A Scheme for Multimodal Component Recommendation

Natacsha Ordones Raposo, Thais Castro[(✉)], and Alberto Castro

GSI Research Group, Institute of Computing,
Federal University of Amazonas, Manaus, Brazil
{natacsha, thais, alberto}@icomp.ufam.edu.br

Abstract. Multimodal interaction has proved to be a promising way for developing more accessible applications, especially for those people with visual impairments, but it still presents many challenges like the use of touchscreen devices and dynamic resources embedded in web content. For HCI, multimodal interaction is characterized by human intervention using a combination of different sensory modalities as input and output channels. In this paper, it is presented a scheme for recommendation of multimodal components for developing accessible applications and it is discussed lessons learned from its use in the development and evaluation of applications using that scheme. This scheme has already been used to guide the development of two software tools, both aimed at improving accessibility for visually impaired people through multimodal interaction. Results have shown evidence that this scheme is fit for purpose and has given some pointers for a framework for multimodal accessible software development.

Keywords: Human-Computer Interaction · Multimodal interaction

1 Introduction

Human interaction with the world is intrinsically multimodal [1, 2]. There are two views on multimodal interaction: the first one focuses on the human side - perception and control [3] and in this context, the word modality refers to human intervention and input and output channels. The second view focuses on the use of two or more modalities for input and output on the computer to build systems that make synergistic use of input or output of these modalities in parallel. The input modalities of many computers and devices can be interpreted as corresponding to the human senses: cameras (sight), tactile sensors (touch) [4], microphones (hearing), olfactory (smell), and even taste [5].

Multimodal interaction systems aim to support the identification of naturally occurring forms of language and human behavior through the use of recognition-based technologies [6, 7]. Considering that humans can process information faster and better when this information is presented in several modalities [8], it has been promising the use of these systems aimed at increasing accessibility to people with disabilities. That is, the interfaces that promote such interaction must be developed in a flexible way in order to support human-computer communication, thus allowing users with different preferences and skill levels to choose how they will interact [9].

The current technological landscape allows digital access to be extended to previously excluded groups or categories of users. However, even with the resources and possibilities offered by the industry to people with disabilities, they still face barriers in the use of accessible software. Second, as the information technology industry focuses on making its offerings accessible to people with disabilities, it is becoming noticeable that moving towards the standards of affordability does not guarantee ease of use for those people with disabilities. This occurs mainly when accessibility criteria are not used during their development or even when these interaction criteria are designed in developer's own way of interaction, that is, particular interaction issues of the individuals who are the target audience for the applications are neglected. According to [11], most developers are not aware of the difficulties encountered by users with disabilities.

Contextualizing this development of multimodal interaction scheme scenario to people with visual impairment, it is evident the importance in relating the specific interaction characteristics to this group of people in order for the system to have the criterion of accessibility in the quality of use that is related to the barriers that prevent visually impaired people from interacting with the system.

This paper is structured in five sections. Next, it is presented some concepts necessary for the proposal. Section 3 presents the scheme for multimodal component recommendation, including the knowledge acquired by designing and evaluating two applications. In Sect. 4 there is a brief discussion about the first steps towards the scheme's validation followed by the conclusions (Sect. 5).

2   Background

Multimodal interaction is a trend topic in interaction design, especially for people with disabilities, because of the possibilities for adapting and customizing content for different devices and resources depending on the abilities each user has. The Web is a natural resource of information, but many websites are not prepared to handle different components for adapting to people's different abilities and for adapting websites or any application for them, one way to act upon it is designing multimodal interfaces. For that reason, in the following subsections, it is presented some usual concepts about web accessibility and multimodal interfaces for people with disabilities.

2.1   Web Accessibility

The term web accessibility is used to emphasise that people with disabilities can access information though interacting on webpages. More specifically, people with disabilities can perceive, understand, navigate, Interact and contribute for the web. Web accessibility also is beneficial for other people as the elderly with reducing capabilities, resulted for the aging process [3].

Considering the principles for web accessibility, W3C created, in 1999, the WAI (Web Accessibility Initiative), formed by work groups dedicated to the elaboration of guidelines, called WCAG - Web Content Accessibility Guidelines (WCAG). Within

the WCAG, there are many guidelines sets for fostering accessibility. The most recent one is WCAG 2.0.

On the developers' side, the accessibility of web content using mobile devices has becoming a more relevant subject mainly because of the increasing number of people using mobile devices for navigating on the web. In navigating, people are transferring most tasks they used to do on computers to their mobile devices, as bank transactions, calendars and receiving and sending e-mails. With the growth in the used of mobile devices for multiple tasks, such devices are becoming more important for people with disabilities and they are using those devices for enhancing their interaction and communication [13]. Regarding accessibility on mobile devices, there are specific guidelines recommended for the applications development for Android, Accessibility Developer Checklist [4] and for iOS, Accessibility Programming Guide for iOS [5].

As aforementioned, there are some guidelines for guiding the development of accessible web content. Even though such guidelines are widely used, they are not enough for gradually supporting each activity involved in the development process of accessible web content, mostly when referring to the development process for a group, such as the group of people with visual impairments, because this development requires the knowledge necessary to understand specific contingencies of the specific group and the existent guidelines are generic.

The challenge posed is to foster web accessibility for each specific group of people with disabilities. For the specific group of people with visual impairments, web content must be accessible for screen readers, what will be discussed later, with the two example tools developed in the context of the proposed scheme.

2.2 Multimodal Interfaces for People with Disabilities

People with disabilities need different input and sometimes output ways to appropriately interact in web sites and computer systems. Multimodal interaction is a concept in Human-Computer Interaction, generally meaning the interaction with virtual and physical environments through more natural modes of communication. The interfaces developed within this concept, have been gradually acquired the capacity of understanding, interpreting and generating specific data in response to the analysis content, differing from the classical applications and multimedia systems which do not take in data semantic (sound, image, video) that they manipulate [6–8]. Such interfaces have becoming promising in software development, because they allow the integration of different modalities for user interaction.

Regarding development purposes, multimodal interfaces are directly associated to different investigation directions [8]:

- developing more natural interactive interfaces, intuitive, efficient and, simultaneously, less awkward, associated to a crescent learning curve [9–11].
- increasing the amount of information transmitted in useful time during an interaction, resulting in a run time decrease [7, 12, 13].

- increasing system robustness, aiming at obtaining a superior intelligibility for recognizing information through signs received from different modalities to solve ambiguities, communication error prevention or fixing [14–16].
- stimulating the commitment of the user in the activity to be developed, promoting his satisfaction [17, 18].
- promoting user's understanding and anticipating his intentions [19, 20].
- allowing greater flexibility in accessibility using different contexts, dissociated from the user, that reveals usability constraints for certain modalities, through the possibility of selecting the most adapted modal channels that suits user's needs, his proficiency level and/or the nature of the task to be performed [21, 22].
- increasing accessibility to computers by people with specific disabilities (whether sensory or motor), by providing them with alternative modalities and styles of multimodality [23, 24].
- providing new ways of computing, previously unavailable [25, 26].
- providing input channels utterance, to prevent cognitive or physical overload for long interaction [21, 27].
- decreasing the cognitive overload associated to a task and, consequently, the attention level necessary to perform it [28, 29].
- fostering information systems adaptation for predominant interaction patterns for each user [30, 31].

Although the advantages of multimodal interfaces are evident, the development of multimodal projects is still a challenge [27], due to the lack of tools that appropriately guide the designer in the design, implementation and evaluation of multimodal interfaces. In addition, there is a need to process heterogeneous user group inputs and integrate multiple output/input modes that can operate in parallel or simultaneously, together with handling recognition and synchronization errors for the generation of efficient multimodal interfaces (apud [32]). The approach of this article extends these conceptions adapting them to the development of each specific application regarding their purpose.

3 Proposal

Considering the context, in this work we present a scheme for recommendation of multimodal components suitable for software development of accessible applications. Multimodal components, in this work, are an independent piece of software with the same functionality configured for all devices and medias available [33].

The proposed scheme, represented in Fig. 1, has a multilayered knowledge base (the block at left) comprising: (i) "theoretical mapping" – a set of heuristics extracted from related literature (recommendations from regulatory associations, scientific studies, multimodal fusion techniques, etc.); (ii) "recommendation history" – structured records of previous recommendations from experimental scenarios; (iii) a multimodal "applications repository" and correspondent "recommendations criteria" – organized according to aspects considered (e.g. physical and cognitive ones).
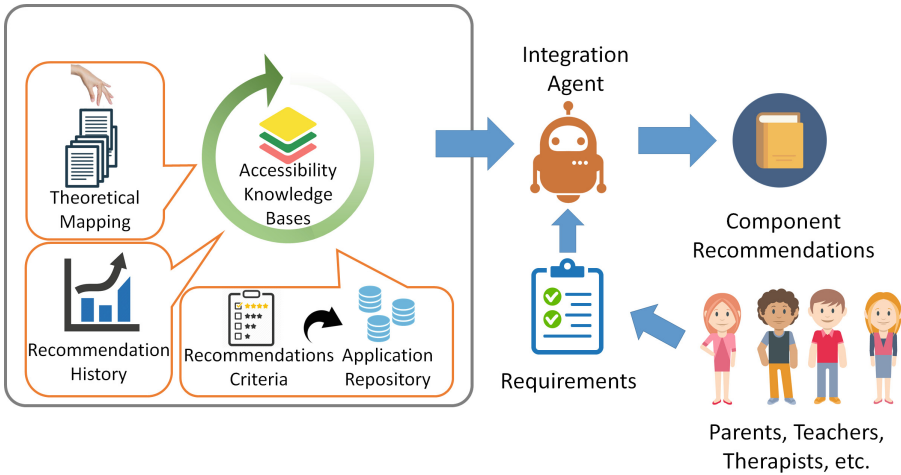
Fig. 1. A scheme for multimodal component recommendation

The resulting body of knowledge, along with a set of requirements elicited from people with disabilities, parents, teachers, therapists and other stakeholders with expertise on the idiosyncrasies of a certain situation, are used by an Integration Agent to generate components recommendations (resources, strategies, user profiles, etc.) recommended to tackle accessibility issues for that situation. In its turn, each situation and its specificities becomes a new case that is stored and retrieved in quite similar way to case-based reasoning.

The integration agent, as defined in [35] is responsible for integrating information from different input devices and/or available media, resulting on a set of recommendations grouped by components. The proposed scheme is a detailed description of components definition on [33, 34], incorporating the artificial intelligence techniques for managing knowledge bases and application repository. The proposed scheme is still on test. The multilayer knowledge block already has two applications for people with visual impairments, using different modalities from mobile phones and desktops or notebooks connected to the internet. From those applications, data about interaction and conformance tests have already been collected.

3.1   Application Repository

The developed applications are the key point for this scheme definition. To develop such applications, we followed recommendations criteria specific to their purpose. The first application developed followed Android's accessibility recommendation criteria and the second application, WABlind, followed the recommendations of WCAG 2.0. The first (RotaColab) is a location-based collaborative tool aimed to support blind people navigating through unconventional paths using mobile devices. The second one (WABlind) is a tool allowing conventional webpages to be locally rebuilt and reorganized in such way they can be read by screen readers. Both software has been tested

in real world conditions or in laboratory settings to assess compliance to usual standards (e.g. WCAG).

RotaColab was developed in the Android platform. This application demonstrates techniques of the use of accessibility features applied to the navigational orientation, for analyzing its contributions related to its communicability with visually impaired people. To develop RotaColab [33], recommendation criteria were adopted for the development of accessible applications of the Android - Accessibility Developer Checklist [4]. These recommendations were related and exemplified with specific interaction characteristics of the application, considering the specificities and limitations of the target audience, the visually impaired. The recommendation criteria were: text field hints, enable focus-based navigation, no audio-only feedback, temporary or self-hiding controls and notifications, controls that change function, supplemental accessibility audio feedback, decorative images and graphics.

Another tool developed for integrating the repository is WABlind (Web Accessible for Blind People) which is a tool for restructuring web content designed for people with visual impairments. Its use is aided by screen readers and voice synthesizers, commonly used by this group of people, on computers and mobile devices. This tool relates WCAG 2.0 guidelines with previous results obtained on how best to present web content to blind people, according to research conducted by the W3C [34].

The tool in its initial execution enables the treatment of 4 of the 15 most problematic interaction items found when browsing web pages, as reported in a recent survey conducted in 2013 by the W3C on the use of Assistive Technologies, Amplifiers and screen readers [34].

The 4 selected items, which are defined as the recommendation criteria, are related to the syntactic property of web pages. That is, at this level, any item related to the semantics of the page was excluded. The following are the listed research items associated with the WCAG 2.0 guidelines that run at this level:

- The presence of inaccessible Flash content - Guideline 1.1.1 Non-textual content (Level A).
- CAPTCHA - Use text image to verify that you are a human user - Guideline 3.3.2 Labels or Instructions (Level A).
- Images with missing or inappropriate description (alt text) - Guideline 1.1.1 Non-textual content (Level A).
- Unidentified foreign language texts - Guideline 3.1.1 Page language (Level A).

The recommendation listed cover the WCAG 2.0 level A guidelines. WCAG 2.0 defines levels for each guideline to provide testable success criteria to allow such guidelines to be used where requirements and compliance tests are required. Three levels of compliance are defined: A (the lowest), AA and AAA (the highest) [35].

The differential of this tool is the set of recommendations applied to the elements, originally inaccessible, of the informed web page, which allows the provision of textual alternatives so that screen readers can identify them while browsing a web page.

Initial Assessment of WABlind

In order to show tool's feasibility, compliance levels tests were carried out with the pages restructured by WABlind with the validator AChecker [36]. Such validator is recommended by W3C [37].

To carry out the initial tests in verifying how much a Webpage improved after applying WABlind, we employed the free services of Alexa[1], which is a Web application that lists the most accessed web sites. We selected the most accessed web sites to verify, how many of them did not meet the needs (in terms of W3C rules not being followed) of visual impaired users. As a result, we gathered information on 50 websites across 5 categories: Education, Health, Entertainment, News and Shopping. In addition to that, we followed the same procedures after applying WABlind and dealing with the accessibility automatically identified problems. Table 1 shows the descriptive statistics of the results before and after WABlind considering all websites. Figure 2 shows the distribution of the results for this initial test without considering outliers (there were cases in which the maximum number of identified problems was 964) in order to facilitate visualization.

Table 1. Descriptive statistics for the results of applying wablind in the most accessed 50 Websites for each of the evaluated categories: Education, Health, Entertainment, News and Shopping.

Category	Measure	Mean	Median	Standard deviation	Minimum	Maximum
Education	Original	13,7	6,0	14,4	0,0	137
	With WABlind	9,7	5,5	9,2	0,0	70
Health	Original	24,4	13,0	22,4	0,0	132
	With WABlind	21,9	12,5	19,9	0,0	132
Entertainment	Original	51,4	31,0	43,0	0,0	379
	With WABlind	39,4	14,5	40,4	0,0	379
News	Original	63,7	31,5	64,6	0,0	848
	With WABlind	28,5	13,5	28,2	0,0	226
Shopping	Original	42,7	7,0	57,1	0,0	964
	With WABlind	44,0	8,5	57,0	0,0	964

Even with the increase of the number of identified problems of AA and AAA types (considering that after fixing A type problems, AA and AAA problems can arise), we can see that in some categories, the number of problems was reduced (i.e. Health, Entertainment and News). These results, suggest that WABlind can increase the quality of the evaluated web site in terms of accessibility. However, we still need to consider the causes for the increase in the number of problems in the Education and Shopping categories. We are investigating to what extent the A type problems considered by WABlind were corrected and which AA and AAA type problems arose.

---

[1] http://www.alexa.com/.
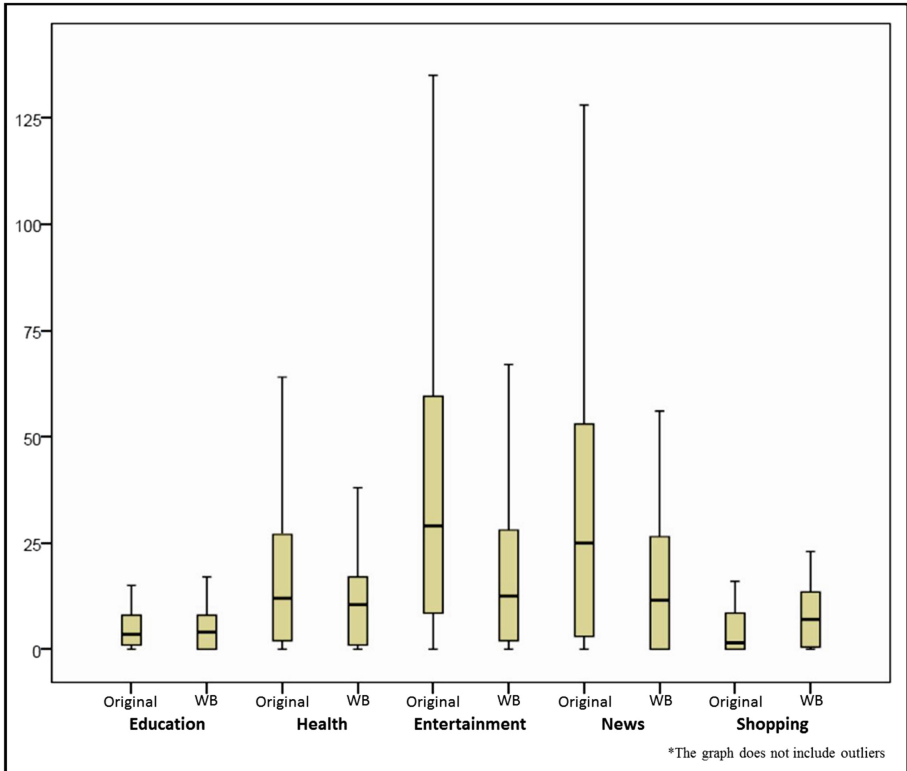
Fig. 2. Boxplot graphs demonstrating the distribution of identified problems both in the original website categories and the categories after using Wablind.

4   Discussion

As the WABlind tool was developed after RotaColab, some information, as well as user behaviors obtained during RotaColab tests were taken into account, such as the layout of elements in the web page in an extended way and the maximization of the voice interaction, since they were perceived difficulties regarding the provision of information and guidance in reading the RotaColab application page during the tests.

This was partially achieved because users are not familiar with touchscreen devices. Therefore, WABlind's development was reflected on the layout of elements in a web page and it was considered to label the elements arranged, both by means of automatic treatments and collaboratively in order to minimize obstacles encountered during navigation, specifically on mobile devices. This last phase of collaboration is still being tested.

For the conformity tests, pages with accessibility problems, in several categories, were collected, according to guidelines of WCAG 2.0. The idea was to find some clues of improvements after using WABLIND. The result of this study suggests that the use of this tool promotes accessibility by attending more conformance elements than without using WABlind, especially in web page elements that do not have auditory feedback.

5   Conclusions

This paper aimed at discussing about the challenge of increasing accessibility for people with visual impairments using multimodal resources as a support for interaction. The approach presented here starts with the development with two different multimodal tools for people with visual impairments, contextualized in a scheme to help developers to design multimodal interface projects for the existent and inaccessible tools based on the knowledge accumulated on the scheme.

The main challenge for the development of multimodal interfaces is the development of interfaces that meet accessibility criteria and services that abstract the needs of the user and not the specific characteristics of the device itself. Investigative studies on the interaction for visually impaired people in mobile devices is a latent demand that requires more and more studies and applied research. In addition to the increasing development of accessible mobile applications, tools for desktop are still used and a project to turn them accessible is necessary.

While standards and guidelines for the development of multimodal interfaces projects are not yet fully defined, individual initiatives such as the research paper reported here attempt to address the lack of guidelines for the development of such applications. As a result, a scheme, not fully consolidated as the proposed one, shows great feasibility from the perspective of development of applications following recommendations criteria appropriated to the development of projects of multimodal interfaces for people with visual impairments.

References

1. Bunt, H., Beun, R.-J., Borghuis, T.: Multimodal Human-Computer Communication Systems, Techniques, and Experiments. LNCS, vol. 1374. Springer, Heidelberg (1998)
2. Quek, F., et al.: A multimedia database system for temporally situated perceptual. Multimedia Tools Appl. **18**, 91–113 (2002)
3. W3C. Acessibilidade para o WAI. **W3C** (2005). http://www.w3.org/WAI/intro/accessibility.php. Accessed 20 Sept 2014
4. ANDROID. Accessibility Developer Checklist (2014). http://developer.android.com/guide/topics/ui/accessibility/checklist.html. Accessed 05 Feb 2014
5. Library, I.D.: Accessibility Guide for IOS (2012). https://developer.apple.com/library/ios/documentation/UserExperience/Conceptual/iPhoneAccessibility/Introduction/Introduction.html. Accessed 03 Mar 2014
6. Nigay, L., Coutaz, J.: A design space for multimodal systems: concurrent processing and data fusion. In: Ashlund, S., et al. (eds.) Proceedings of the INTERCHI 1993 Conference on Human Factors in Computing Systems, pp. 172–178. IOS Press, Amsterdam (1993)
7. Bourguet, M.-L.: Towards a taxonomy of error–handling strategies in recognition-based multimodal human-computer interfaces. Sig. Process. J. **86**, 3625–3643 (2007)

8. RAFAEL, S.I.F.D.S.: Tese de Doutorado. Para uma Taxonomia da Multimodalidade [S.l.]: [s.n.] (2014)

9. Cohen, P.R., et al.: Synergistic use of direct manipulation and natural language. In: Proceedings of CHI 1989 Conference on Human Factors in Computer Systems, New York, USA, pp. 227–233 (1989)

10. Oviatt, S.L., Angeli, A.D., Kuhn, K.: Integration and synchronization of input modes during multimodal human–computer interaction. In: Proceedings of CHI 1997, pp. 415–422 (1997)

11. Vernier, F., Nigay, L.: A framework for the combination and characterization of output modalities. In: Palanque, P., Paternò, F. (eds.) DSV-IS 2000. LNCS, vol. 1946, pp. 35–50. Springer, Heidelberg (2001). doi:10.1007/3-540-44675-3_3

12. Wahlster, W.: Pointing, language and the visual world: towards multimodal input and output for natural language dialog systems (Panel). In: Proceedings of the 10th International Joint Conference on Artificial Intelligence, Milan, Italy. Morgan Kaufmann, Agosto 1163 (1987)

13. Cohen, P.R., et al.: QuickSet: multimodal interaction for distributed applications. In: ACM International Conference on Multimedia, Seattle, WA, pp. 31–40 (1997)

14. Bretan, I., Karlgren, J.: Synergy effects in natural language based multimodal interaction. In: Proceedings of the ERCIM 1993 Workshop on Multimodal Human–Computer Interaction, Nancy, France (1993)

15. Hall, D.L., Llinas, J.: An introduction to multi-sensor data fusion. In: Proceedings of the IEEE International Symposium on Circuits and Systems, pp. 6–23 (1997)

16. KO, T.H.: Untethered human motion recognition for a multimodal interface. Massachusetts Institute of Technology, Cambridge, Massachusetts, USA (2003)

17. Anastopoulou, S.: Investigating multimodal interactions for the design of learning environments: A case study in science learning. Thesis submitted to The University of Birmingham for the degree of Doctor of Philosophy, Birmingham, United Kingdom (2004)

18. Lisowska, A.: Multimodal Interface Design for Multimedia Meeting Content Retrieval. Switzerland, Geneva (2007)

19. Lee, J.C.: Spatial user interfaces: augmenting human sensibilities in a domestic kitchen. Massachusetts Institute of Technology, Cambridge, Massachusetts, USA (2005)

20. Pelachaud, C.: Multimodal expressive embodied conversational agents. In: Proceedings of the 13th Annual ACM International Conference on Multimedia, Singapore, pp. 683–689 (2005)

21. Reeves, L.M., et al.: Guidelines for multimodal user interface design. 1. ed. [S.l.]. Commun. ACM **47**, 57–59 (2004)

22. D'ulizia, A., Ferri, F.: Formalization of multimodal languages in pervasive computing paradigm. In: Paper Presented at the Third International Conference on Signal-Image Technology e Internet–Based Systems (Sitis 2006) (2006)

23. Gepner, B., Buttin, C., De Schonen, S.: Face processing in young autistic. Infant Behavior and Development, 661 (1994)

24. Mynatt, E.D.: Transforming graphical interfaces into auditory interfaces for blind users. Hum.-Comput. Interact. **12**, 7–45 (1997)

25. Oviatt, S.: Designing robust multimodal systems for diverse users and environments. In: Workshop on Universal Accessibility of Ubiquitous Computing: Providing for the Elderly (2001)

26. Dumas, B., Lalanne, D., Oviatt, S.: Multimodal interfaces: a survey of principles, models and framework. Hum. Mach. Interact., 3–26 (2009)

27. Oviatt, S., Coulston, R., Lunsford, R.: When do we interact multimodally? Cognitive load and multimodal communication patterns. In: Proceedings of the 6th IEEE International Conference on Multimodal Interfaces, State College, PA, USA, pp. 129–136 (2004)

28. Anthony, L., Yang, J., Koedinger, K.R.: Evaluation of multimodal input for entering. In: ACM Conference on Human Factors in Computing Systems (CHI 2005), Portland, OR, USA, pp. 1184–1187 (2005)
29. Anthony, L., Yang, J., Koedinger, K.R.: Entering mathematical equations multimodally: results on usability and interaction patterns. Technical report CMU–HCII–06–101 (2006)
30. Xiao, B., Girand, C., Oviatt, S.L.: Multimodal integration patterns in children. In: Proceedings of ICSLP 2002, pp. 629–632 (2002)
31. Oviatt, S.: Advances in robust multimodal interface design. IEEE Comput. Graph. Appl. **23**(5), 62–68 (2003)
32. Talarico Neto, A. Uma abordagem para projeto de aplicações com interação multimodal na web. Tese de Doutorado. São Carlos: [s.n.], 199 p (2011)
33. Raposo, N., Rios, H., Lima, D., Gadelha, B., Castro, T.: An application of mobility aids for the visually impaired. In: Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia (MUM 2014), New York, NY, USA, pp. 180–189 (2014)
34. W3C. Resultados preliminares - Pesquisa sobre uso de Tecnologias Assistivas: Ampliadores e leitores de tela, 21 Janeiro (2013). http://acessibilidade.w3c.br/pesquisa/resultados-preliminares/
35. WCAG. Web Content Accessibility Guidelines (WCAG) 2.0 (2014). https://www.w3.org/Translations/WCAG20-pt-br/. Accessed 15 Nov 2014
36. ACHECKER. Web Accessibility Checker (2011). http://achecker.ca/checker/index.php. Accessed 03 Nov 2014
37. W3C. Web Accessibility Evaluation Tools List. Web Accessibility Initiative (2006). https://www.w3.org/WAI/ER/tools/. Accessed 05 July 2014