

MIMBCD-UI

User Testing Guide

Francisco Maria Calisto
francisco.calisto@tecnico.ulisboa.pt

09/04/2018

Prototype: [prototype-breast-cancer](#)
Milestone: [1.0.6-alpha](#)

Version: [v1.0.6-alpha](#)
Release: [v1.0.6-alpha](#)

DICOM: [dicom-server](#)
Commit: [d889ba07d715a3f47bc01634f2163af30b147a20](#)

Deployment Environment: Test
Link: [breastscreening.isr.tecnico.ulisboa.pt](#)

Deployment Server: Test

Main Server: [Test](#)
Private IP: [10.0.1.23](#)
Private Domain: [cromo.isrnet](#)

Port: 8081
Public IP: [193.136.138.62](#)

DICOM Server: [Test](#)
Private IP: [10.0.1.23](#)
Private Domain: [cromo.isrnet](#)

Port: 8043
Public IP: [193.136.138.62](#)
From: 8143

1 Introduction

This document aims to describe the protocol performing a set of tests in the scope of [v1.0.6-alpha](#) version from the [prototype-breast-cancer](#) repository of the [Medical Imaging Multimodality of Breast Cancer Diagnosis User Interface \(MIMBCD-UI\)](#) project using a traditional devices (mouse and keyboard). The goal of the test is to understand the user, performance, efficiency and efficacy metrics. With the session, the sessions will be recorded via video on the computer and using a record, heat-map and triggered event tools. It is guaranteed the confidentiality of the recordings, which will be used only for academic purposes.

Dividing the activity session into three distinct phases of the three tasks. In all three tasks, by supporting our traditional devices, the interaction with mouse and keyboard. The first phase is the act of doing those tasks. On the second phase, we will do a small questionnaire at the end of each task using [NASA-TLX](#) [9]. Finally, the third phase we will have a final survey regarding the Usability of the system. The well-known scale called [System Usability Scale \(SUS\)](#) [8].

2 Material

For the material and apparatus, it is essential to capture the session apprehending the user interactions. In our case, we will record this interaction by using the [QuickTime Player Version 10.4 \(928.5.1\)](#) to obtain all interactions. We will pair this video tool with a user watch tool called [Hotjar](#). This tool serves the purpose of using several logs of the interaction and gives us visualisation over it. Both instruments will help us to capture where are users interacting. By looking at the test participant's reactions, we find a lot of information regarding the prototype design.

The tools that we choose for the material and apparatus of this User Testing Guide are low-cost and easy to use. Our equipment is a cost-effective and, by using our laboratory materials, bringing it to the radiology room, we enable to capture not only what the user is doing on the screen, but on the body language supported by the interviews and observation.

The material used in the test sessions for the user interface consists of:

- MacBook Pro: it will allow the user to interact with the keyboard and a wireless mouse;
- Wireless Mouse: it will allow the user to interact with a mouse and will complement the keyboard;

2.1 Technical Details

To produce this traditional environment, and since we can simulate with a laptop, the mouse and keyboard interaction, we are using a Microsoft Mobile Mouse 4000 together with the [MacBook Pro](#) (Retina, 13-inch, Early 2015) with a standard integrated keyboard on the laptop.

2.2 Software

To track our user interactions across our system, we are using [Hotjar](#). This tool is an analytic package allowing us to follow our users remotely. It also provides two critical pieces of functionality, among others, that can aid in remote user testing. First of all, the heatmaps allow us to see where users are clicking, tapping and scrolling on our system. Second, it records a video playback of the entire user session. The tool shows evidence of being useful for our studies while we successfully used it in the past.

To record the task activities and the interview, we used [QuickTime](#) [10]. The [QuickTime \(Apple Computer\)](#) tool is available for [MacBook Pro](#) to movie, audio and screen recording. Despite of have an overall of features, we just used it for our user's screen recording. It provides this functionalities at minimum requirements and compatible to our apparatus.

3 Description

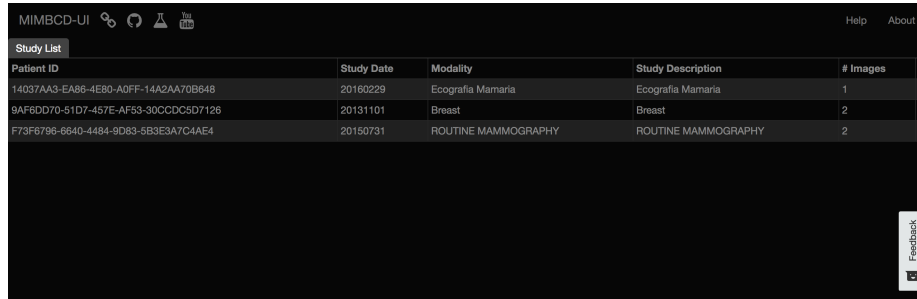
To verify our work, we identified measurable and explicit targets. By having several goals, including that a value percentage of the users should be able to operate the tasks without the need of help. On the same rate value, the user should be able to start and complete the medical diagnosis tasks over the system with little errors or mitigating those errors. Measuring the expected number of errors with a relation between our laboratory pilot tests. On the laboratory pilot tests we aim to test our prototypes with Researchers. The Researchers are in the context of the system and know well the functionalities so that we need to expect a percentage value over their results compared to Clinicians and not the same benefits. Last but not now least, both users (Researchers and Clinicians) should be able to understand in a similar time amount the meaning of all visible controls. By the similar amount of time, it is expected to have a variance of the percentage value between Researchers and Clinicians of the same value percentage of the early goals described in this paragraph.

We tested each objective in early laboratory and field tests so that we could take the appropriate corrective actions. Also, we expect to run early field tests with Researchers and Clinicians to highlight issues that we overlooked and ignored during the prototyping phase. To support interaction use by the Clinicians, we will try to emphasise several key factors on our user tests. The tasks must be simple, low intrusive, support for natural interaction and the system must always give visibility and the task current-state.

3.1 Devices

Traditional interaction remains the most common way to interact with user interfaces in a clinical environment. Unfortunately, most of this interaction is made by low profile equipment that makes users produce more errors and take more time interacting with those user interfaces.

On Figure 1 the user can select the list of patients. The list has a table with several patient information. The first column is the *Patient ID*; we used it as an identifier of the patient. That way we can have anonymised information with no reference to the patient name. The second column is the *Study Date*, the third column is the *Modality* of the used **DICOM** image, the fourth column is the *Study Description* of the used study and the last column is the number of *Images*.



Patient ID	Study Date	Modality	Study Description	# Images
14037AA3-EA86-4E80-A0FF-14A2AA70B548	20160229	Ecografia Mamaria	Ecografia Mamaria	1
9AF6DD70-51D7-457E-AF53-30CCDC5D7126	20131101	Breast	Breast	2
F73F6796-664D-4484-9D83-5B3E3A7CAAE4	20150731	ROUTINE MAMMOGRAPHY	ROUTINE MAMMOGRAPHY	2

Figure 1: List of patients.

As we can see in Figure 2, it shows the first task in our User Interface (UI), where the patient's breasts are on a small left column. The options are in a short row near of the viewport and described below. We also have the tabs where the user can change the patient. The centre viewport shows the **DICOM** image, and it can be configured to display a number up to four **DICOM** images. The viewport has some text information on it (yellow) with the details of the metadata.

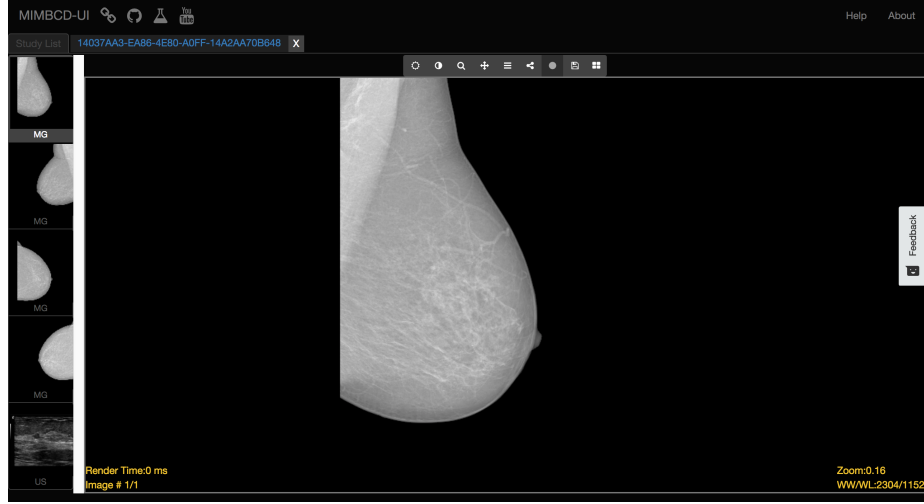


Figure 2: Viewer of the **DICOM** images.

Manual annotation is adopted by us thanks to Freehand ROI and Probe annotation features, both from CornerstoneJS. According to the CornerstoneJS Library, the user can create an annotation by setting up consecutive landmarks around a Region of Interest (ROI). The markers finish a lesion annotation when it interconnects the historical. Additional features available in our User Interface (UI) includes on-demand increment of the number of landmarks, and throw transformations of the shape of an annotation.

3.2 User Interactions

The systems have several buttons (Figure 3) that allows the user to interact or access to a set of user interface features. Each item of the following list represents each metaphoric icon of Figure 3.



Figure 3: Toolbar of the System available features.

The buttons are (from left to right of Figure 3) as follows:

- WW/WC
- Invert
- Zoom
- Pan
- Stack Scroll
- Freehand
- Probe (Deactivated)
- Save
- Window Controller

3.3 Usability Evaluation Technique

Usability and functionality are the significant elements which can significantly affect the performance of a medical system. While some prior studies [1] have investigated the functionality of healthcare systems, the usability issue has mostly been overlooked in the existing Health Informatics (HI) literature regarding Human-Computer Interaction (HCI).

The following Table 3.3 is presenting six evaluation questions to have in mind during evaluation. The purpose of this questions is to facilitate systematic user studies in a clinical environment and support the identification of usability problems. The proposed issues involve various aspects of workload combined with either need for satisfaction or division of attention.

Number	Issues of Content Key Questions
1	How do you perceive this activity?
2	Could it be done in a more intuitive way?
3	What are the consequences?
4	Why did you do as you did with this activity?
5	Is this activity relevant for you?
6	Could you suggest another way to do this activity?

Table 1: Usability Evaluation Questions

The influence of perceived activity [5] is an important variable for our empirical analysis. In fact, the trust of the user increases when the user perceived that the system is usable and that there will be a consequent increase of the clinician trust in our system. This arguments explains the first question, the *How do you perceive this activity?* question. For the second question, the *Could it be done in a more intuitive way?* question, we aim to conclude if there is some solution for a more intuitive way of perceive the activity. Third, we intend to filter possible consequences of the clinician workflow by asking *What are the consequences?* directly to the clinician. The fourth question, underlines the reasons why the clinician did that way, with the question *Why did you do as you did with this activity?* we can understand the process of achieving the activity goal and the clinician’s interpretation of it. On the fifth question, where we ask *Is this activity relevant for you?*, we aim to understand the potential relevance of our system to the clinician. Last but not least, the six question is present to give the clinician opportunity to suggest improvements, reflecting the reasons why we ask the *Could you suggest another way to do this activity?* question.

To conclude this section, by doing this questions, we aim to support our user studies by giving our users, the clinicians, the opportunity of improving our empirical analysis regarding user’s *open answers*. However, the results should be treated with caution. Several bias exists since we are doing here an ambiguous approach.

4 System

The hereby prototype used is the [v1.0.6-alpha](#) version of our [prototype-breast-screening](#). The purpose of this prototype is to integrate a web-based tool by using medical imaging for the breast screening diagnosis. This web-based tool was created with a front-end and bak-end architecture utilising common programming languages, libraries, frameworks and tools including [JavaScript \(JS\)](#) [4], [NodeJS](#) [12], [HammerJS](#), [CornerstoneJS](#) [6] and [Orthanc](#) [7]. The central component of this prototype is a web-based [PACS](#) [2] pairwise with ubicous web technologies and based on the **Open Source (OS)** [3] [CornerstoneJS](#) [6] library.

4.1 Environments

This section describes the user environment over interaction, the so called **Radiology Room (RR)** (Figure 4). This guide is based on soft-copy diagnosis using computer workstations in their current reading room environment. It will be here where we take impressions regarding the efficacy of radiologist’s reading room design, and their recommendations based on their experience for improvements on the soft-copy reading environment. Several studies demonstrated that radiologist fatigue levels and performance are related to environmental factors such as monitor brightness and ambient room light in addition to workstation enhancements. Supported by this guide, our research aims to conduct an investigation for the several environmental variables. We expect to analyse the noise, temperature, seating ergonomics and so on.



Figure 4: Radiology Room

4.2 Case Studies

The functionality of the prototype will be best demonstrated by a series of case studies. By describing the expected workflow and capabilities of the research study at the **RR** specific environment and changes of the workflow by using our system prototype. The study implies the the evaluation of medical imaging features of several breast lesions. The primary goal of this case studies analysis is to generate a receiver operating characteristic to evaluate the performance and validation of our system. Let us consider a list of hypothetical use cases for the research investigation that evaluates the interaction and usability performance of the prototype. Therefore, the following list will show the preliminary case studies.

List of case studies to analyse our solution prototype:

- Multi-modality Imaging Features of a Breast Cancer Diagnosis;
- Radiologist Learning Curve Assessment;
- Remote Access;
- Intuitive Controls Analysis;
- Performance Measurement Values Acquisition;
- Usability Measurement Values Acquisition;
- Radiologist Validation;

We expect to demonstrate several uses through a series of case studies, including implementation of our research prototype for a multi-modality view studies and other imaging research, and creation of a novel tool for the purpose. By creating a set of questions, we will try to achieve and feed this case studies. It is automatically associated with all cases. The Radiologist may interact with our system manipulating the medical imaging and report to us difficulties and improvements. The number of questions is not restricted to the present document, since the interview will be open and suggestive. There is no limit to the number of questions that can be asked per case.

The data will be collect from the study video and observations into a spreadsheet for further analysis. We will report the results of this tests and conclusions. This guide and respective use cases will be iteratively improved.

5 Procedures

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

5.1 Initial Questionnaire

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

- Sex, Age, Literacy.
- Experience with ...
- Experience with ...
- Experience with ...

5.2 Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

5.3 Training Session

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

5.4 Execution of Tasks

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

5.5 Final Survey

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

6 Tasks

During our usability tests, we need to ask participants to provide a subjective assessment of their experience using our system. There are several widely used questionnaires giving us different pros-and-cons. However, in most cases, a [single question instrument](#) [11] is the right method for a quantitative usability testing. By taking less time and effort to answer, participants are pursuing to this phase after task while it is minimally disruptive.

In our **User Testing Guide** a set of tasks is necessary and carefully crafted. Our usability studies involve asking participants to perform the following tasks. By looking at what our user need to do with our system, our tasks are realistic as possible. We are not describing the exact steps participants need to take. We achieve that by avoiding the precise language used as labels in our system. The tasks are emotionally neutrals. And we did several [pilot tests](#) to prevent misleading situations saving us from wasting resources by accidentally use a lousy task or from getting bad data. The tasks are as follows.

List of stand alone tasks:

Task 1: Annotate the **US** modality from the **20160229** date;

Task 2: Annotate the last but one medical image of the **Breast** study description;

Task 3: Annotate the second **MR** frame number **10** of the last patient;

7 Measurements

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

- Measurement 1;
- Measurement 2;
- Measurement 3;
- Measurement 4;
- Measurement 5;
- Measurement 6;
- Measurement 7;
- Measurement 8;
- Measurement 9;

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

- Difficulty of ...
- Difficulty of ...
- Difficulty of ...
- Degree of ...

References

- [1] Francisco M. Calisto, Alfredo Ferreira, Jacinto C. Nascimento, and Daniel Gonçalves. Towards touch-based medical image diagnosis annotation. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, ISS '17*, pages 390–395, New York, NY, USA, 2017. ACM.
- [2] Robert E Cooke Jr, Michael G Gaeta, Dean M Kaufman, and John G Henrici. Picture archiving and communication system, June 3 2003. US Patent 6,574,629.
- [3] Joseph Feller, Brian Fitzgerald, et al. *Understanding open source software development*. Addison-Wesley London, 2002.
- [4] David Flanagan. *JavaScript: the definitive guide*. ” O’Reilly Media, Inc.”, 2006.
- [5] Carlos Flavián, Miguel Guinalíu, and Raquel Gurrea. The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & management*, 43(1):1–14, 2006.
- [6] Jason Hostetter, Nishanth Khanna, and Jacob C Mandell. Integration of a zero-footprint cloud-based picture archiving and communication system with customizable forms for radiology research and education. *Academic radiology*, 2018.
- [7] S. Jodogne, C. Bernard, M. Devillers, E. Lenaerts, and P. Coucke. Orthanc – A lightweight, RESTful DICOM server for healthcare and medical research. In *Biomedical Imaging (ISBI), IEEE 10th International Symposium on*, pages 190–193, San Francisco, CA, USA, April 2013.
- [8] Konstantina Orfanou, Nikolaos Tselios, and Christos Katsanos. Perceived usability evaluation of learning management systems: Empirical evaluation of the system usability scale. *The International Review of Research in Open and Distributed Learning*, 16(2), 2015.
- [9] Anjana Ramkumar, Pieter Jan Stappers, Wiro J Niessen, Sonja Adebahr, Tanja Schimek-Jasch, Ursula Nestle, and Yu Song. Using goms and nasa-tlx to evaluate human–computer interaction process in interactive segmentation. *International Journal of Human–Computer Interaction*, 33(2):123–134, 2017.
- [10] Melissa R Rowell, Frank M Corl, Pamela T Johnson, and Elliot K Fishman. Internet-based dissemination of educational audiocasts: a primer in podcasting-how to do it. *American Journal of Roentgenology*, 186(6):1792–1796, 2006.
- [11] J Sauro. 10 things to know about the single ease question (seq). *Measuring U*, 2012, 2012.

- [12] Jim Wilson. *Node.js & the Right Way: Practical, Server-side Javascript that Scales*. Pragmatic Bookshelf, 2018.