e-Learning Environment with Multimodal Interaction

A proposal to improve the usability, accessibility and learnability of e-learning environments

André Constantino da Silva
Institute of Computing (PG)
UNICAMP, IFSP
Campinas, Brazil, Hortolândia, Brazil
acsilva@ic.unicamp.br

Heloísa Vieira da Rocha
Institute of Computing, NIED
UNICAMP
Campinas, Brazil
heloisa@ic.unicamp.br

***Abstract*—The Human-Computer Interaction is challenging the use of many modalities to interact with an application. The e-Learning environments interfaces are been exposed to this diversity of modalities, but they are designed to be used with a limited set. The impact is that users have interaction problems caused by the cross modality. The e-Learning environments need to evolve allowing users to interact with a more broadly interaction styles. One solution is adopt Multimodal Interaction concepts, improving the usability and accessibility of the e-Learning environment and make possible to embrace better learning contexts, property that we define as learnability.**

Keywords-Human-Computer Interaction; Interaction Styles; Multimodal Interaction; Electronic Learning Environment.

I. INTRODUCTION

The Human-Computer Interaction is challenging the replacement of the mouse and actual interfaces for interfaces that works with natural interaction, non-tactile interfaces, speech recognition, facial and movement recognition and gestures [1][2]. So, there are many ways to interact with digital artifacts and applications, like keyboards, mouse, small, medium and big displays, voice, touch and gestures. Many hardware components are available supporting different interaction styles.

The Multimodal Interaction is a solution to possibility the use of an application in this diversity of interaction styles, allowing users interact with computers by many input modalities (e.g., speech, gesture, eye tracking) and output channels (e.g., text, graphics, audio) [3]. Multimodal Interaction is proposed to turn the human-computer interaction more natural, i.e., more close to the human-human interaction. The main benefices are the increase of application's usability, accessibility, flexibility and convenience [3]. But, building a multimodal interaction system is not a trivial task yet, because the literature does not have sufficient information about how to design this kind of system and there is a lack of technologies to support them.

e-Learning environments like Moodle [4], SAKAI [5], TelEduc [6], Ae [7] are applications that use the Web infra-structure to support teaching and learning activities. The e-Learning environments are designed to support a variety of users and learning contexts, but they are designed to support a limited interaction styles, usually keyboard and mouse as input and a medium screen as output.

To attend this demand, the e-Learning environment needs to have good usability, accessibility, mobility and learnability. Considering all these dimensions is not a trivial task. Does the multimodality can improve these requirements on e-Learning environments?

Section II presents a literature review about e-Learning environments, multimodality and multimodal interfaces. Section III presents the research problem that we want to deal, and Section IV our hypothesis and methods. Section V some preliminary results and expected contributions.

II. LITERATURE REVIEW

The actual versions of e-Learning environments take advantages of the Web to offer content with text, images, audios and videos in a hypertext document. Tools like chat, forums, portfolios, repositories are widely used, and tools those explore the audio and video resource to user communication, such as instant messenger and video-conferences, are becoming common among the environments.

Due the diversity of users whom may use the e-Learning environments, these systems need to have good usability so that the user interface does not prejudice the teaching and learning activities. Accessibility is another important requirement to allow disabled people to use the environment. So usability and accessibility are two desired requirements to the e-Learning environments.

Devices, such as smartphones and tablets, are becoming increasingly popular; most of them have touch screen displays, access to the Internet and enough computing power to process Web pages. So, Web sites and Web applications, initially developed to be used with keyboard, mouse and a medium size display, are been accessed by small touchscreen devices. This is another aspect of accessibility, so the environments´ development teams are building solutions to provide access on mobile devices. Three kind of solution are emerging: specific device application; web site specific for mobile devices; and improve the web site for mobile and desktop access [8].

Two motivations allow the participants interacting anytime and anywhere with the content and each other; but, due to the device restrictions, there are needs to obtain better design solutions. The actual user interface design techniques take account just a limit set of input and output hardware, limited to the context, such as techniques to design user interface for desktop or for mobile platforms. But, there is a

 ISBN: 978-1-61208-250-9

lot of input or output hardware in these devices and these techniques are asked to consider all of them. Some input and output devices are: touchscreen, microphones, pen sensitive screen, touchpad, TrackPoint, accelerometers, joysticks, loudspeakers, small screen, large screen, printers, etc.

Another e-Learning environment characteristic is to be used in many of learning context, e.g., teacher training, undergraduate courses, and team training in all areas of knowledge. We call these as learnability. But, the actual hardware increases the difficulty to use the environment to produce content for any area and support student activities.

To attend this demand, the e-Learning environment needs to be usable and accessible for many users in many social, physical, technological and learning contexts. So, e-Learning environment needs to be evaluated in the usability, accessibility, mobility and learnability dimensions, a not trivial task.

Since the e-Learning environments were building to Web, they have a common architecture: the client-server. Client is responsible to render the user interface through a browser. It is in the client side that the user interacts with the system using input and output hardware. The server is responsible to process client´s requests and data persistence. The server knows few about the input and output devices in client side.

Ae is an e-Learning environment developed by a consortium of Brazilian universities using J2EE technology and component-based development process. Layered component-based software architecture was defined for Ae environment [9] with the following layers:

Presentation layer: provides the application user interface;

System layer: provides an interface for the application functionality, that is, it is a façade for the application business rules;

e-Learning layer: provides the component interfaces that implement the application's business rules, which can be used by various applications and which use services and functionalities from the infrastructure layer to implement the business rules;

Infrastructure layer: implements a set of infrastructure services such as, for example, data persistence;

Common services layer: has the public services that can be utilized and accessed by all other architecture layers, except the presentation layer.

Multimodal interaction is a research proposal to turn the interaction between humans and machines more natural, i.e., more close to the interactions between two humans, and have the benefits to increase the usability, flexibility and convenience [3].

Modality is the used term to define a mode what the user data input or a system output is expressed. The communication mode refers to the communication model used by two different entities to interact [10]. Nigay and Coutaz [11] define modality as an interaction method that an agent can use to reach a goal, and it can be described in general terms such "speech" or in specific terms such "using microphones".

For monomodal systems, designers are not limited to choose only one modality. But, in multimodal systems, they can choose many modalities, that used together, increasing the system flexibility and gives other benefices. Interfaces with this characteristic are called as multimodal interfaces and the system are called multimodal interaction systems. Mayes [12] defines multimodal interaction systems as a system with the capacity to communicate with the user by many different communication modes, using more than one modality, automatically gives or extracts mean.

According to Oviatt [10] "Multimodal interfaces process two or more combined user input modes (such as speech, pen, touch, manual gesture, gaze, and head and body movements) in a coordinated manner with multimedia system output".

Lalanne *et al.* [3] describe multimodal interaction systems, or multimodal systems, allow users to interact with computers though many data input modalities (e.g., speech, gesture, eye gaze) and output channels (e.g., text, graphics, sound, avatars, voice synthesis).

Bangalore and Johnston [13] say the critical advantage of multimodal interfaces is that they allow user input and system output to be expressed in the mode or modes to which they are best suited, given the task at hand, user preferences, and the physical and social environment of the interaction. Allowing users interact with many modes, it is possible to improve the accessibility because a multimodal interface can be used for a disabled person using the interaction mode that she can handle.

Multimodal content is common on multimedia systems. The research problem that we want proposes a solution is to use multimodality on user interaction and get benefits of multimodal content too.

Dumas, Lalanne and Oviatt [14] present a generic architecture for multimodal systems, turning more easy to understand the mainly components of the multimodal systems: a fusion engine, a fission module, a dialog manager and a context manager, which all together form what is called the "integration committee". The authors define "input modalities are received though various recognizers, which output their results to the fusion engine in charge of giving a common interpretation of the inputs. A fusion machine gives an interpretation for the data and it communicates it to the dialog manager, in charge of identifying the dialog state, the transition to perform, the action to communicate to an application, and/or the message to return through the fission machine. The fission machine returns a message to the user through the most adequate modality or combination of modalities, depending on the user profile and context of use. For this reason, the context manager, in charge of tracking the location, context and user profile, closely communicates any changes in the environment to the three other components, so that they can adapt their interpretations". Multimodal systems need to take account of all input done by the user to identify and process the solicited action.

Developing multimodal interaction systems is a complex task [14]; but, to turn more easy the development there are some frameworks, such as the ICARE framework [15], FAME [16] and special approaches [17][18].

 ISBN: 978-1-61208-250-9

Bouchet, Nigay and Ganille [15] define formally the CARE properties to characterize and assess aspects of multimodal interaction: the Complementarity, Assignment, Redundancy, and Equivalence that may occur between the interaction techniques available in a multimodal user interface. To aim build multimodal system, the authors propose the ICARE framework.

Larson [19] shows three general questions to response with a web application will be improved with a new mode of input: the new mode needs to add value to the web application, the application leverages the strengths of the new mode and avoids its weaknesses, and the users need to have the required hardware and software.

To implement multimodal system for web it is necessary consider both architecture: for multimodal systems and for web systems. Gruenstein, McGraw and Badr [20] present a framework to develop multimodal interfaces for web, the WAMI Toolkit. The framework defines tree client-side components (Core GUI, GUI Controller and Audio Controller) and more four server-side components (Web Server, Speech Recognizer, Speech Synthesizer, and Logger). The user interact with the Core GUI, described at HTML and JavaScript, and the Audio Controller, a Java Applet to receive the audio input. The collected data is sent to server to be treated by the Speech Recognizer and the Web Server components.

Zaguia *et al*. [21] present an approach to develop multimodal systems using fusion machines dispose on web services in such a way the user can choose the modalities that she sees fit to her situation instead of already pre-defined modalities from the beginning.

But we need to not only build a multimodal system; we are worry about the environment usability so that the interface does not prejudice the teaching and learning activities. Nielsen [22] defines usability as a combination of five elements: easy to learning, efficient, easy to remember, low probability of users do mistakes and user satisfaction.

III. THE RESEARCH PROBLEM

Since the number of devices accessing the Web grows, the e-Learning environments are exposed to a variety of interaction styles, including ones that are not considered in the design time. Just supporting these interaction styles causes cross modality interaction problems [8], limitation in the multimodality use, and do not take advantages from the interaction style in use. So it is necessary to develop a solution without these limitations.

Thinking about the learning domain, we ask "how do the users interact with a multimodal e-Learning environment? Is this a solution to allow an application be accessed by different devices with a diversity of users, physical and social contexts avoiding cross modality problems and get the better use of the interaction styles?"

Since there are no one multimodal interaction e-Learning environment to aim us to response these questions, we need built one. So we want to know how to develop a multimodal interaction e-Learning environment? Which are the cross modality problems related with this context? How can we identify?

It is the main problem, but several others derive: Is it possible to improve an application to be used by many interaction styles and get advantages? Which kind of modifications is necessary to get the best use of an interaction style? How to guaranties that the usability will not be affected? After the modifications, the application has a better accessibility? Allowing many interaction styles, do we have a new kind of application? How to distinguish the applications that have these characteristics from the other that does not have?

Due the multimodality allows users interact with many modes, maybe the user will use this mode not only to interact with the application, but to create content. How this impact in the environment architecture?

The special interest in the learning domain is the necessity to improve the e-Learning environment to better attend the teaching and learning activities in the variety of learning context, reaching out the maximum of learnability.

IV. THE RESEARCH HYPOTHESIS AND METHODS

We argue that it is possible improve the usability, accessibility and learnability of an e-Learning environment adding multimodality concepts in the environment´s user interface.

We planned to use empirical method to validate our hypothesis, building a multimodal interaction e-Learning environment prototype and doing user study, collecting interaction data and user opinions by observation and questionnaires. Not one interaction style will be studied per time; we want study various interaction styles being used to complete some tasks in the environment. The collected data will be used to verify the e-Learning environment in three dimensions: usability, accessibility and learnability.

Due to the quantity of interaction styles and the complexity to build a multimodal system, we need to define an incremental process to build the prototype, taking one or a limited number of interaction styles per time. But, it is important to reduce the development efforts for the next iteration, when another interaction style will be selected. Here, we will apply some software engineering techniques, like software components. We proposed these steps for this process:

1. Select an interaction style;
2. Do user studies to collect interaction problems using the selected interaction style and how the manner to realize the tasks change;
3. Redesign the software user interface to defining better solutions for identified problems;
4. Analyze the software to identify the components responsible to user interface and components that manipulate user data;
5. Find or implement recognizers and synthesizers for the selected interaction style;
6. Change the software architecture to have the main components of a generic multimodal system;
7. Do tests to collect errors and fix them.

We believe if we can do these steps two or three times; so, there is possible to repeat it until all interaction styles is considered.

 ISBN: 978-1-61208-250-9

To prove our solution, we propose to implement the Multimodal Interaction concepts in the Ae e-Learning environment [7], and use this new kind of environment to research advantages and disadvantages of multimodal interfaces in learning. We planned study the interaction using pen, touch and gesture in two tools of the Ae e-Learning environment: the Weblog and the Whiteboard. Both tools are used to construct the content and are selected based on our premises that they are good choices to study the multimodality.

V. PRELIMINARY RESULTS AND EXPECTED CONTRIBUTIONS

One of the preliminary results were some problems in the TelEduc environment that happened when users interact with touchscreen devices [8], i.e., some problems happened due the platform changing (when user access the environment using a smartphone) and some problems happened due the modality changing. The problems identified in TelEduc due the modality changing will happen on Ae environment.

Other contribution is related with how the multimodal concept changes the architecture of an application and how to find solutions. The Ae architecture needs to be changed to adopt the multimodal concepts. Considering the Web-Accessible Multimodal Interfaces architecture [20] and the architecture of multimodal systems [14], we redesign the Ae architecture (Fig. 1). Due to the fact that the browser has the responsibility to show the GUI components in the client side, some components are added to treat the data input from the input devices. The user interaction data is sent to the server, who have the responsibility to process this data and gives an interpretation for the received data. After the input data interpretation process, the correspondent action is sent to the system component. So, the fusion and fission machines will be on the server. The fusion machine will be called when the server receive the data from the client-side components, and the fission machine will be called when the system response the request, after the data processing. Since, we are using a component-based architecture, the presentation layer and the multimodal components can run in a proper server, increasing the scalability and performance.

The main expected contribution is to know more about the relationship between usability, accessibility and multimodality. For the educational context, we want to know if the support more interaction styles there is an impact in the learning contexts, so define the learnability.

The prototype is another contribution, because there is no one multimodal interaction e-Learning environment, we will call it |Ae|.

ACKNOWLEDGMENT

The authors thank the CAPES and CNPq for financial support and for FAPESP through TIDIA-Ae Project (n. 05/60572-1), which provided the equipment.

REFERENCES

[1] L. Kugler. “Goodbye, Computer Mouse”. Communications of ACM, vol. 51, n. 9, September 2008, pp. 56.

[2] B. Hayes. “Gartner´s Seven IT Grand Challenges”. Communications of ACM, vol 51, n. 7, July 2008, pp. 10.

[3] D. Lalanne, L. Nigay, P. Palanque, P. Robinson, J. Vanderdonckt, and J. Ladry. “Fusion Engine for Multimodal Input: A Survey”. Proc. of the 11th International Conference on Multimodal Interfaces (ICMI-MLMI´09), ACM Press, Nov. 2009, pp. 153-160, doi: 10.1145/1647314.1647343.

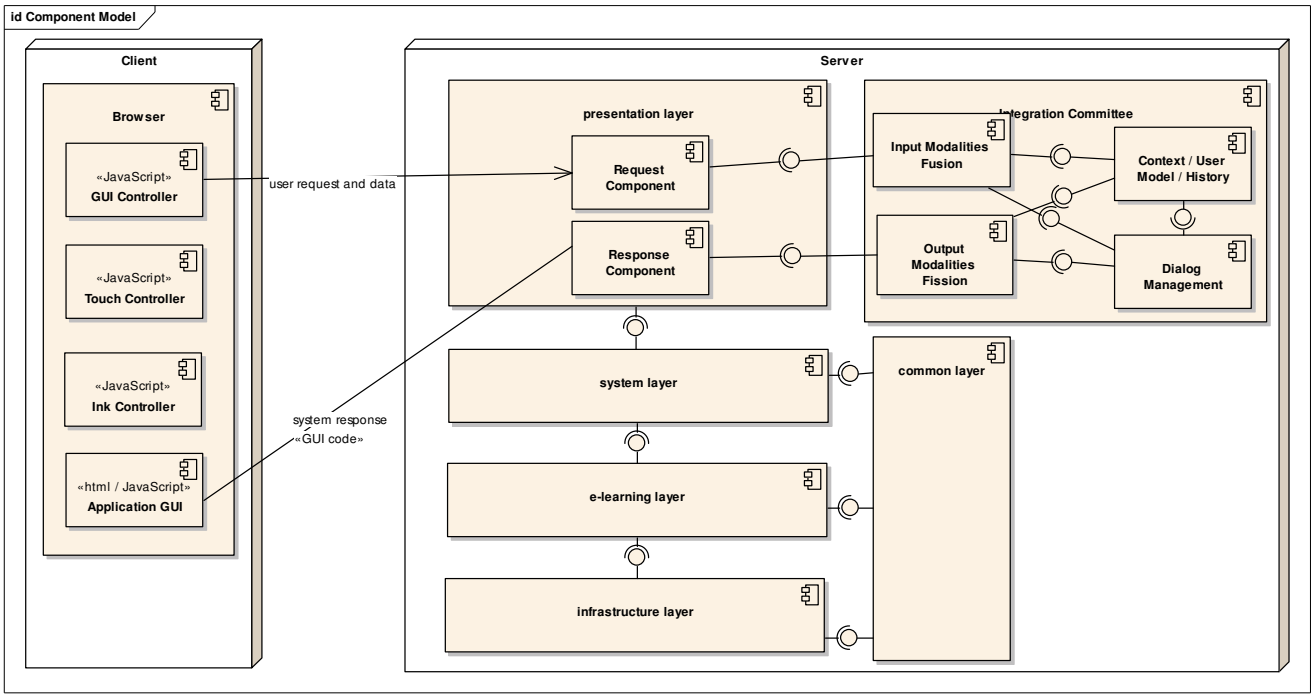

Figure 1. Ae architecture with multimodal components.

 ISBN: 978-1-61208-250-9

[4] Moodle Trust. “Moodle.org: open-source community-based tools for learning”. Available at <http://moodle.org>. [retrieved: Jan. 2013]

[5] SAKAI Environment. “Sakai Project | collaboration and learning - for educators by educators”. Available at <http://sakaiproject.org>. [retrieved: Jan. 2013]

[6] TelEduc Environment. “TelEduc Ensino à Distância”. Available at <http://www.teleduc.org.br>. [retrieved: Jan. 2013]

[7] Ae Project. “Ae - Aprendizado Eletrônico Environment”. Available at <http://tidia-ae.iv.org.br/>. [retrieved: Jan. 2013]

[8] A. C. da Silva, F. M. P. Freire, and H. V. da Rocha. “Identifying Cross-Platform and Cross-Modality Interaction Problems in e-Learning Environments”. Proc. of the 6th International Conference on Advances in Computer-Human Interactions (ACHI 2013), IARIA, 2013. / in press /

[9] D. M. Beder, A. C. da Silva, J. L. Otsuka, C. G. Silva, and H. V. da Rocha. “A Case Study of the Development of e-Learning Systems Following a Component-based Layered Architecture”. Proc. of the 7th IEEE International Conference on Advanced Learning Technologies (ICALT 2007), IEEE Press, Jul. 2007, pp. 21-25, doi: 10.1109/ICALT.2007.4.

[10] S. L. Oviatt. “Advances in Robust Multimodal Interface Design”, in IEEE Computer Graphics and Applications, vol. 23, no. 5, Sep. 2003, pp. 62-68, doi: 10.1109/MCG.2003.1231179.

[11] L. Nigay and J. Coutaz. “A Generic Platform for Addressing the Multimodal Challenge”. Proc. of 13th Conference on Human Factors in Computing Systems (SIGCHI 1995), ACM Press / Addison-Wesley Publishing Co., May 1995, pp. 98-105, doi: 10.1145/223904.223917.

[12] T. Mayes. “The ‘M’ Word: Multimedia interfaces and their role in interactive learning systems”, in Multimedia Interface Design in Education, A. D. N. Edwards and S. Holland, Eds. Berlin: Springer-Verlag, 1992, pp. 1-22, doi: 10.1007/978-3-642-58126-7_1.

[13] S. Bangalore and M. Johnston. “Robust Understanding in Multimodal Interaction”. Computational Linguistic, Cambridge, Massachusetts, vol. 35, n. 3, Sep. 2009, pp. 345-397.

[14] B. Dumas, D. Lalanne, and S. Oviatt. “Multimodal Interfaces: A Survey of Principles, Models and Frameworks”, in Human-Machine Interaction, D. Lalanne and J. Kohlas, Eds. Berlin: Springer Berlin / Heidelberg, 2009, pp. 3-26, doi: 10.1007/978-3-642-00437-7_1.

[15] J. Bouchet, L. Nigay, and T. Ganille. “ICARE: Software Components for Rapidly Developing Multimodal Interfaces”. Proc. of 6th International Conference on Multimodal Interfaces (ICMI 2004), ACM Press, Oct. 2004, pp. 251-258, doi: 10.1145/1027933.1027975.

[16] C. Duarte and L. Carriço. “A Conceptual Framework for Developing Adaptive Multimodal Applications”. Proc. of 11th International Conference on Intelligent User Interfaces (IUI 2006), ACM Press, Jan. 2006, pp. 132-139, doi: 10.1145/1111449.1111481.

[17] A. Stanciulescu, Q. Limbourg, J. Banderdonckt, B. Michotte, and F. Montero. “A Transformational Approach for Multimodal Web User Interfaces based on UsiXML”. Proc. of 7th International Conference on Multimodal Interfaces (ICMI 2005), ACM Press, Oct. 2005, pp. 259-266, doi: 10.1145/1088463.1088508.

[18] A. T. Neto, T. J. Bittar, R. P. M. Fortes, and K. Felizardo. “Developing and Evaluating Web Multimodal Interfaces – A Case Study with Usability Principles”. Proc. of 24th ACM Symposium on Applied Computing (SAC 2009), ACM Press, Mar. 2009, pp. 116-120, doi: 10.1145/1529282.1529306.

[19] J. A. Larson. “Should You Build a Multimodal Interface for Your Web Site?”. Available at < http://www.informit.com/articles/article.aspx?p=29024>. [retrieved: Jan. 2013]

[20] A. Gruenstein, I. McGraw, and I. Badr. “The WAMI Toolkit for Developing, Deploying, and Evaluating Web-Accessible Multimodal Interfaces”. Proc. of 10th International Conference on Multimodal Interfaces (ICMI 2008), ACM Press, Oct. 2008, pp. 141-148, doi: 10.1145/1452392.1452420.

[21] A. Zaguia, M. D. Hina, C. Tadj, and A. Ramdane-Cherif. “Using Multimodal Fusion in Accessing Web Services”, in Journal of Emerging Trends in Computing and Information Sciences, vol. 1, n. 2, Oct. 2010, pp. 121-137.

[22] J. Nielsen. Usability Engineering. EUA: Morgan Kaufmann, 1993. 362 p.

 ISBN: 978-1-61208-250-9