks taking in average 2 hours and 20 minutes in total.

The usability test script followed by participants was: 1) reading the presentation of the usability test and explaining the basics of AudioImagem system navigation; 2) starting screen and audio recording; 3) answering the pretest questionnaire; 4) familiarizing with the system in doing the first task; 5) choosing between pen or finger touch on tablet; 6) performing the other 11 tasks; 7) answering the task questionnaire after each task; and 8) answering the posttest questionnaire.

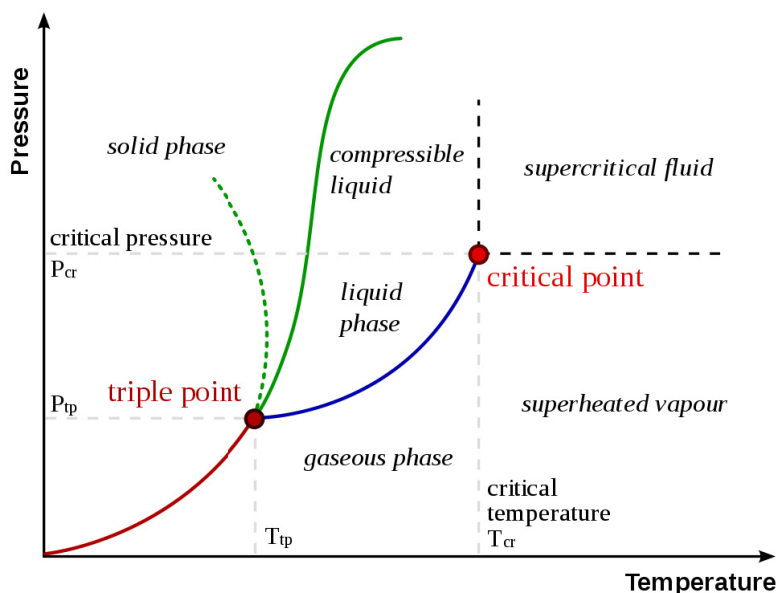


Fig. 4. Image of a phase graph of a chemical substance

The pretest questionnaire asked information about the participant, such as name, age, gender, education level, motivation to participate in the test, kind of visual disability, origin of disability, proficiency in reading Braille, other disabilities, and kinds of assistive technology used – type, model, version and use time of each technology.

The task questionnaire asked the final answer to the respective task, the degree of difficulty of each task done (very easy, easy, more or less, difficult, very difficult) and its reason, the system helpfulness to perform the task (essential, helpful, more or less, little helpful, not at all helpful) and its reason, and the degree of trust in the answer given.

The post-test questionnaire presented questions about the difficulties faced by users, their suggestions to make the system more helpful for them, about what they liked most in using it, and also asked whether the system could be helpful in education, receiving the unanimous answer: yes. In participants' words: "Of course it is" and "For sure".

5 Results

A brief summary of some quantitative results of the usability test is presented in the Table 1. For example, participant P2 is blind and has experience using screen reader. In the study, this user concluded 5 tasks, partially concluded 1 task, and did not succeed to conclude 4 tasks, and took 1 hour and 58 minutes using the finger (F) and pen (P) on the tablet. This user first chose to use finger touch, but when faced problem in finding out smaller marked areas in the image the user changed to pen. The same

happened with participant P3 who is also blind. It reveals that the pen gives more precision in navigating on the image. It may also happen that the use of pen allows her or him to put the fingers on the tablet while moving the pen over the tablet, since when pen mode is activated the finger touch is turned off. Otherwise, when the finger touch is activated the user can only use one finger to move on the tablet and so the arm must be kept suspended in order to avoid touching with other fingers. After some time this causes fatigue and can become uncomfortable.

Table 1. Summary of some quantitative results

Users	Kind of disability	Screen reader experience	Tasks concluded	Tasks part. concluded	Tasks not concluded	Time test (hour)	Tablet (Pen or Finger)
P1	Low vision	Yes	11	0	1	1:59	Finger
P2	Blind	Yes	5	1	4	1:58	F and P
P3	Blind	Yes	4	4	2	2:43	F and P
P4	Low vision	Yes	9	0	2	2:21	Pen
P5	Low vision	No	5	0	7	2:49	Pen

In the fourth task, for instance, it was displayed a map of Brazil and users were asked about the number of states that compose the northeast region of Brazil. This region has nine states and many of them are small areas in the map. So all users found very difficult to navigate on that image and figure out where all states are.

6 Findings and Implications for Design

The usability evaluation has brought important findings and implications for design which are discussed around four themes: 1) the interface for interacting with the image; 2) audio descriptions and previous knowledge; 3) the borders or limits of delimited areas within an image; and 4) modes and strategies of navigation on the image.

Although five users are indeed a very small sample of visually impaired people, clearly the two blind users had more difficulty in navigating on the task images. Any visual hint such as shadows, blurry spots, colored blots on the image displayed in the screen ends helping low vision user to navigate on the image. Blind users cannot count on such hints, only on the audio feedbacks while wandering out spatially on the image. For this reason the two blind users did not accomplished the twelve tasks designed avoiding getting too tired.

6.1 Mouse, Tablet and Other Interface Devices

In the usability test, the mouse was regarded as inappropriate to visually impaired users since it provides less control over the cursor localization on the image. This conclusion was based on preliminary tests. However, this is not always true. Participant P1 showed that low vision users may opt to use a mouse. Specifically, this user had great difficulty in using the finger on the tablet in absolute mode, and preferred

finger touch in relative mode since the participant was used to the notebook mousepad, which is by default always in relative mode. This user said that colored blots helped very much to find out the small areas on the image as in the ninth task, displayed in Figure 4.

The comparison between pen and finger touch is not conclusive yet. While touching with the finger seems to provide more freedom to navigate on the tablet, at the same time it is not as precise as the pen, especially on small image areas. It also seems to be somehow stressing since finger touch obliges the user to keep her or his arm suspended over the tablet, since absolute mode is the only way to use finger touch on tablet, as the cursor on the screen cannot be controlled by two fingers at the same time.

6.2 Audio Descriptions and Previous Knowledge

The usability test revealed that describing an image properly for visually impaired students is not as trivial as it seems to be at first. First, the long audio description is very important to orient users spatially before they try “walking over” the image. Some task images were not suitably described with the long audio description, such as in the autumn landscape photo of the second task, as shown in Figure 2.

It also revealed the need of knowing more precisely the previous knowledge of users about any given image. The deeper is the user’s knowledge, the less need there is to extend the description in the long audio description option, before the user navigates on the image. For instance, participant P4 knew the periodic chemical element table and remembered roughly the position of metals, semimetals, non metal and noble gases in the table. So this user had an advantage over the other users. Thus, for people who already know, even roughly, the elements of an image the long audio description must be different from those who are newbie in a given subject matter that is shown in the image.

6.3 Image Marked Areas and Their Borders

A problem of AudioImagem system readily identified by all users was the response speed of the audio feedback that is slightly slow and needs to be faster as users usually pass on an area at a higher speed and miss the audio description of the respective area. This was especially true in relation to small image areas.

Another problem was that as the user moved to a marked area and listened to its audio description, the user could not know where that marked area finished. Only when the user reached another marked area and listened to its associated audio description, could the user know that the previous marked area had finished. So they moved many times from one side to the other within a given marked area in order to know its extension within the image. There should be a way, perhaps an audio feedback or another key, to warn that the user is still in the same marked area.

A closed related problem was to find out the borders of the marked areas. One strategy of some users was to travel in zigzags between two areas to localize spatially the contours of the given marked areas.

This findings show clearly the need of a feedback of the border of marked areas. One suggestion is to create an “audio line” – an audibly feedback to reveal the contours of an area within the image for the visually impaired user. The same could be done regarding the contours of the whole image, as one participant suggested. So, instead of seeing the borders of the image and its marked areas, the user could listen to such contours.

6.4 Navigation Modes and Strategies

Two participants adopted the following strategy to scan the images while performing the tasks: they traveled with finger or pen in horizontal lines over the tablet, changing at each time only the height the user traversed the image, so that the user could avoid missing small areas and ensure they were finding out all marked areas of the image. The same strategy was also performed in vertical lines.

This finding reveals the need for a sort of directed navigation to a given delimited image area. The present navigational mode of audio description is a sort of free navigation on the image. The user must wander over the image to find out the marked areas. The only help is the long audio description option in the static mode of navigation. Another suggestion is to create another alternative to the navigational mode, a “directed navigation” that directs the user to a selected marked area. So the user could choose in a list of areas that one he or she wants to find on the image. The system could then direct the user to that area giving an audio feedback. The closer the user is to the area the louder is the audio feedback, and so the user could find smaller areas of the image more easily.

7 Conclusion

We discussed and described a platform independent alternative of the technique of multimodal presentation of images, implemented as a prototype of a web system called AudioImagem and designed for visually impaired people. Also we presented the usability test of the AudioImagem, performed with 5 participants, and its main results. The videos recorded have still much information to be more in depth examined. Generally the prototype fulfilled its purpose to allow visually impaired people to interact with an image and understand what is displayed in it, as revealed by the number of tasks concluded, 70% in average. Nevertheless it revealed some weaknesses, such as the speed of audio feedback and the problem of border in marked areas. However it also suggested some solutions to them, such as the “audio line” to allow visually impaired users to listen the contours of the area of an image, and the “directed navigation” to direct users to an area of the image. Future works include the improvements of AudioImagem, creation of didactic material using audio described images, and new usability tests with this assistive technology and the didactic materials created.

Acknowledgements. We thank all the participants for their most valuable contributions to this work. We also thank CNPq for the financial support to this research via Grant 458795/2013-5.

References

1. WHO – World Health Organization, In: Visual Impairment and Blindness, Fact Sheet No 282 (June 2012)
2. IBGE, In: IBGE Census, General characteristics of the population, religion and people with disabilities (2010)
3. Petrie, H., Harrison, C., Dev, S.: Describing images on the web: a survey of current practice and prospects for the future. In: 3rd International Conference on Universal Access in Human-Computer Interaction, part of HCI International 2005 (2005)
4. Caldwell, B., Cooper, M., Guarino Reid, L., Vanderheiden, G.: “Web Content Accessibility Guidelines 2.0”, In: W3C Recommendation. (2013), <http://www.w3.org/TR/WCAG20> (accessed in November 2013)
5. Power, Jürgensen: “Accessible presentation of information for people with visual disabilities”. Universal Access in the Information Society 9.2, 97–119 (2009, 2010)
6. Techtudo: “Smartphone conceitual premiado em concurso da LG investe em telas táteis” In: Techtudo Curiosidades, reportagem de dezembro de 2012 (2012), <http://www.techtudo.com.br/curiosidades/noticia/2012/12/smartphone-conceitual-premiado-em-concurso-da-lg-investe-em-telas-tateis.html> (accessed in October 2013)
7. Poh, S.P.: Talking diagrams., Master’s Thesis, The University of Western Ontario (1995)
8. Landau, S.: “Tactile graphics: strategies for non-visual seeing”, In: Thresholds. MIT School of Architecture (1999)
9. Landau, S., Gourgey, K.: Development of a talking tactile tablet. Information Technology Disability 7(2) (2001), <http://www.rit.edu/~easi/itd/itdv07.htm>
10. Xu, H.: “A Support System for Graphics for Visually Impaired People”, Master’s Thesis in Computer Science, The University of Western Ontario, Canada (2013)
11. Touch Graphics Inc., “Talking Tactile Tablet 2”, tablet para pessoas com deficiência visual (2013), <http://touchgraphics.com/OnlineStore/index.php/featured-products/talking-tactile-tablet-2-ttt.html> (accessed in November 2013)
12. Freitas, A.C.F.: Study and proposal of an assistive technology of audio description of images for visually impaired persons. Undergraduate final year dissertation, Amanda Cristina Ferreira de Freitas, Computer Science, UFLA (2010)
13. Da Silva, R.C.: Study and proposal of a spatial navigability model for a software of audio description of images addressed to visually impaired people. Undergraduate final year dissertation, Renato Corcova da Silva, Computer Science, UFLA (2011)
14. Polaris, Polaris Inovações em Soluções Web (2011), <http://www.polarisweb.com.br/>