Towards Touch-Based Medical Image Diagnosis Annotation

Francisco M. Calisto
Institute for Systems and Robotics (ISR/IST), LARSyS
Instituto Superior Técnico
Universidade de Lisboa

Jacinto C. Nascimento
Institute for Systems and Robotics (ISR/IST), LARSyS
Instituto Superior Técnico
Universidade de Lisboa

Alfredo Ferreira
INESC-ID
Instituto Superior Técnico
Universidade de Lisboa

Daniel Gonçalves
INESC-ID
Instituto Superior Técnico
Universidade de Lisboa

{*francisco.calisto, jacinto.nascimento, alfredo.ferreira, daniel.j.goncalves*}*@tecnico.ulisboa.pt*

ISS '17, October 17–20, 2017, Brighton, United Kingdom

ACM ISBN 978-1-4503-4691-7/17/10
https://doi.org/10.1145/3132272.3134111

Abstract
A fundamental step in medical diagnosis for patient follow-up relies on the ability of radiologists to perform a trusty diagnostic from acquired images. Basically, the diagnosis strongly depends on the visual inspection over the shape of the lesions. As datasets increase in size, such visual evaluation becomes harder. For this reason, it is crucial to introduce easy-to-use interfaces that help the radiologists to perform a reliable visual inspection and allow the efficient delineation of the lesions. We will explore the radiologist's receptivity to the current touch environment solution. The advantages of touch are threefold: (i) the time performance is superior regarding the traditional use, (ii) it has more intuitive control and, (iii) for less time, the user interface delivers more information per action, concerning annotations. From our studies, we conclude that the radiologists still exhibit a resistance to change from traditional to touch based interfaces in current clinical setups.

Author Keywords
Touch-Based, Medical Image Diagnosis, Medical Visualization, Interaction Design, Human-Computer Interaction

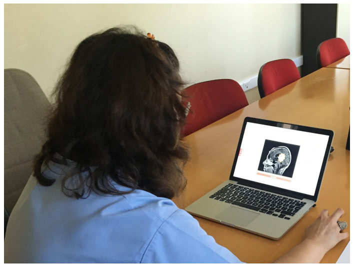

Figure 1. Radiologist interacting with traditional environment.

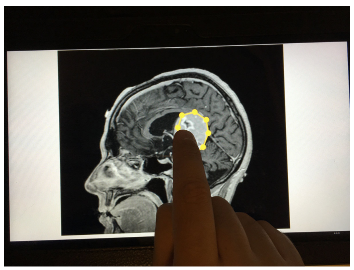

Figure 2. Touch environment interaction.

Phase	Order 1	Order 2
1	Trn (M)	Trn (T)
2	Tst (M)	Tst (T)
3	Q (M)	Q (T)
4	Trn (T)	Trn (M)
5	Tst (T)	Tst (M)
6	Q (T)	Q (M)
7	QF	QF

Table 1. Study Ordering.

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation]: User Interfaces.

Introduction

Medical imaging diagnosis is a topic of great interest, that has been the subject of intensive research in the field of medicine, and more specifically in Radiology [3, 4, 7]. However, current tools do not fully support annotation collection. These annotations and their relationships are crucial for a proper diagnosis. Namely, to have the morphological time evolution of potential lesions that may be present in some human organs.

In this paper, we present a performance and experience analysis conducting a study of touch and traditional environments (Figures 1 and 2) that are aligned to the previous mentioned goals. Both environments are comfortable to interact with and fast enough to solve the annotation tasks. For each environment, the same simple image feature (e.g. delete/correct annotation) is considered.

The performance differences of traditional and touch input devices have been studied [10] in terms of speed and accuracy for a variety of clinical and medical interactive tasks. This includes features such as Regions of Interest (ROI) annotations and length measurements, among others. Although, user experience while inserting annotations is important when evaluating the input devices, there is still a lack of research examining how the experience compromises the interactive surfaces based on the input type. To tackle this issue, our main goals focus primarily on cognition and time-performance. The experience is also considered as a means to validate the sensation, affect and value of interaction.

Indeed, while user experience analysis and evaluation has been applied to clinical systems [2], time-measure and error scales are still missing. Moreover, a touch interactive surface application can improve medical and clinical competences, since it is able to reduce the running time figures to complete common tasks in the diagnosis loop.

Evaluation

Table 1 shows the order of the user (radiologist) tests. Both Conditions (Tst (M) = Mouse Test, Tst (T) = Touch Test) took 5 minutes. Training (Trn) and Questionnaire (Q) took less then 1 minute. Participants complete a final questionnaire (QF = Final Questionnaire) that took less then 5 minutes. It can be seen that the "Order 1" column is the opposite regarding the "Order 2" column. In the "Order 1" column, the user trains on the mouse-based environment (Trn(M)), then executes a set of pre-defined tests (Tst(M)), followed by a questionnaire (Q(M)). The QF in the table, is related to a questionnaire that concerns issues such as the environment, functionality usability and seasickness issues among others. This procedure is then repeated for the touch environment. In the "order 2" column is similar, with the users training and performing their tasks in the touch setting first. Half of the users first perform the mouse and keyboard test, and the other half perform first the touch tests.

Task

Radiologists interact with this user interface by doing some annotations in certain regions of interest areas of the image. The images are chosen to be

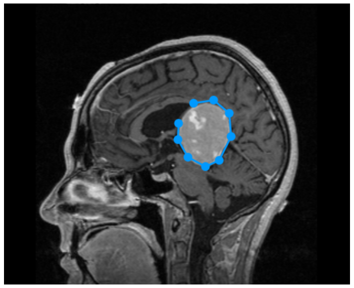

Figure 3. Pre-Annotated DICOM image.

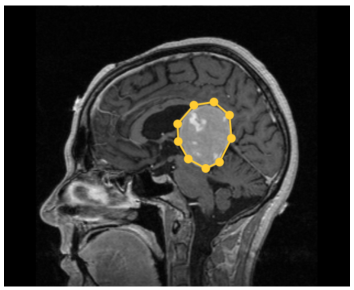

Figure 4. Post-Annotated DICOM image.

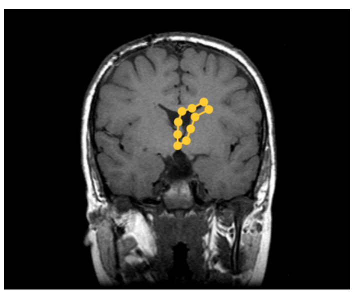

Figure 5. Hard difficulty DICOM annotated image.

representative to the domain field and to understand the different levels of the difficulty in these images. In our work, the first task is to perform the annotations on the DICOM Image along with the brain lesion near of the cortex, as we can see on Figure 3 and Figure 4. The second image is almost similar of the first one (just a different frame) so the task is mainly the same. The last image we put radiologists doing some more difficult annotation as it is shown in Figure 5.

Measures
The following scales are considered: (i) Starting by the Positive and Negative Affect Scale [9] (PANAS), (ii) Intrinsic Motivation Inventory [6] (IMI) and (iii), Experience Needs Satisfaction [1] (ENS). When we achieve the end of the condition, radiologists are asked to provide information and rank each condition, providing any comments.

Performance
We measure accuracy by computing the Hit Rate Score (HRS). This measure is defined has the percentage of annotations that lie on three areas (Figure 6). These three areas are measured from the ground-truth (see black line in Figure 6). The first area, **Area A**, is the area delimited by the two **Green Lines**, having a width of 2ϵ, where ϵ is the perpendicular distance from a point in the ground truth (black line) to the **Green Line**[1]. The second area, **Area B**, has also a width of 2ϵ. However, **Area B** embraces two regions, each region falling between the **Green Lines** (inner boundary) and **Red Lines** (outer boundary).

[1] Here, the ϵ is the diameter of the annotation point.

User Experience
To evaluate the user experience we embed, the IMI [6], PANAS [9] and ENS [1] measures in a 5-point Likert-scales, ranging from 1 (strongly disagree) to 5 (strongly agree).

Results

Since the data collected does not meet the applicability pre-conditions required for ANOVA, i.e., does not follow a Gaussian distribution and also the number of samples is small, we resort to use the Kruskal-Wallis Analysis of Variance [8] to test the performance and experience that is useful for testing between-subjects effects for three or more conditions [5].

Table 2 shows the overall results of the Kruskal-Wallis test. This result shows that all the mean differences for each of the radiologists is significant in the above 11 conditions (see Tables 3, 4 and 5).

A small number of radiologists is used in the tests. Three of the five radiologists (1, 2 and 5) have shown that touch environment is better classified then traditional environment. Although, we must consider and understand the other two radiologist (3 and 4) results. The Kruskal-Wallis test[2]revealed that radiologists have a significant main effect on touch environment ($\chi^2_{\text{tou}} = 1.711$, $\alpha_{\text{tou}} = 0.789$), against traditional environment ($\chi^2_{\text{tra}} = 4.587$, $\alpha_{\text{tra}} = 0.332$).

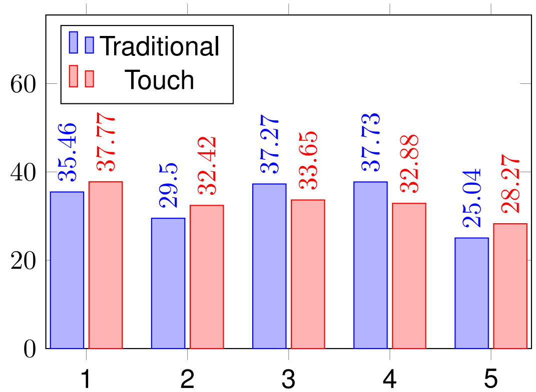

Table 2. Kruskal-Wallis.

In the following sections (Table 2 - Table 5) we describe the results in terms of the first and second order statistics, using the mean (M_{env}) and standard deviation (σ_{env}), as well as the standard error of the mean[3] that is defined as:

$$sem_{\text{env}} = \sigma_{\text{env}} / \sqrt{N}$$

Performance
Table 3 shows the mean time performance of the tasks as described in section Task.

Time. A significant main effect of the traditional input environment (M_{tra}=153.4s, σ_{tra}=40.88, $\text{sem}_{\text{tra}}(\bar{x}) = 18.28$) with radiologist maximum time spent of 210 seconds and radiologist minimum time spent of 109 seconds. For the touch input environment (M_{tou}=114.20s, σ_{tou}=51.71, $\text{sem}_{\text{tou}}(\bar{x}) = 23.12$) the radiologists take less time.

Number of Interactions (NI). The mean of the interactions on traditional (M_{tra}=39.20) is less than the touch environment (M_{tou}=40).

Hit Rate Score (HRS). It is shown that the HRS in traditional environment (M_{tra}=75.2, σ_{tra}=22.72, $\text{sem}_{\text{tra}}(\bar{x}) = 10.16$), has a higher value then touch environment (M_{tou}=50.6, σ_{tou}=15.35, $\text{sem}_{\text{tou}}(\bar{x}) = 6.86$).

User Experience
Responses to the final questionnaire suggest that both interactions are adequate for analyzing medical images. Moreover, no radiologist reported discomfort, dizziness or fatigue in neither environments.

Motivational
The motivational results are presented as *Competence*, *Autonomy*, *Relatedness* and *Immersion* (Table 4) options.

Competence. The main outcome on traditional environment (M_{tra}=4.2, σ_{tra}=0.83, $\text{sem}_{\text{tra}}(\bar{x}) = 0.37$), has shown that radiologists feel more comfortable in traditional environment, as it can be expected. On the other hand, for touch environment (M_{tou}=3.4, σ_{tou}=1.67, $\text{sem}_{\text{tou}}(\bar{x}) = 0.74$), radiologists showed a rated above average, however, lower than traditional environment.

Autonomy. The traditional environment reflected (M_{tra}=4.4, σ_{tra}=0.89, $\text{sem}_{\text{tra}}(\bar{x}) = 0.4$) a high average level of autonomy with three radiologists choosing a $\max(x_{\text{tra}})$ = 5. For the touch environment we also have

[2] *env*: environment (*env*) with both traditional (*tra*) and touch (*tou*) options;
χ^2_{env}: Chi-Square of the environment (*env*);
α_{env}: Alpha value of the environment (*env*);
[2] *N*: the number of users (radiologists);
sem_{env}: standard error of the mean for the environment (*env*);
M_{env}: mean value of the environment *env*;
σ_{env}: standard deviation of the environment *env*;

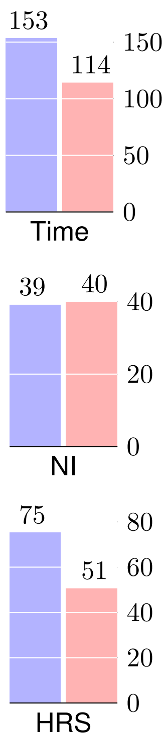

Table 3. Time Performance Mean; Number of Interactions (NI) Mean; Hit Rate Score (HRS) Mean.

values above average (M_{tou}=3.4, σ_{tou}=1.14, $\mathrm{sem}_{tou}(\bar{x}) = 0.5$), we also have a $\max(x_{tou})$ = 5 but just one radiologist scored this value.

We present the results of *Affect*, *Enjoyment* and *Intuitive Controls* (Table 5).

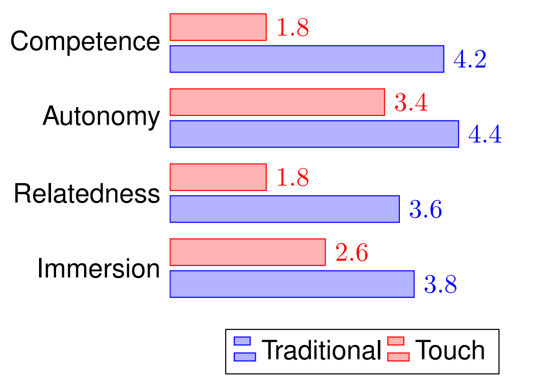

Table 4. Motivation Experience

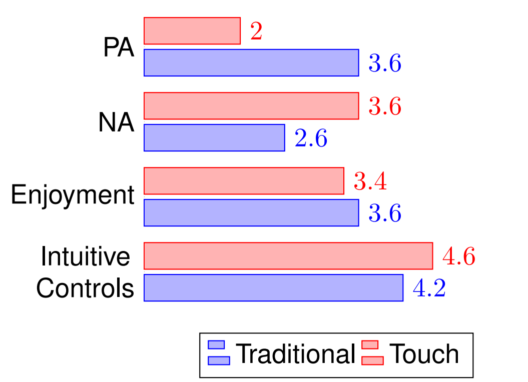

Table 5. Overall Pleasure and Enjoyment of the Task

Relatedness. Radiologists relatedness show us a touch environment below the average (M_{tou}=1.8, σ_{tou}=0.44, $\mathrm{sem}_{tou}(\bar{x}) = 0.2$), where the $\max(x_{tou})$ = 2 and the $\min(x_{tou})$ = 1. On the other hand, the traditional environment the mean (M_{tra}=3.6) is above average of the touch environment.

Immersion. For motivational experience, the immersion values of traditional environment are (M_{tra}=3.8, σ_{tra}=1.3, $\mathrm{sem}_{tra}(\bar{x}) = 0.58$). Here, we have a $\max(x_{tra})$ = 5 and the $\min(x_{tra})$ = 2, however two radiologists voted x_{tra} = 5 the mean value (M_{tra}=3.8) was not that high. The touch environment provided (M_{tou}=2.6, σ_{tou}=1.51, $\mathrm{sem}_{tou}(\bar{x}) = 0.67$).

Pleasure and Enjoyment

Affect. The affecting computing has placed an emphasis on understanding the affective and cognitive-affective responses that radiologists have to their technological interactions. Traditional environment is perceived as more positive (M_{tra}=3.6, σ_{tra}=0.54, $\mathrm{sem}_{tra}(\bar{x}) = 0.24$), whereas the marginal negative affect for traditional environment (M_{tra}=2.6, σ_{tra}=0.54, $\mathrm{sem}_{tra}(\bar{x}) = 0.24$) showed a typical correlation. The touch environment is perceived as more negative (M_{tou}=2, σ_{tou}=0.7, $\mathrm{sem}_{tou}(\bar{x}) = 0.31$) then positive ($M_{tou}$=3.6, σ_{tou}=1.14, $\mathrm{sem}_{tou}(\bar{x}) = 0.5$).

Enjoyment. Intrinsic motivation is foundations of the enjoyment of interactive experiences, and can be attributed to volition and achievement inherent in the person. The traditional environment provides (M_{tra}=3.6, σ_{tra}=1.14, $\mathrm{sem}_{tra}(\bar{x}) = 0.51$) which is almost the same as in the touch environment (M_{tou}=3.4, σ_{tou}=1.67, $\mathrm{sem}_{tou}(\bar{x}) = 0.74$).

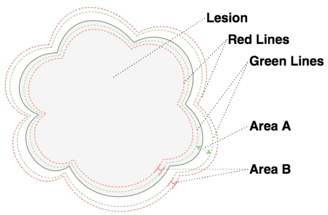

Figure 6. Annotation Classification Areas.

Intuitive Controls. Controls are intuitive when they do not interfere with one's sense of presence being easily mastered. For the intuitive controls the overall system was high rated by radiologists, however the behavior of traditional environment (M_{tra}=4.2, σ_{tra}=0.44, $\text{sem}_{tra}(\bar{x}) = 0.2$) was less 8% than touch environment (M_{tou}=4.6, σ_{tou}=0.54, $\text{sem}_{tou}(\bar{x}) = 0.24$).

Acknowledgments

This work was partially supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013 and Instituto Superior Técnico (IST-ID) through the FCT/UID/EEA/50009/2013 project, BL89/2017-IST-ID grant. We would like to thank Hospital Fernando Fonseca (HFF) for the collaboration.

References

[1] Broeck, A., Vansteenkiste, M., Witte, H., Soenens, B., and Lens, W. Capturing autonomy, competence, and relatedness at work: Construction and initial validation of the work-related basic need satisfaction scale. *Journal of Occupational and Organizational Psychology 83*, 4 (2010), 981–1002.

[2] Crisan, S., and Tarnovan, I. G. Optimization of a multi-touch sensing device for biomedical applications. In *Advanced Engineering Forum*, vol. 8, Trans Tech Publ (2013), 545–552.

[3] Doi, K. Current status and future potential of computer-aided diagnosis in medical imaging. *The British journal of radiology 78*, suppl_1 (2005), s3–s19.

[4] Doi, K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics 31*, 4 (2007), 198–211.

[5] McFarlane, D. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction 17*, 1 (2002), 63–139.

[6] Ryan, R. M. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology 43*, 3 (1982), 450.

[7] Seibel, E. J. Medical imaging, diagnosis, and therapy using a scanning single optical fiber system, Dec. 13 2005. US Patent 6,975,898.

[8] Theodorsson-Norheim, E. Kruskal-wallis test: Basic computer program to perform nonparametric one-way analysis of variance and multiple comparisons on ranks of several independent samples. *Computer methods and programs in biomedicine 23*, 1 (1986), 57–62.

[9] Watson, D., and Clark, L. A. The panas-x: Manual for the positive and negative affect schedule-expanded form.

[10] Watson, D., Hancock, M., Mandryk, R. L., and Birk, M. Deconstructing the touch experience. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*, ACM (2013), 199–208.