M 2014

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

GESTURE METRICS

OPTIMIZING THE INTERPRETATION OF FUSAMI DATA FOR USABILITY ANALYSIS

ANA CATARINA CARDOSO SALGUEIRO FERNANDES

DISSERTAÇÃO DE MESTRADO APRESENTADA
À FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO EM
MULTIMÉDIA

U.PORTO

Gesture Metrics- Optimizing the interpretation of FUSAMI data for usability analysis

Ana Catarina Cardoso Salgueiro Fernandes

Mestrado em Multimédia da Universidade do Porto

Orientador: Luís Teixeira (PhD)

Co-orientadora: Ana Correia de Barros (PhD)

Junho de 2014

Gesture Metrics- Optimizing the interpretation of FUSAMI data for usability analysis

Ana Catarina Cardoso Salgueiro Fernandes

Mestrado em Multimédia da Universidade do Porto

Aprovado em provas públicas pelo Júri:

Presidente: José Miguel Santos Araújo Carvalhais Fonseca (PhD)

Vogal Externo: Paula Alexandra Gomes da Silva (PhD)

Orientador: Luís Filipe Pinto de Almeida Teixeira (PhD)

Abstract

There is plenty of documentation and instructions on how to use and interpret data recorded through eye-tracking equipment: eye-movement metrics. However there is still lack of information on metrics associated with gestures. The main purpose of this study is to extract meaning from gestures and define new gesture metrics for usability evaluation and inference of users' state and behaviour, using as reference the existing metrics of eye tracking. Frustration was chosen as the first user state to study because i) it is an important state in the field of human-computer interaction studies, and ii) because it is relatively easy to induce and there is already significant related work on inducing and measuring frustrating episodes. The intention of this work was twofold: first, to induce the state of user frustration and then to collect data about users' gestures under frustration. The methodology adopted is based on mixed-methods approach drawing from experimental procedures, to collect quantitative and qualitative data. Tests were performed with seventeen participants who rated their emotional state before and after the test, and performed several tasks using an application with some intentional bugs and problems to induce the frustrating episodes. The retrospective think aloud method was used after the test to help collect qualitative data about the user interaction. The results show that the most meaningful measure for the study was the number of gestures performed which shows significant variations according to the nature of the tasks and the level of frustration reported by the participants. The decrease of time between gestures can be an indicator of impatience and irritation of the user, indicating that he/she has encountered a problem or obstacle during their interaction. The analysis of the navigation patterns suggests that in the easy and intuitive tasks, users behave in a similar way with slight differences in their interaction and, in the frustrating tasks, the patterns are smaller and fragmented, suggesting that there was not a common behaviour during navigation. The frequency measure shows no conclusive results.

Contents

List of Figures

List of Tables

Abbreviations

AICOS	Assistive Information and Communication Solutions
FUSAMI	Fraunhofer Usage Mining
HCI	Human Computer Interaction

Chapter 1

Introduction

There is plenty of documentation and instructions on how to use and interpret data recorded through eye-tracking equipment: eye-movement metrics. These help researchers analyse the user´s interaction data and extract meaning from users' actions (e.g. long fixations, fixation special density, gaze, saccades, scan paths, transition matrixes…).

The proliferation of mobile devices gave rise to different tools to record users' interaction with the applications. Rather than relying on eye-gaze, these tools rely mostly on the record of users' gestures. Alas, all these new tools still lack studies to provide support in interpreting the data, and researchers do not yet know what specific sequences or patterns of gestures mean.

The goal of this study is to combine the use of FUSAMI (a web-based platform to perform advanced analytics on real-time mobile applications usage data[1]) with qualitative research to extract meaning from gesture patterns. The first user state that we chose to study was frustration because it is an important state in the field of human-computer interaction studies. Frustration with technology is one of the major causes that lead people to hesitate or avoid using the computer.

1.1 Objectives and expected results

It is hypothesized that 1) touch gesture patterns can be identified from logging users' interaction and that 2) these patterns may relate to certain behaviours of users, as is the case nowadays with eye-tracking. The main objective of this study is to identify new gesture metrics to help interpret gesture-related data on mobile devices. Due to time constraints for this study, we chose to evaluate only one state of the user, the frustration state. To achieve the expected

[1] http://fusami.projects.fraunhofer.pt/

objectives several experiments with different users were performed and the results of those tests analysed. This set of experiments intends to answer two questions:

- Can user frustration with a smartphone be induced?
- Are there gesture patterns that can be associated with user frustration?

As a result we expected, after inducing frustration, to find some differences in the gesture patterns that would allow us to identify this state.

1.2 Methodology

The methodology adopted for this study is based on a mixed-methods approach drawing from experimental procedures. The work is organized in two phases, the first of which constitutes a literature review on eye-tracking, remote usability testing, heat maps and associated metrics, gesture analysis and interpretation, and users' behaviour models. During the second phase the user tests were performed. The results were analysed in order to identify gesture patterns, gesture metrics and associated meaning.

First phase:

- Literature review on eye-tracking, remote usability testing, heat maps and associated metrics;
- Literature review on gesture analysis and interpretation.

Second phase:

- Design and conduct a set of experiments using FUSAMI and Think Aloud protocols/Contextual Inquiry;
- Identify gesture patterns, gesture metrics and associated meaning;

1.3 Dissertation structure

The dissertation is organized in 5 chapters:

- This first chapter introduces the context of this dissertation, objectives, and expected results and methodology.

- The second chapter describes the affective states that the user can experience, then defines gestures and their main characteristics and identifies the gestures commonly used in interaction with mobile devices, and finally explains the eye-tracking metrics and visualization.
- The third chapter characterizes the state of frustration, and presents different studies related to this state.
- The fourth chapter presents the objectives, participants, procedure and results for the tests with the users and discusses the main findings.
- Finally, the fifth chapter presents and overview of the results of this research, and future work to be developed.

Chapter 2

State of the art

This chapter presents the state of art. The objective of this study is to identify new gesture metrics based on the interpretation of gesture-related data on touch enabled mobile devices to help researchers understand users' behaviour. Since the main purpose of the study is to understand the user´s behaviour, the first section describes attentional and affective states that the user can experience when testing and using an application. This is relevant because by detecting the user's state it is possible to have a better understanding of his/her behaviour. The data that will be used to study the user´s behaviour will be extracted from gestures, so it is important to understand the definition and the main characteristics of a gesture. Therefore, the second section is focused on gestures and their importance in human-computer interaction. The third and last section presents information about eye tracking, a widely used technique to study usability. In order to identify new gesture metrics, it is important to understand the metrics that are already used to study usability issues. For this reason, the last chapter presents the metrics used in eye tracking techniques, their interpretation and the visualization of the data extracted.

2.1 Behavioural state of the user

The analysis of user behaviour is an important field of study in HCI (Hudlicka, 2003; Picard, 1999; Scherer et al., 2012). Knowing the user's behaviour and its associated meaning can be a key feature in usability research. It is important to recognize the reaction and the feedback of the user when interacting with an application, since his/her behavioural state may contain significant and relevant information on the application in question. Understanding this behaviour can help minimize the cognitive load and maximize usability, reducing the user's mental effort when performing tasks and achieving his/her intended goals (Whitenton, 2013).

State of the art

Despite the difficulties in analysing user behaviour in the real world, it is possible to estimate the user state using body and hand movements, eye gaze or facial expressions (Shaker, Asteriadis, Yannakakis, & Karpouzis, 2013). These techniques are used to extract data from the physical behaviour of the user and establish a connection between his/her actions and his/her state. Most studies use visual material (e.g. face recognition) or speech signals. Some studies look at bi-modal-information, combining visual and audio, in order to get closer to the human sensory system. In face to face interactions, humans rely on voice and facial expressions to interpret non-verbal communicative signals in terms of affective states expressed (Pantic & Rothkrantz, 2003). The bi-modal approach in HCI tries to simulate this kind of communication to recognize the user´s emotional state. There are no studies, to our knowledge, that try to infer users' behavioural states through gestural interaction with a touchscreen.

First, it is important to define what an affective state is and how it affects the user's behaviour. The affective state is closely connected to emotions. One well-known way of organising emotions is into basic ones (e.g., happiness, anger, sadness, surprise, disgust, fear), that can be recognized cross culturally. Some researchers claim, however, that emotions cannot be defined in terms of a small set of emotion categories; instead they must be characterized in terms of a multidimensional affect space. The definition of emotion is not consensual. According to Hudlicka (2003), an affective state is a range of conditions including basic emotions (happiness, anger, sadness, surprise, disgust, fear), complex emotions (shame, guilt, jealousy) and simple bi-polar reactions, such as like and dislike, boredom and excitement. These states can be dependent both on the individual, his/her temperament or individual history, and on the situational context.

Although essential to individual and social development, the study of emotions became a field of interest in HCI, due to the discovery that emotions play a critical role in the rational aspects of behaviour, such as perception, decision-making, learning, planning and action selection (Hudlicka, 2003). This recent research has identified affective states as crucial factors influencing decision making and performance, regarding the nature of these influences on perceptual, cognitive, and motor processes. These influences occur both at lower levels of processing (attention orientation, working memory), and at higher levels involving goals and expectations. (Hudlicka, 2001). Therefore, it is possible to conclude that emotion influences cognitive processing and plays a central role in the control of behaviour.

Many studies have been done in the field of affect or emotion recognition (Chen, 2000; Zeng et al., 2007; Zeng, Pantic, Roisman, & Huang, 2009). In the aforementioned studies, the authors recognize six basic emotions: fear, anger, sadness, happiness, disgust and surprise. They also identify four non basic affective states focused on HCI contexts, denominated by HCI related affects: interest, boredom, confusion, frustration. According to Zeng et al., (2007, 2009), in HCI research these states are the most relevant because they indicate the user's cognitive/motivational state. In the context of learning and tutoring, Lehman, Matthews, D'Mello, & Person (2008) focused their study not only on the basic emotions, but in learning-centred emotions like confusion, frustration, anxious, contempt, *eureka* and curiosity.

As seen before, the nature and the purpose of the research influences the choice of the affective states to be analysed. Different states are selected regarding the topic of study. The present study aims to define new gesture-metrics for usability evaluation, therefore it is crucial to select the most relevant states for this purpose. Since this is focused on usability issues, it is important to take into account the effectiveness, efficiency and satisfaction with which users achieve specified goals in particular environments. Effectiveness measures the accuracy and completeness of a task performance and the efficiency measures the resources expended in relation to the accuracy. To analyse this two measures it is relevant to know the user´s cognitive effort performing the task. The states confused and attentive can provide information about that. If the user is confused, it can indicate poor comprehension or problems in the respective task. On the other hand, the attentive state can indicate that the user is focused and engaged in the task at hand and is not experiencing any problems. Satisfaction measures the comfort and acceptability of the system and can be analysed by the frustrated and confident states. If the user is frustrated during or after performing a task, it can indicate that he/she is not pleased with the system. Although, if he/she is confident, that can show that he/she is pleased because he/she accomplished all tasks successfully.

The *interested* state occurs when the user is making a remarkable effort to interact and is focused in his/her tasks (Asteriadis, Tzouveli, Karpouzis, & Kollias, 2009; Scherer et al., 2012). Visual evidences such as eyes looking at the screen or eyes wide open for a long period of time can be extracted from face recognition in order to identify this state. Using the eye tracking techniques, it is possible to recognize the same state using metrics like fixations or saccades. For example, longer fixations in a particular area indicate that this might be an important area and fixations concentrated in a small area indicate focused and efficient search. In section 3.1, eye tracking metrics are explained in further detail. The *confused* state is recognized when the user is perplexed by conflicting situations or statements, or is having difficulties during the activities (Picard, 1999; Zeng et al., 2004). This state can be evidenced by examining the saccades and scan paths during the eye tracking test. It is possible to conclude that regressive saccades indicate the presence of less meaningful clues, long-lasting scan paths indicate less efficient scanning, and longer scan paths indicate less efficient searching. *Frustration* can be identified when the user experiences the feeling of being upset or annoyed as a result of being unable to achieve something. Visual signs like eyes not looking at the screen or eyes blinking, head moving or frowning at the screen can be evidence of frustration. Longer fixation durations can be understood as difficulty of information extraction and interpretation and more overall fixations indicate less efficient search. *Confident* is related to belief in being able to perform the task at hand correctly.

2.2 Mobile and surface gesture

2.2.1 Gesture characterization

The definitions of gestures are particularly associated with the communicational aspect of the human hand and body movements. Gestures can provide a unique link between actions and thoughts. Gestures form an integrated system with speech, as listeners are more likely to understand the message conveyed in speech if it is accompanied by a gesture (Cartmill, Beilock, & Goldin-Meadow, 2012). To begin characterizing the gesture it is essential to distinguish it from unintentional movements. Unlike gestures, unintentional movements do not carry any meaningful information. The gesture is closely connected with speech, it has meaning and always conveys a thought or an intention (van den Hoven & Mazalek, 2011). After this first characterization, gestures can be classified into two distinct categories: manipulative and communicative gestures. Manipulative gestures are the ones used to act upon objects in an environment, for example object movement or rotation. Communicative gestures are usually accompanied by speech and have an inherent communicational purpose. They can be either acts or symbols. Symbols have a linguistic role – they represent some referential action or are used as modellers, often of speech. Acts are gestures that are related to the interpretation of the movement itself, for example imitating some actions or pointing acts (Pavlovic, Sharma, & Huang, 1997; van den Hoven & Mazalek, 2011). According to Efron (1941), human gestures are a dynamic process, and it is possible to identify three phases based on the temporal characteristics of gestures. The first phase is called preparation phase and describes the preparatory movement in which hands are moved to the location where the gesture will take place. The next phase, the stroke phase, occurs when the actual gesture is performed. Finally, the retraction phase takes place when the gesture is finished and the hands move away from the gesture location and return to the resting position (Pavlovic et al., 1997; van den Hoven & Mazalek, 2011).

2.2.2 Gesture in HCI

In HCI, a gesture can be used as a mean of communication between the user and the computer. Gestures are useful for computer interaction since they are the most primary and expressive form of human communication (Pavlovic et al., 1997). Despite their communication purpose, in the HCI context, these gestures are manipulative because they communicate with the machine by giving it instructions. Currently, gesture detection, recognition or tracking are topics that are often covered in HCI studies. Gesture recognition allows humans to interact with computers.(Yousefi, Abedan Kondori, & Li, 2013). In the HCI context, gesture recognition can provide an intuitive and more convenient way of interaction, since the use of gestures is a natural mean of communication between humans (Panwar & Mehra, 2011).

2.2.2.1 Gesture recognition

To explore the use of gestures in HCI, movements from hand, arm, and even other parts of the human body, have to be measured and interpreted by the machine (Pavlovic et al., 1997). Gesture recognition is a technique that allows the identification and recognition of gestures originated from body movements, from face or hands. Gestures made by the user are used to send the information to the machine (Sonam & Ubale, n.d.). In the article previously mentioned the authors describe the gesture recognition as being composed of several stages:

- Data Acquisition: In this step the input data, the gestures, are collected;
- Gesture Modelling: This stage employs the fitting and fusing of the input gesture into the model used;
- Feature Extraction: Extract data from the gestures input. These features can be hand/palm/fingertips location, stroke size and angles, etc.;
- Recognition State: After the analysis of the gesture, the recognition state declares a command or a meaning for the gesture in order to complete the communication purpose.

2.2.2.2 Touch analytics

“Each user interaction behaviour on touchscreen can be quite unique” (Frank, Biedert, Ma, Martinovic, & Song, 2013, p.2). Current research explores the additional sensory information that an interactive screen can provide. The extraction of biometrics from keystroke dynamics for authentication has been researched as an alternative of traditional authentication schemes based on passwords. The main idea of this new type of authentication is to monitor the user’s interaction with the device and estimate his/her behaviour, in order to identify him/her using only the biometrical information from the touch. Frank and co-workers (2013) conducted several experiments in which the phones recorded the users’ touch data. The features extracted from the recorded data were divided into individual strokes. The stroke begins with touching the screen and ends with lifting the finger. Each stroke has its own features, such as location, time stamp, pressure on the screen, area occluded by the finger, the orientation of the finger, and the orientation of the phone. The analysis and interpretation of these biometrics aim to study and understand the user’s behaviour during the interaction using data extracted from the gestures performed on the screen.

Another tool that performs analytics on real-time usage data collected in smartphones is *Fraunhofer Usage Mining* (FUSAMI). The FUSAMI system provides information from real user interaction when testing an application, recording the information when the user clicks in an application screen so that people analysing the results can understand the user's behaviour and find out more about usability issues. In order to accomplish that, FUSAMI offers a set of different analytical tools, composed of screenshots of the real application. Heat maps are one of the tools provided by the FUSAMI system and they can be generated by screen or widget. Heat map visualization can show the click frequency to indicate which screens are often used and which are not, and it also allows to visualize the time spent in each screen and the latency between widget events. Sequential patterns and transitions are another type of tools provided by this system. This analytical tools allows the system to describe frequent usage patterns in order to define how a user navigates through an application. FUSAMI also provides analytical tools for gesture analysis. One of these tools is the heat map resulting from the gesture interaction with the application, which reveals the screen areas where the user touched more often. A set of common gestures displayed in each screen is another tool available in the screen. This aggregates the raw data of the gestures performed and generates a pattern of the most common gesture location and orientation.

FUSAMI was the tool chosen for this study because, besides being a program developed at Fraunhofer AICOS, other existing solutions do not allow for a detailed analysis of gestures. Other programs available for gesture analysis aggregate multiple gestures into a single image, marginalizing the dynamic aspects of the interaction. On the other hand, FUSAMI allows a more detailed analysis of gestures, which can be helpful to understand whether or not users are able to perform a given gesture, whether or not they need some attempts, their level of accuracy or how they have explored the application (Schaefers, Ribeiro, & de Barros, 2013).

2.2.3 Commercially available gesture sets

In order to interpret data from gesture recording, such as using FUSAMI, it is important to understand what type of gestures there are to interact with mobile devices and what they are used for in different operating systems.

The most common operating systems used in mobile devices are iOS, Android and Windows Phone. In the table below are identified the most relevant gestures in these operating systems, the physical action that is performed to complete the gesture and the resulting action in each system.

Gesture	Action	iOS[2]	Android[3]	WindowsPhone8[4]
Tap	Press, lift	To press or select a control or item.	Triggers the default functionality for a given item.	Opens or launches whatever you tap.
Double Tap	Two touches in quick succession	To zoom in and centre a block of content or an image. To zoom out (if already zoomed in).	Scales up a standard amount around the target with each repeated gesture until reaching maximum scale. For nested views, scales up the smallest targetable view, or returns it to its original scale. Also used as a secondary gesture for text selection.	Zooms in or out in stages.
Swipe/Drag	Press, move, lift	To scroll or pan—that is, move side to side. To drag an element.	Scrolls overflowing content, or navigates between views in the same hierarchy. Swipes are quick and affect the screen even after the finger is picked up. Drags are slower and more precise, and the screen stops responding when the finger is picked up.	Moves through screens or menus at a controlled rate
Tap and Hold	Press, wait, lift	In editable or selectable text, to display a magnified view for cursor positioning.	Enters data selection mode. Allows you to select one or more items in a view and act upon the data using a contextual action bar. Avoid using long press for showing contextual menus.	Opens a context-specific menu (like right-clicking with a mouse).
Pinch / Spread	2-finger press, move outwards or inwards, lift	Pinch open to zoom in; pinch close to zoom out.	Zoom into or out of contents	Gradually zooms out or in gradually on a map, webpage, or picture.
Flick	Swipe your finger quickly in the direction you want the screen to move.	To scroll or pan quickly.	To scroll or pan quickly.	Scrolls rapidly through menus or pages, or moves sideways in Hubs.

Table 1- List of gestures and consequent action for each operating system (Leitão, 2013)

2.3 Eye tracking

Eye tracking is a useful tool for studying usability issues in an HCI context. It is a technique where eye movements are measured in order to evaluate the user behaviour. For usability analysis, the user's eye movements while using the system are recorded and later analysed. From the data collected the researches can know where the user is looking at, at a given time, how much time he/she spent looking at something and the sequence in which the eyes shifted from one location to another (Poole & Ball, 2005). Eye tracking is an effective method for tracking the user's behaviour because eye movements recordings provide a trace of where the person's attention is being directed (Poole & Ball, 2005).

---

[2] https://developer.apple.com/library/ios/documentation/UserExperience/Conceptual/MobileHIG/MobileHIG.pdf
[3] http://developer.android.com/design/patterns/gestures.html
[4] http://www.windowsphone.com/en-us/how-to/wp8/start/gestures-flick-pan-and-stretch

2.3.1 Eye movement metrics

2.3.1.1 Fixations

Fixations occur when the eyes are relatively stationary, taking or encoding information (Poole & Ball, 2005). Fixations can have different interpretations depending on the context. In an encoding task, higher fixation frequency on a particular area can be indicative of a greater interest or it can be a sign that the target is complex and more difficult to encode. However, in a search task a higher number of single fixations is often a manifestation of grater uncertainty in recognizing a target item (Jacob & Karn, 2003). Some metrics can be derived from fixation:

- Overall number of fixations: According to Goldberg & Kotval (1999) (cited in Jacob & Karn, 2003; Poole & Ball, 2005), more overall fixations indicate less efficient search. A larger number of fixations indicates less efficient search possibly resulting from a poor arrangement of display elements.

- Fixations per area of interest: Poole et al. (2004) (cited in Poole & Ball, 2005) affirm that more fixations on a particular area indicate that it is more noticeable, or more important, to the viewer than other areas.

- Fixation duration: For Just & Carpenter (1976) (cited in Poole & Ball, 2005), a longer fixation duration indicates difficulty in extracting information, or it means that the object is more engaging in some way.

- Gaze: As described by Mello-Thoms et al. (2004) and Hauland (2003) (cited in Jacob & Karn, 2003; Poole & Ball, 2005), gaze is usually the sum of all fixation durations within a prescribed area. The proportion of time looking at a particular display element could reflect the importance of that element. It is best used to compare attention distributed between targets. It can also be used as a measure of anticipation in situation awareness if longer gazes fall on an area of interest before a possible event occurs.

- Fixation spatial density: According to Cowen et al.(2002) (cited in Poole & Ball, 2005), fixations concentrated in a small area indicate focused and efficient searching. Evenly spread fixations reflect widespread and inefficient search.

- Repeated fixations: Goldberg & Kotval (1999) concluded that higher numbers of fixations off-target after the target has been fixated indicate that it lacks meaningfulness or visibility (cited in Poole & Ball, 2005).

2.3.1.2 Saccades

A saccade is an eye movement occurring between fixations. The purpose of most saccades is to move the eyes to the next viewing position (Poole & Ball, 2005). Saccades cannot evaluate complexity or salience of an object in the user interface because no encoding takes place during them. However, regressive saccades can be used to measure difficulty during encoding. Saccade-derived metrics are explained below:

- Number of saccades: Goldberg & Kotval (1999) concluded that more saccades indicate more searching (cited in Poole & Ball, 2005).

- Saccade amplitude: According to Goldberg et al. (2002) (cited in Poole & Ball, 2005), considering the saccade amplitude it is possible to conclude that larger saccades indicate more meaningful cues.

- Regressive saccades: As described by Sibert et al. (2000) (cited in Poole & Ball, 2005), regressions are used to measure difficulty and can indicate the presence of less meaningful cues.

- Saccades revealing marked directional shifts: Cowen et al.(2002) (cited in Poole & Ball, 2005) affirmed that any saccade larger than 90 degrees from the previous saccade shows a rapid change in direction, indicating that the user’s goals have changed or the user interface layout does not match the user’s expectations.

2.3.1.3 Scan path

The scan path is an eye tracking metric, usually composed by a complete sequence of fixations and interconnecting saccades (Poole & Ball, 2005). This metric can indicate the efficiency of the arrangement of elements in the user interface (Jacob & Karn, 2003). The following metrics are derived from scan path:

- Scan path duration: As described by Goldberg & Kotval (1999), a longer-lasting scan path indicates less efficient scanning (cited in Poole & Ball, 2005).

- Scan path length: For Goldberg et al. (2002) (cited in Poole & Ball, 2005), a longer scan path indicates less efficient searching.

- Spatial density: Goldberg & Kotval (1999) (cited in Poole & Ball, 2005) concluded that smaller spatial density indicates more direct search.

- Transition matrix: According to Goldberg & Kotval (1999), the transition matrix reveals search order in terms of transitions from one area to another (cited in Poole & Ball, 2005).

- Scan path regularity: For Goldberg & Kotval (1999) (cited Poole & Ball, 2005), after defining the "cyclic scanning behaviour", deviation from a normal scan path can indicate search problems due to lack of user training or user interface layout failures .

- Scan path direction: Altonen et al. (1998) (cited Poole & Ball, 2005), affirmed that the scan path direction can determine a user´s search strategy regarding menus, lists and other user interface elements.

- Saccade/fixation ratio: Goldberg & Kotval (1999) (cited in Poole & Ball, 2005) concluded that higher ratio of the comparison between time spent searching (saccades) and time spent processing (fixations) indicates more processing or less searching.

2.3.2 Eye tracking visualization

2.3.2.1 Heat maps

Heat maps (Figure 1) are the most popular visualization technique to present eye tracking data (Tsang, Tory, & Swindells, 2010). They are composed by semi-transparent, multi-coloured layers that cover areas of higher attention with warmer colours and areas of less attention with cooler colours (Blignaut, 2010). This kind of visualization is commonly used to provide a qualitative impression of the distribution of user's attention over the display space (Andrienko, Andrienko, Burch, & Weiskopf, 2012; Kurzhals & Weiskopf, 2013). Heat maps can be created using the fixation length, the time people spend looking at that area, or the fixation count, the number of fixations targeted at that area (Andrienko et al., 2012; Pernice & Nielsen, 2009).

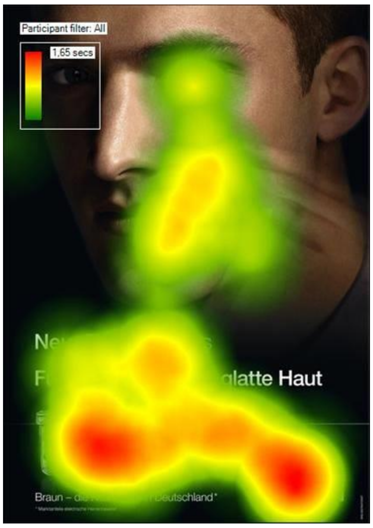

Figure 1- Heat map example
[source: http://eyetracking.com.ua/eng/]

2.3.2.2 Gaze plots

Another visualization technique is the gaze plot (Figure 2). Gaze plots display movement sequence, order and duration of gaze fixation. This technique captures the sequence of eye fixations using ordered circles where circle size is proportional to duration and the number indicates the order and connects consecutive fixations by lines (Tsang et al., 2010). This method is not suitable for large data due to enormous over plotting (Andrienko et al., 2012).

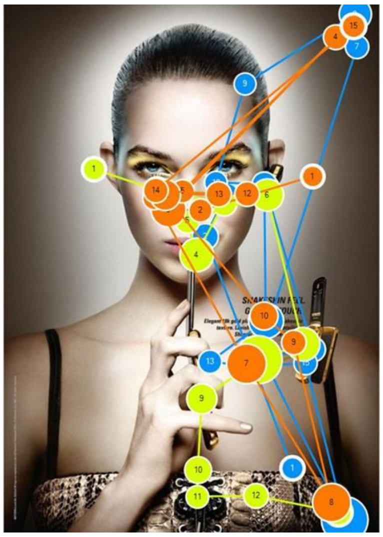

Figure 2- Gaze plot example
[source: http://eyetracking.com.ua/eng/]

2.3.2.3 Areas of interest

Defining areas of interest (Figure 3) and counting the number of fixations in each one is another useful feature in eye tracking data visualization for doing quantitative analysis (Pernice & Nielsen, 2009). The first stage is to choose the areas that are significant to analyse with an appropriate metric, then the system compares and calculates how often and how long these items were looked at, and gets some statistics. A limitation found in this approach is the fact that it neglects sequence information, the order in which items of interest were visited, and therefore only provides a summary of the content viewed (Tsang et al., 2010).

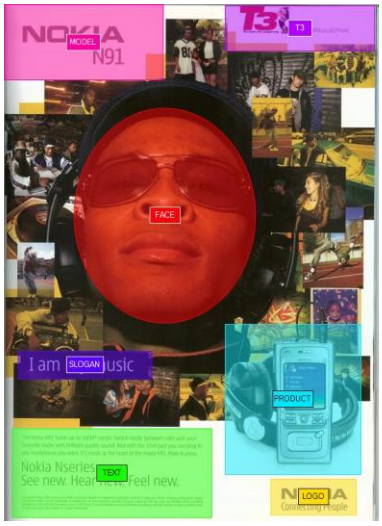

Figure 3- Area of interest example
[source: http://eyetracking.com.ua/eng/]

Chapter 3

Frustration state

“Even with current usability practices, computer interfaces continue to induce negative states on human users”(Mentis & Gay, 2002, p. 406). States like confusion, anxiety and frustration may impede productivity, creativity and cognitive capacity and can interfere with the interaction between humans and technology (Klein, Moon, & Picard, 2002). Since frustration is, as previously described, one of the most important states in usability studies, this dissertation is focused in this state. This chapter presents the definition and other associated topics related to frustration. Firstly, the emotional state of frustration is defined, then are presented the major causes of frustration with technology. And in the last section, some previous work related to user frustration is described.

3.1 Definition of frustration

Frustration is an emotional state resulting from the occurrence of an obstacle that prevents the user to achieve his/her goals (Bessiere, Ceaparu, Lazar, Robinson, & Shneiderman, 2003; Klein et al., 2002). The user experiences the feeling of being upset or annoyed as a result of being unable to change or achieve something. The level of frustration experienced by an individual can be different, depending on the circumstances surrounding the frustrating experience and on the individuals themselves. One important factor related to frustration is the goal commitment that has a strong relationship to performance and can depend on two factors: the importance of the task and the belief that the goal can be accomplished (Bessiere et al., 2003). The more the importance of the goal increases, the more committed the individual is, affecting the strength of the reaction to the interruption of the goal (Lazar, Jones, Bessiere, Ceaparu, & Shneiderman, 2005). According to Bessiere et al., (2003), the intensity of the frustration felt depends on several factors which are grouped into two categories: the incident and the individual factors. The incident factors are based in goal theory and include the level of goal commitment, the severity of the interruption and the strength of the desire to attain the goal. The goal commitment and the strength of desire for the goal are measured in terms of the importance of the task to the user. The severity of the interruption is measured as the amount of time spending fixing the problem combined with the time lost due to the problem. Individual factors are the elements related to the user, including technology experience variables, mood, psychological factors and cultural influences (Bessiere et al., 2003).

3.2 Technology frustration

"Frustration is a common theme among computer users. As technology rapidly advances, we the users must deal with the ensuing error messages that invariably result, as well as annoying delays, incompatible files, and indecipherable menus" (Bessiere et al., 2003, p.2). Frustration with technology is said to be one of the major causes that lead people to hesitate or avoid using technology to reach their goals (Lazar et al., 2005). The level of frustration faced by the users depends on several factors, such as time loss, importance of the goal and the user´s self-efficacy and his/her previous experience. The most common causes related to user frustration include hardware or software failures, poor user interface design or the user´s lack of experience (Bessiere et al., 2003). In a study conducted by Mentis (2007), using memory as indication of where frustration occurs, the majority of the frustrating incidents that users recalled were incorrect auto-formatting, computer error or bugs, slow or dropped internet connections and unwanted pop-ups (Shneiderman, Plaisant, Cohen, Jacobs, & more, 2009). Another study, which attempted to measure the frequency, cause, and the level of severity of frustrating experiences, determined that the most frequent frustration causes cited by the users were confusing error messages, timed out/dropped/refused connections, application crashes, long download time, long system response time and missing/hard to find features (Ceaparu, Lazar, Bessiere, Robinson, & Shneiderman, 2004).

3.3 Previous work

This section presents some related work to this dissertation that might serve as a comparison and influence on this study. There are several studies that attempt to identify and measure user frustration, using different types of sources to collect this information.

In their work, Scheirer, Fernandez, Klein, & Picard (2002) used a deliberately slow computer-game user interface to induce frustration and tried to developed a strategy for coupling the data collected with real word events. They used two physiological sensors, one to measure galvanic skin response (GSR) and another for blood volume pressure (BVP), as input to understand the users' emotional state, and combine these physiological measures with another variable, the mouse click behaviour. With the aim of inducing the state of frustration in the user, they created different puzzle games, and the task was to click on the right figure of the puzzle in order to advance to the next screen. The game was timed, scored and worked as a competition with a final prize for the winner, to incentive the participants. At irregular intervals of time, a delay occurred during which the mouse appeared not to work, prevented the user to achieve his/her goal, inducing frustration with the intentional delay. The authors analysed the variables throughout the periods following mouse failures and compared them to episodes of normal activity (no mouse failures). A pattern-recognition strategy was applied to the physiological signs to test whether or not frustration could be automatically discriminated. Regarding the mouse click behaviour variable, they computed the number of mouse-clicks following each frustrating episode, and their results revealed four different types of behavioural responses. The first type was a person who usually just waited without clicking, occasionally clicked one extra time and rarely clicked more than that. In the next two types, this behaviour shifted to a higher number of clicks. And in the last type, the user always made superfluous clicks, usually many of them. The data collected with the sensors and the mouse behaviour were combined, in order to create models to predict frustration (Scheirer et al., 2002a).

Another work concerned with the automatic prediction of frustration was conducted by (Kapoor, Burleson, & Picard, 2007). The main goal of this study was to create an automated method to predict frustration and use the information gathered to provide support to the user in a learning tutor system. The main focus of this work was to address the problem of recognizing the state in which a child who begins a problem-solving activity on the computer is. The activity chosen was solving the Towers of Hanoi. To collect the data, the authors used an environment with sensors that were able to measure video from the face, postural movement from the chair, skin conductance from the wireless sensor on the non-dominant hand and pressure applied to the mouse. They used users' self-labelling as an indication of being frustrated and collected their behavioural data leading up to that. The users were presented with two different buttons, labelled "I´m frustrated" or "I need some help" and they could choose to ignore these buttons or click one of them. If a person clicked in the "I'm frustrated button", they labelled the segment leading up to the click as frustration, discriminating two classes (frustrated vs. others). The pressure mouse data were processed to obtain the features activity, skew, mean and variance. From the camera that analysed the face, they obtained the following features: presence of head nod, presence of head shake, probability of fidget, probability of smile, presence of blink, head velocity and head tilt. With the posture analysis seat data they obtained the ratio of forward to back posture feature and activity level feature. After the analysing the data, they could demonstrated assessment of a kind of "pre-frustration" using only non-verbal cues. They concluded that, while these signs alone did not provide perfect classification of behaviour, they significantly outperformed a random classifier (Kapoor et al., 2007).

Another highly relevant work in the study of frustration, conducted by (Dennerlein, Becker, Johnson, Reynolds, & Picard, 2003), aimed to demonstrate that there is aims to demonstrate there is a link between frustration and physical measures. They quantified physical measures of upper extremity stress, force, posture, and muscle activity. The test subjects were asked to complete a web-based survey with five pages. The page with the largest number of questions was designed to reset and delete the answers to the questions in an attempt to frustrate the user. After answering all questions in the page and pressing the submit button, an error message appeared, informing the user that he/she had completed an erroneous field incorrectly and had to go back to correct it. After that, the page would reset and delete all the previous responses, in order to force the user to fill in that page again. After completing the task, the participants completed a questionnaire evaluating the performance and usability of the web-survey. They recorded the electromyographic signals from seven muscles, the wrist, arm and shoulder posture, and the forces applied to the side and button of the mouse were measured with a custom design sensing mouse. The test group was divided into two different groups, according to the answers to the usability questionnaire: a high response group who had higher scores indicating more dissatisfaction and a low response group who had lower scores indicating less dissatisfaction. For the high response group, forces applied to the mouse increased significantly and muscle activity of wrist extensor muscles also increased after the resetting of the page, but no significant trends were observed for the low response group.

As seen in the works described above, mouse pressure is an important variable when studying frustrating episodes. Different studies attempted to establish a connection between pressure and user's frustration. One example is the work done by (Qi, Reynolds, & Picard, 2001), based on a hypothesis that subjects tend to apply excessive pressure to the mouse after encountering frustrating events. In order to test their hypothesis, they mounted eight pressure sensors on a mouse and collected pressure signals from the subjects who filled in web forms containing usability bugs. They divided the data gathered into two regions: mouse pressure where the task is proceeding smoothly and mouse pressure following a usability bug. In the tests, the subject was asked to fill a multiple-page web-form, and to increase the level of frustration induced, time was made salient to the subjects by emphasizing that their timing was very important. To induce frustration, while filling in the web form, an error message alerted the users to the fact that some information was incorrect, and forced the users to fill in that page again. With their results, the authors observed that refilling the web-form is strongly correlated with a pattern change in the mouse pressure signals (Qi et al., 2001).

The following two works used a touchpad to collect the pressure data and associate it with user negative behaviour. In the work by Mentis & Gay (2002), the main goal was to study behavioural signals as indicators of user affect. They chose the touchpad pressure as the behavioural measure because users would not be aware of the monitoring. During the test, each subject was asked to complete a set of tasks using Microsoft Word. Some tasks were chosen due to their moderate difficulty and the possibility of causing the user to make errors. After completing the tasks, the subjects were asked by the experimenter what they thought of the tasks and whether they were frustrated by any of the tasks. Each user's finger pressure on the touchpad before and after a frustrating incident was compared with a t-test. From the analysis of the results, they found a difference in the mean pressure levels before and after the frustrating incidents (Mentis & Gay, 2002). The other study addressing touchpad pressure as a variable to measure frustration was conducted by McLaughlin, Chen, Park, Zhu, & Yoon (2004) and aimed to examine the utility of pressure and positional data to discriminate between baseline and frustrating user circumstances. The experimental tasks were to assemble Lego pieces following a set of pictures provided on a laptop. The first task was a baseline task – the instructions manual contained all the information need for completing the tasks including pictured steps, so the users could easily finish the task if they followed all the steps. The second task was designed to induce frustration – the given instructions manual was missing two crucial steps and two other steps were presented twice, making the task more difficult to the user. Data from the touchpad was captured and logged; in addition, the subjects' activity was videotaped. Based on their initial observations, the authors concluded that the pressure data was more important than the positional data, therefore they emphasized the pressure data in their analyses (McLaughlin et al., 2004).

Chapter 4

Implementation, tests and results

This chapter describes the experimental methodology and procedures used on the tests with the users. Based on the works described in the previous chapter, a set of experiments was designed and conducted with the main objective of inducing frustration in the user, collecting the data from the gestures performed when interacting with a mobile application and analysing whether or not there was a relation between the users' affective state and the gestures they performed.

4.1 Objectives

The main objective of this study is to combine the use of FUSAMI (a web-based platform to perform advanced analytics on real-time mobile applications usage data) with qualitative research to extract meaning from gesture patterns and help define new gesture-metrics for usability evaluation patterns. This set of experiments intends to answer two questions:

- Can user frustration with a smartphone be induced?
- Are there gesture patterns that can be associated with user frustration?

To answer the first question, an Android mobile application was developed, with intentional bugs and other problems deliberately created. These bugs and problems were devised based on the analysis of the literature regarding what constitutes a frustrating event in HCI. The answer to the second question is found by analysing the data collected during and after user interaction with the application.

4.2 Measures

In order to extract relevant information from gestures, we selected a set of measures that we wanted to identify and analyse. The selection of these measures is based on the previous works performed to measure and identify frustration. In these works, the researchers chose measures such as physiological signals (Dennerlein et al., 2003; Kapoor et al., 2007; Scheirer, Fernandez, Klein, & Picard, 2002b), the number of clicks (Scheirer et al., 2002), the pressure applied to the mouse or touchpad (Dennerlein et al., 2003; Kapoor et al., 2007; McLaughlin et al., 2004; Mentis & Gay, 2002; Qi et al., 2001). As seen before, the most common measures used to identify frustration are the pressure applied on the mouse or on the screen and physiological signals. For this study it was not possible to measure these two variables, because we did not have the appropriate sensors and neither the FUSAMI nor the Android functionalities available can measure the pressure value. Therefore, the selected measures that can be extracted from FUSAMI and which were used in this study are:

- Number of gestures
- Frequency of gestures
- Time between each gesture
- Gesture location (gesture heat maps)
- Navigation patterns and transitions

4.3 Mobile application design

With the main purpose of inducing frustration on the user, we developed an Android application with some intentional bugs and problems. Is a simple image search application, where the images are organized into categories and sub-categories.

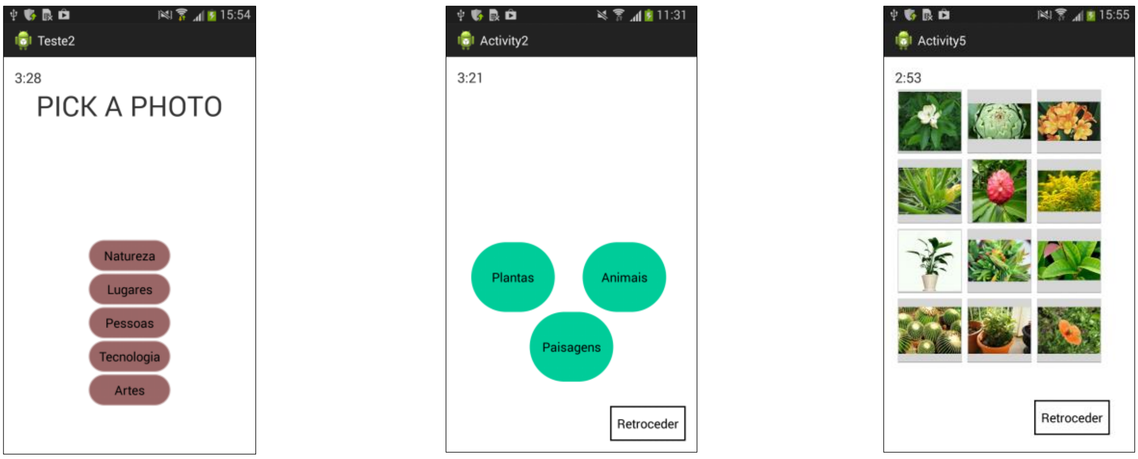

Figure 4- Screenshots of the application pages.

As mentioned in the previous chapter, the most common causes related to user frustration include hardware or software failures, poor user interface design or the user´s lack of experience, error or bugs, slow or dropped internet connections and unwanted pop-ups. In order to convert the navigation in the application into a frustrating experience for the user, we deliberately implemented some of these conditions. While navigating the application to find images, the user is confronted with delays, error messages, and consecutive confirmation pop-up messages. The application layout was designed to be confusing, the menus did not have the same appearance from one category to the next, one of the sub-categories menu was not very explicit and one of the images did not open when selected.

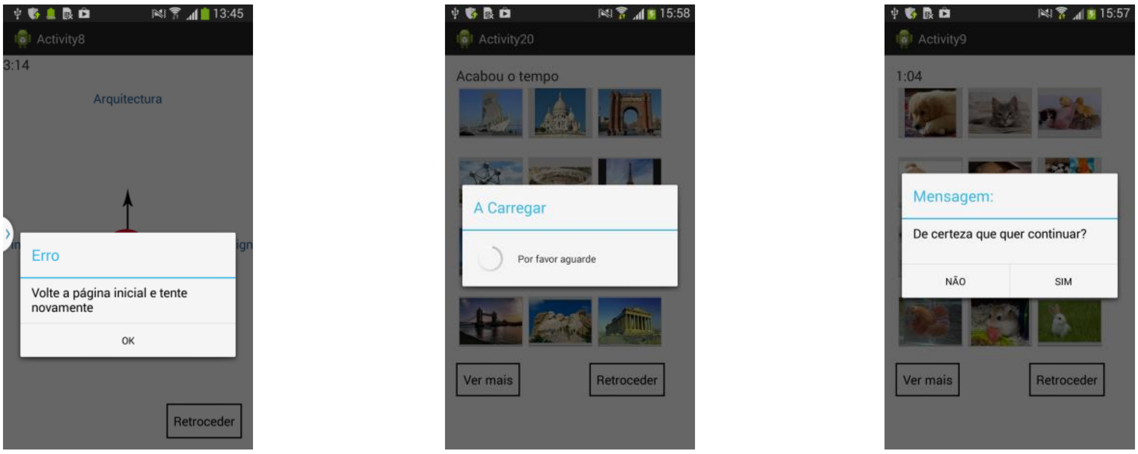

Figure 5- Screenshots with examples of an error, a delay and a confirmation pop-up.

Figure 6- Screenshots with examples of different menu designs.

Frustration is evidenced when a problem occurs and the user has to fix it, or when poor user interface usability leads to loss or failure to attain goals (Lazar, Jones, Hackley, & Shneiderman, 2006). Therefore, the errors, delays and pop-up messages work as obstacles to disrupt the navigation through the application and make it more difficult for the user to achieve the goal. The poor layout designs were implemented to confuse the user and make navigation less obvious. The image that always failed to open was implemented because, as referred before, the interruption of a goal or a task causes frustration to individuals (Lazar et al., 2006). In this case, the user knows that the image is there, but the application bug prevents him/her from opening it (achieving the goal). With the implementation of these elements we aimed to frustrate the users who would be using our application.

4.4 Participants

Seventeen adults (12 male and 5 female) aged from 22 to 51 years (Mean =28.7) voluntarily participated in this study. Participants were recruited from FEUP (Faculty of Engineering –University of Porto) and from Fraunhofer AICOS. Fifteen participants (88.4%) owned a smartphone, two of them used Windows Phone (13.33%), eight used Android (53.33%) and the remaining 5 used the iOS system (33.33%).

4.5 Procedures

The main task of the test was to use the application to find the largest number of images in the shortest possible time. The materials used were a Samsung Galaxy Note II with a screen size of 720 x 1280 pixels, 5.5 inches, and Video Camera PANASONIC / HDC-TM700 to recording the sessions. Figure 7 shows an image of testing room set-up and Figure 8 shows a real image captured during the test.

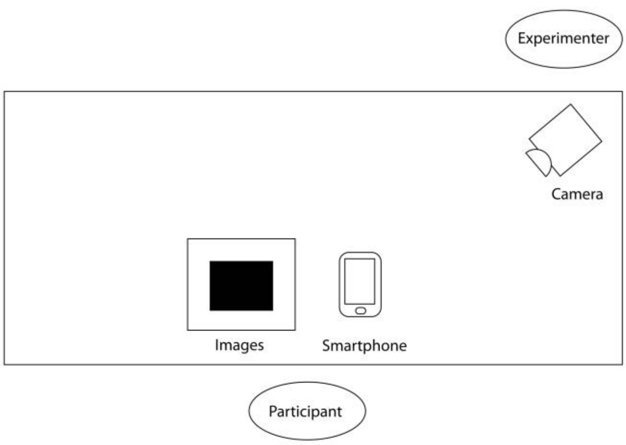

Figure 7- Image of testing room set-up

Figure 8- Image captured during the tests

The real objective of the test – inducing frustration – was never disclosed to participants. They were only told that the test served to collect information about the user experience using the application. Before starting the test, the users filled in a socio-demographic survey and a questionnaire which assessed their emotional state at that time. After completing all tasks, they filled in another questionnaire about their emotional state after the test. These two questionnaires served to help understand whether there was any change in the state of the user caused by the use of the application during the test. The most common two dimensions to measure and describing emotions are arousal and valence. *Arousal* constitutes the activation or excitement level and the *valence* is the positive or negative quality of emotion (Scheirer et al., 2002).According with these two dimensions, frustration can be defined by increased arousal and negative valence (Prendinger & Ishizuka, 2005). In order to measure the arousal, we asked the users to classify their state with a five-point Likert scale – from Quiet to Nervous in the pre-test questionnaire, and from Calm to Irritated in the post-test questionnaire. To measure the valence, we requested the users to classify their satisfaction with the application in the post-test questionnaire. A poor experience using the application can suggest negative valence. The test session was recorded, and we used the video to perform retrospective think-aloud in order to collect more information about the users' emotional state at each point. After performing the test, the users watched the video and verbalised their main difficulties, identifying frustrating episodes during the interaction.

To make users more committed to the tasks and increase their level of frustration, the task had a maximum time limit, 3:30 minutes, to find all the 10 given images. This was meant to look like a challenge or a competition. Participants were also told that nobody before them managed to finish the challenge, to increase their commitment to the task and thus increase their frustration when they were prevented from reaching their goal.

To begin the test, we asked the user to find a given image using the application. The first five tasks were easy and the user should not have any problem finding the images. Then, after this, we started to introduce the manipulated tasks in order to make it difficult, or impossible, for the user to achieve the goal. The tasks were divided into two groups, easy and frustrating, to allow for later comparison of the results. The details of the tasks are described below:

- Tasks 1-5: Ask the user to navigate the application to find the given image. The user should not have any trouble finding the image.
- Task 6: Ask the user to find an image and make the loading times longer than usual.
- Task 7: Ask the user to find an image and make one of the sub-categories (where it is obvious to the user that the image is there) always display an error message, forcing the user to always go back to previous page to solve the problem.
- Task 8: Ask the user to find an image and, as he/she navigates through the application, make the application display a confirmation pop-up message every time the user selects something.
- Task 9: Ask the user to find an image and when he/she tries to open it, it does not open.
- Task10: Ask the user to find an image and make one of the sub-categories menu not explicit (do not use specific terms).

4.6 Results

In this section we analyse and discuss the results. First, the data related to the user emotional state, collected with the surveys, is analysed to compare the level of frustration, irritation and nervousness, and the level of satisfaction reported by the users. Then, the quantitative data collected with FUSAMI is presented, first considering only the users who completed all tasks, except task 10 that no one managed to finish, and then we analyse the numbers related to the most frustrating task according to all users (task 7). After that, some qualitative results are presented, describing the behaviour of users in each task. Finally, in the discussion, the most significant results are interpreted. These results are based only on tap gestures, because, since the application interaction was based on buttons, only this type of gestures was recorded.

4.6.1 User state results

The results comparing the data about the user state before and after the tests are presented below. In Figure 9, we compare the level of nervousness before the test and the level of irritation after the test reported by the users in the surveys. For eight of the participants the value increased, for other eight people the value remained and decreased only for one user. Since the level of nervousness and irritation can be a measure of the arousal dimension of frustration, the increase of these values can be a sign of frustration. For the eight people that showed an increase in this value, two reported frustration level of 5, three reported level 4, one described her frustration state in level 3, and the remaining two classified their frustration level as 1

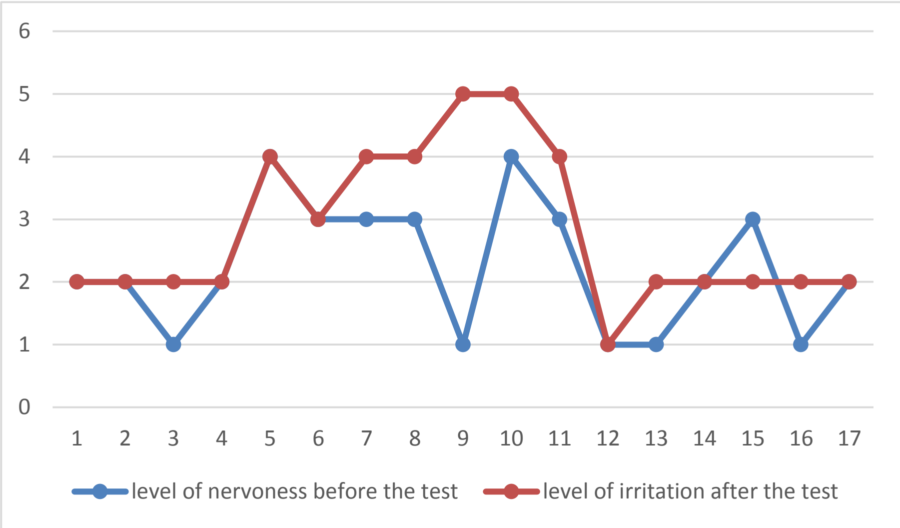

Figure 9- Comparison between the level of nervousness before the test and the level of irritation after the test.

The results presented in Figure 10, show the comparison between the level of frustration reported by the user and the level of satisfaction with the application. For the satisfaction level values, 1 means very satisfied with the application and 5 not satisfied at all.

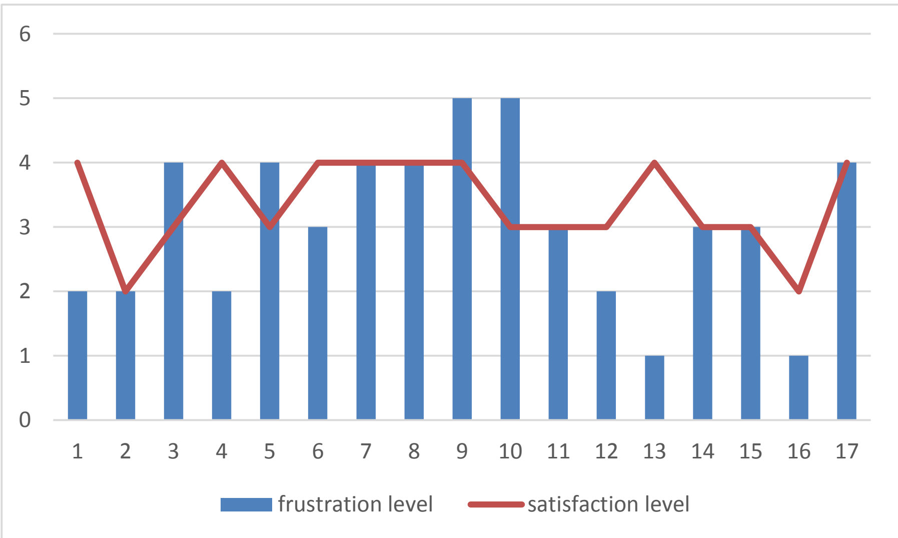

Figure 10- Comparison between the level of frustration and the level of satisfaction reported by the users.

4.6.2 Completed task results

In this sub-section, we present the results for users who were able to complete all tasks without giving up any. We analyse the results of these seven users, considering the number of gestures, time between gestures and frequency of gestures.

4.6.2.1 Number of gestures

Figure 11 shows the results comparing the minimum number of gestures needed to complete each task and the average number that users made. Regarding to the first five tasks, considered to be not frustrating, we can observe that there is not a large variation in the minimum number and the average number of gestures made. In frustrating tasks, it can be seen that this variation is more pronounced, having the task 7 and 9 the biggest difference between these two numbers.

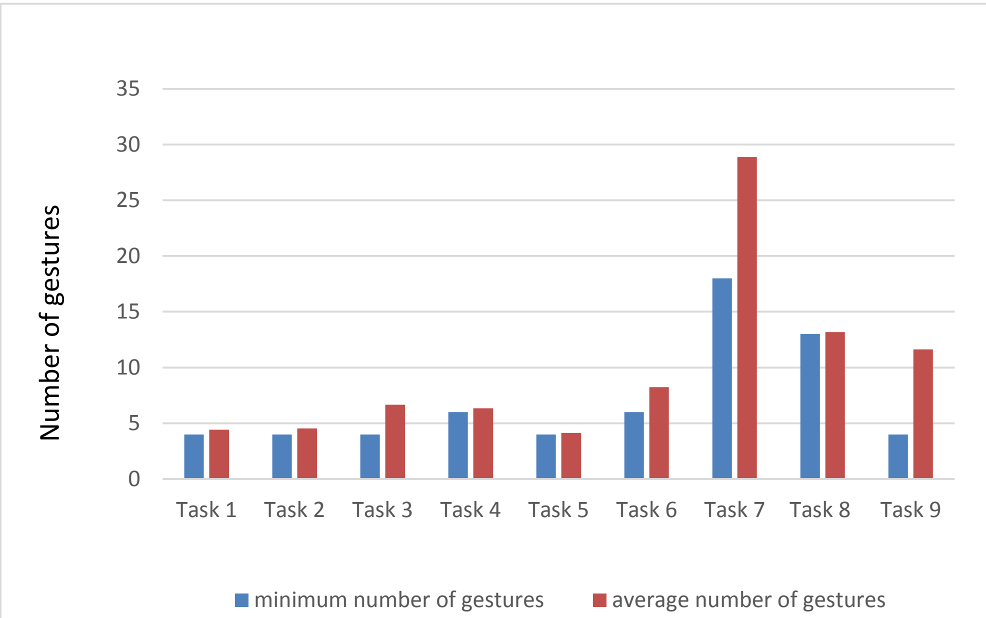

Figure 11- Comparison between the minimum number of gestures to complete the task and the average number of gestures performed by the users for each task.

	Non-frustrating tasks	Frustrating tasks
All users	+ 0.29	+ 4.36
Frustration level 1-3	+ 0.05	+3.25
Frustration level 4-5	+ 0.6	+ 5.83

Table 2- Values of the difference between the minimum number and the average number of gestures comparing the nature of the tasks and the level of frustration of the user.

Values relative to the difference between the minimum number and the average number of gestures made are shown in Table 2. Figure 12 shows these values for all users. Once again we can observe that this value increases in the frustrating tasks as compared to non-frustrating tasks. The average value for this difference increases 0.29 for the first five tasks and 4.36 for the last four tasks.

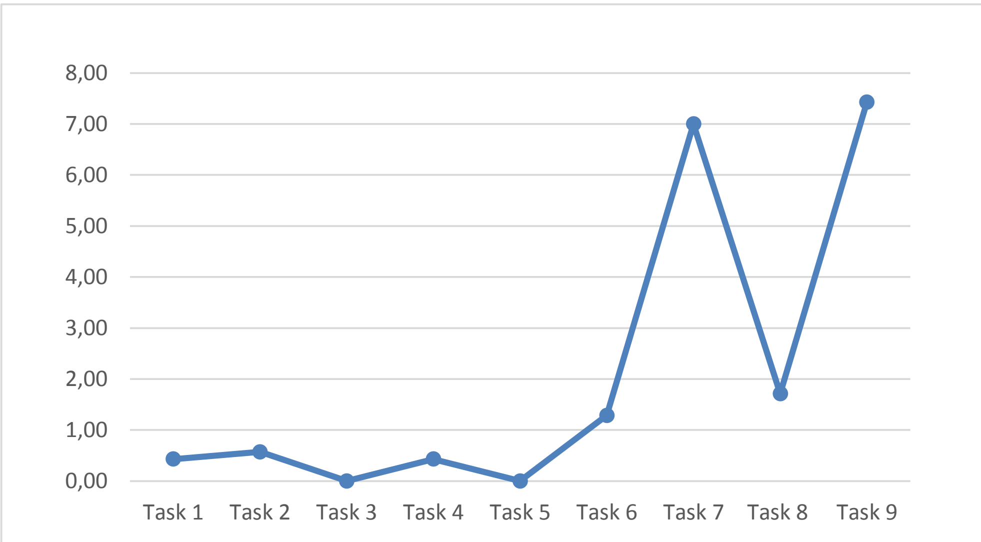

Figure 12- Difference between the minimum number of gestures and the average number for each task.

In the following analysis, users were divided into two groups according to the level of frustration reported in the questionnaires. Figure 13 shows the comparison between the minimum number of gestures and mean values for all users and for the two groups: the first group considering the four users that reported the level of frustration between 1 and 3 and the remaining three, which reported the level of frustration between 4 and 5.

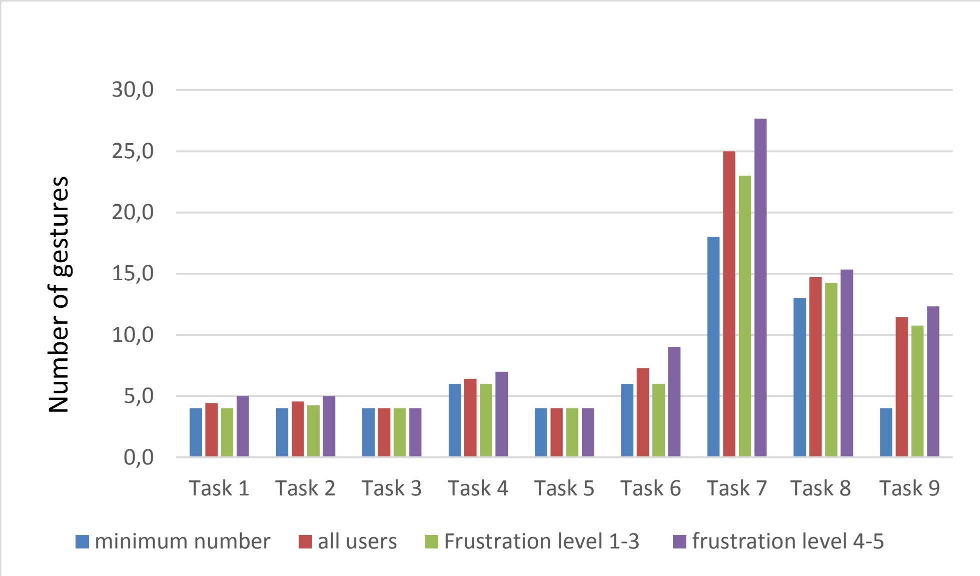

Figure 13- Comparison between the minimum number and the average number of gestures sorted by level of frustration.

For the first group, with frustration level between 1 and 3, the average difference between the minimum number and the number of gestures performed increased 0.05 for the non-frustrating tasks and it increased 3.25 for the frustrating ones. In the other group these values show a more significant increase: more 0.6 for the non-frustrating tasks and 3.83 for the frustrating. Figure 14 shows the comparison between the values of this difference for all users and for each group.

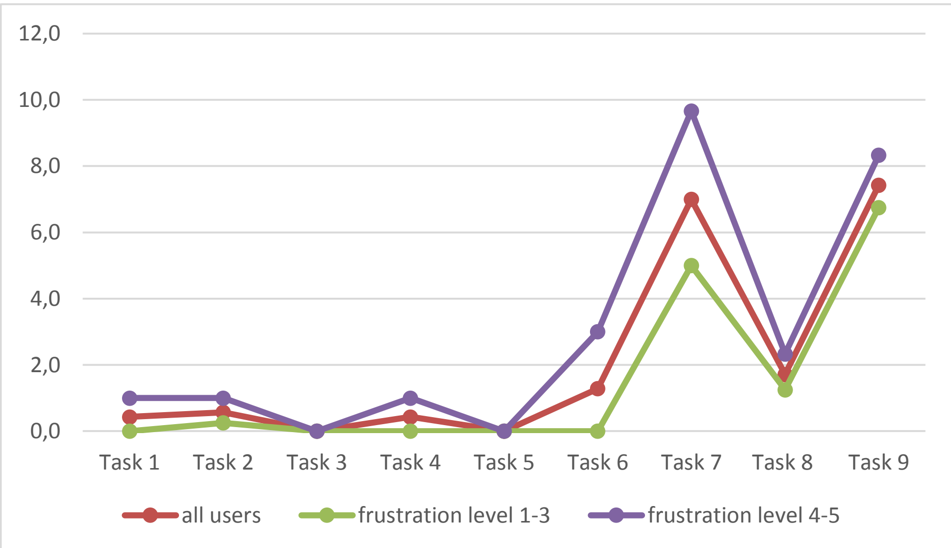

Figure 14- Comparison of the difference between the minimum number and the average number of gestures sorted by the level of frustration.

4.6.2.2 Frequency of gestures

The next results are related to the frequency of gestures for each task. To calculate the frequency of the number of gestures for each task were divided them by the time that the user took to complete the task. Figure 15 provides the mean frequency of gesture for each task. The frequency has higher values for tasks 1, 5 and 7 and the lowest value is the one in the task 6.

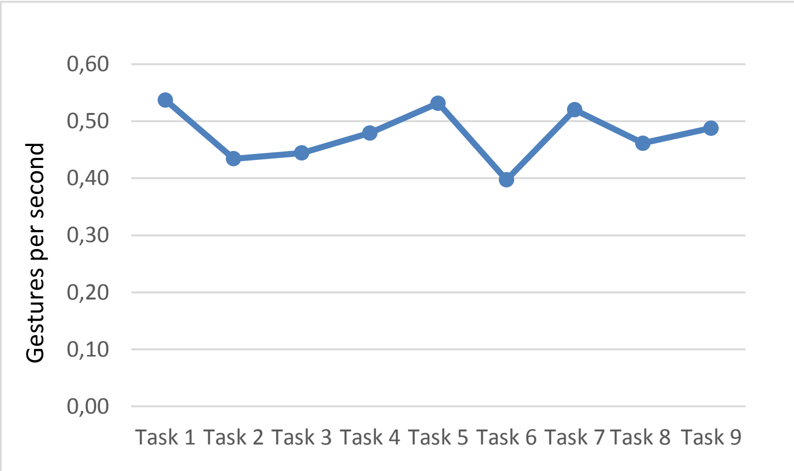

Figure 15- Average values of frequency of gestures for each task.

Comparing this frequency values according to the levels of frustration reported, we can observe that there is no major difference between the mean of the two groups and the average of all users. These results are presented in Figure 16.

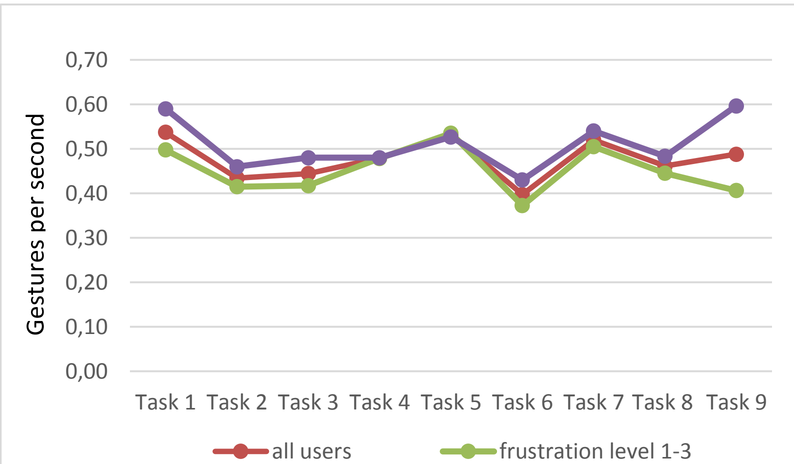

Figure 16- Comparison between average values of frequency for each task for different levels of frustration**.**

4.6.2.3 Time between gestures

The metric that we are going to analyse next is the time between gestures for each task. Figure 17 presents the average values for this measurement for the seven users. We can observe that, for the non-frustrating tasks, values are decreasing. They increase again after the first frustrating task and decrease significantly in task 7.

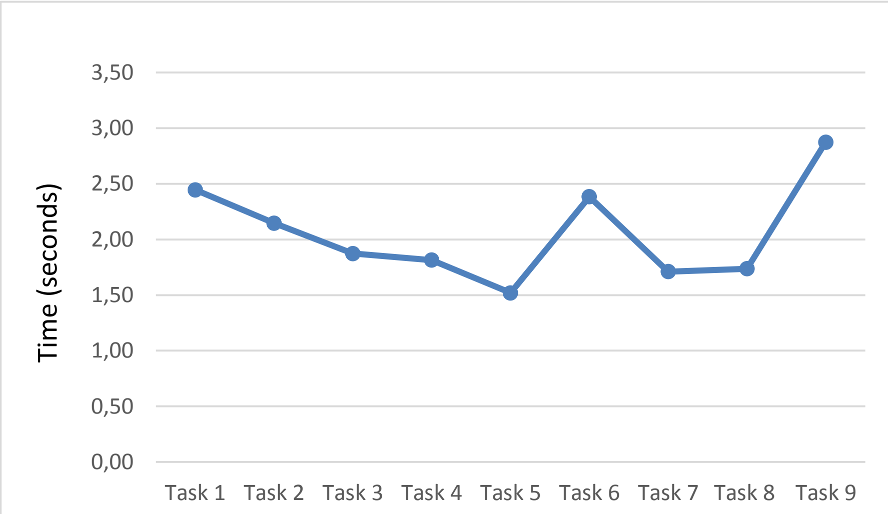

Figure 17- Average values of time between gestures for each task.

Figure 18 shows the comparison of the average time between gestures, for all the seven users and the same measure for the two groups with different levels of frustration. There is no significant variation in these values, only in task 9 this difference is relevant. For this task, the average value for all the users was 3.09 seconds, slightly less when compared to the group with the lowest levels of frustration, where the average value is 3.89 seconds and and slightly higher comparing to the group with the highest levels of frustration, where the average value is 2.04. Comparing these two groups, the average time between gestures increases 1.85 seconds for the group with lower levels of frustration.

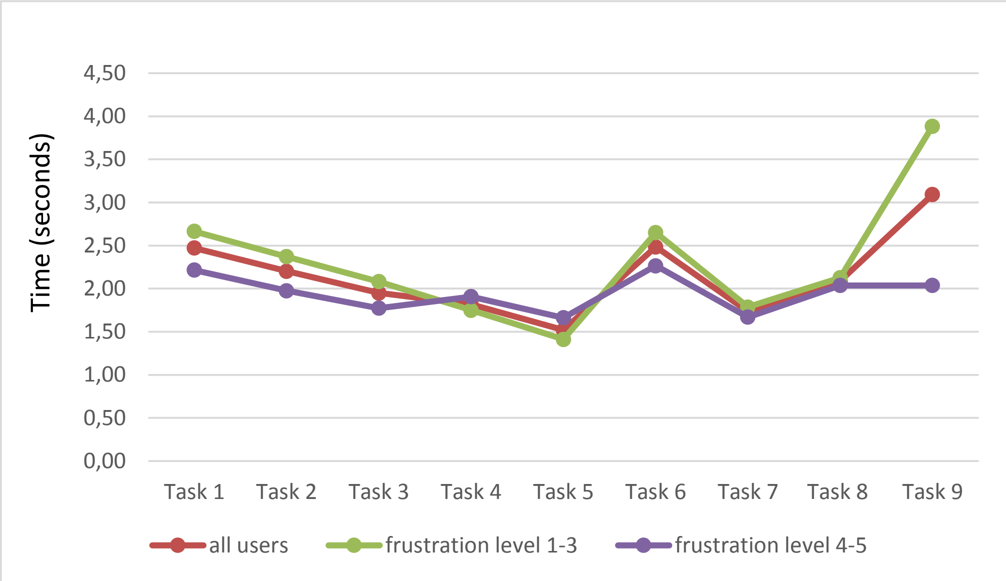

Figure 18- Comparison between average values of time between gestures for each task for different levels of frustration.

4.6.3 Most frustrating task results

All users reported task 7 as the most frustrating task of the test. In this sub-section, the results related to this task are presented. Fourteen users finished this task, the other three dropped out without completing the task. We analyse the number of gestures, frequency and time between it.

4.6.3.1 Number of gestures

The results presented below in Table 3 show the number of gestures performed in task 7, for the fourteen users that finished this task. Figure 19 shows the comparison between the minimum number of gestures to perform the task, eighteen, and the number of gestures that each user performed, and the level of frustration reported by the user. The average difference between the minimum number and the number made by the users is 13.29.

	Most frustrating task (task 7)
All users	+ 13.29
Frustration level 1-3	+ 8.75
Frustration level 4-5	+ 19.33

Table 3-Values of the difference between the minimum number and the average number of gestures in task 7 comparing the level of frustration of the user.

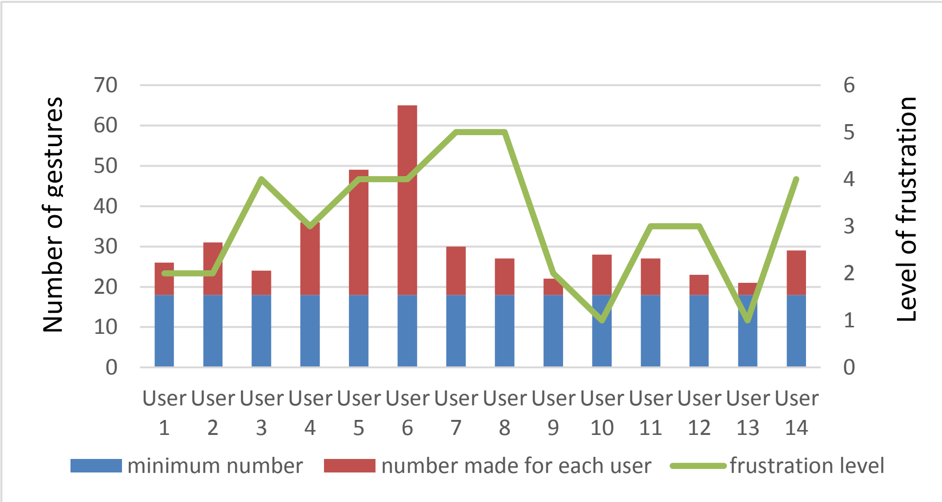

Figure 19- Comparison of the minimum number, the performed number of gestures and the level of frustration for task 7.

The next two figures show the same results but sort by the level of frustration. Figure 20 shows the difference between the minimum number and number of gestures made by the users who reported frustration levels from 1 to 3. For this group the average difference value between the minimum and the performed is 8.75. For the other group, represented in Figure 21, this average value increases significantly to 19.33.

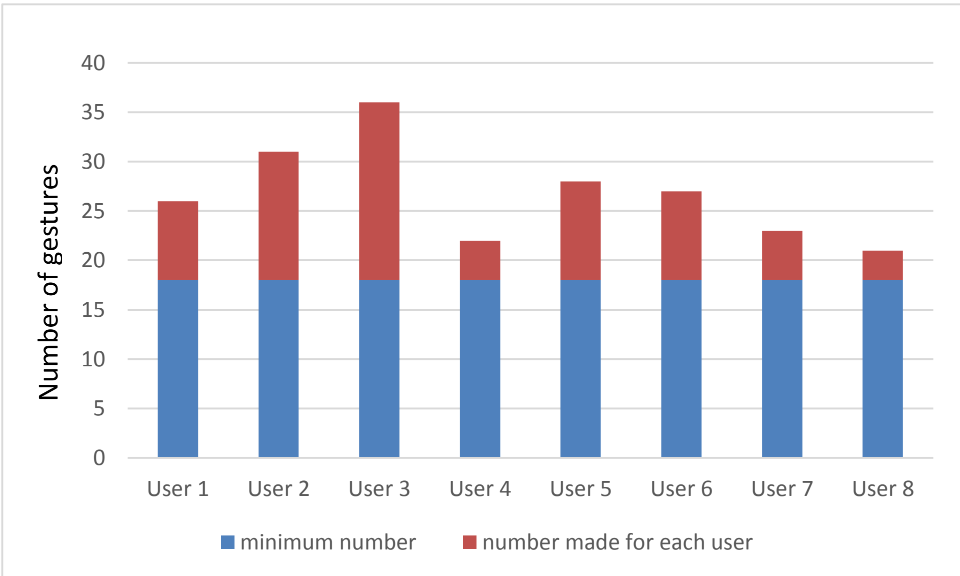

Figure 20- Comparison of the minimum number and the performed number of gestures for task 7 of users with frustration levels between 1 and 3.

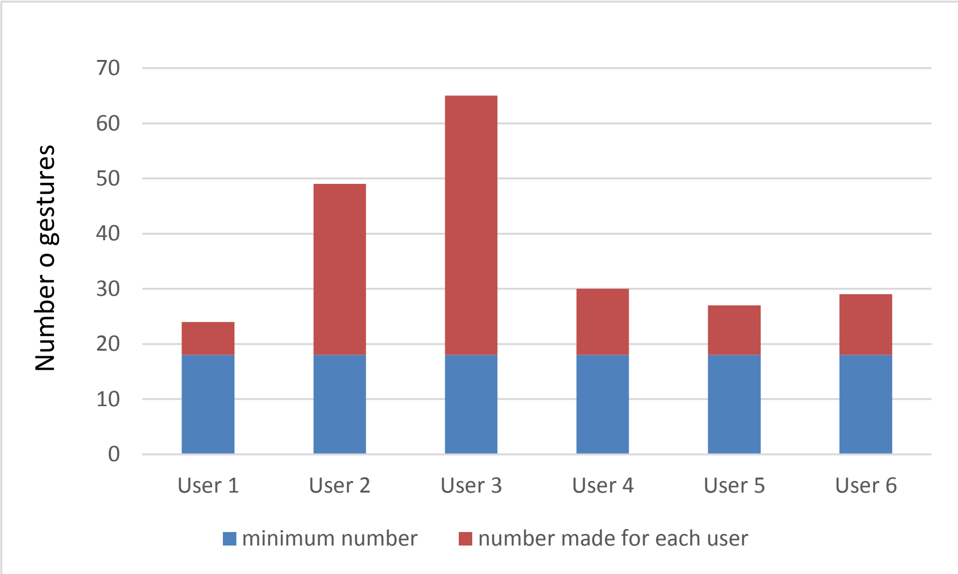

Figure 21- Comparison of the minimum number and the performed number of gestures for task 7 of users with frustration levels between 4 and 5.

4.6.3.2 Frequency of gestures

The average value of frequency for this task is 0.55 gestures per second. Comparing to other task, where the average value is 0.47 for the non-frustrating tasks (task 1 to 5) and 0.49 for the remaining frustrating tasks (tasks 6, 8 and 9), there is no significant variation for this measure.

4.6.3.3 Time between gestures

Regarding the time between gestures, the average value for this task was 1.65 seconds, slightly less when compared to the non-frustrating tasks, where the average value is 2 seconds and to the others non frustrating task, that has an average value of 2.13 seconds.

4.6.4 Navigation patterns and transitions

Analysing the patterns of navigation provided by FUSAMI, we can observe that the non-frustrating tasks have larger patterns comparing to the frustrating tasks. This is due to the fact that most people follow the same steps in the first tasks. The patterns relating to frustrating tasks are smaller in size and larger in quantity indicating that participants had more variations in their actions during the interaction. Figure 22 shows an example of the largest pattern found by FUSAMI. This pattern shows the interaction between tasks 1 and 6 and the beginning of the task 7, and has a size of 27 steps.

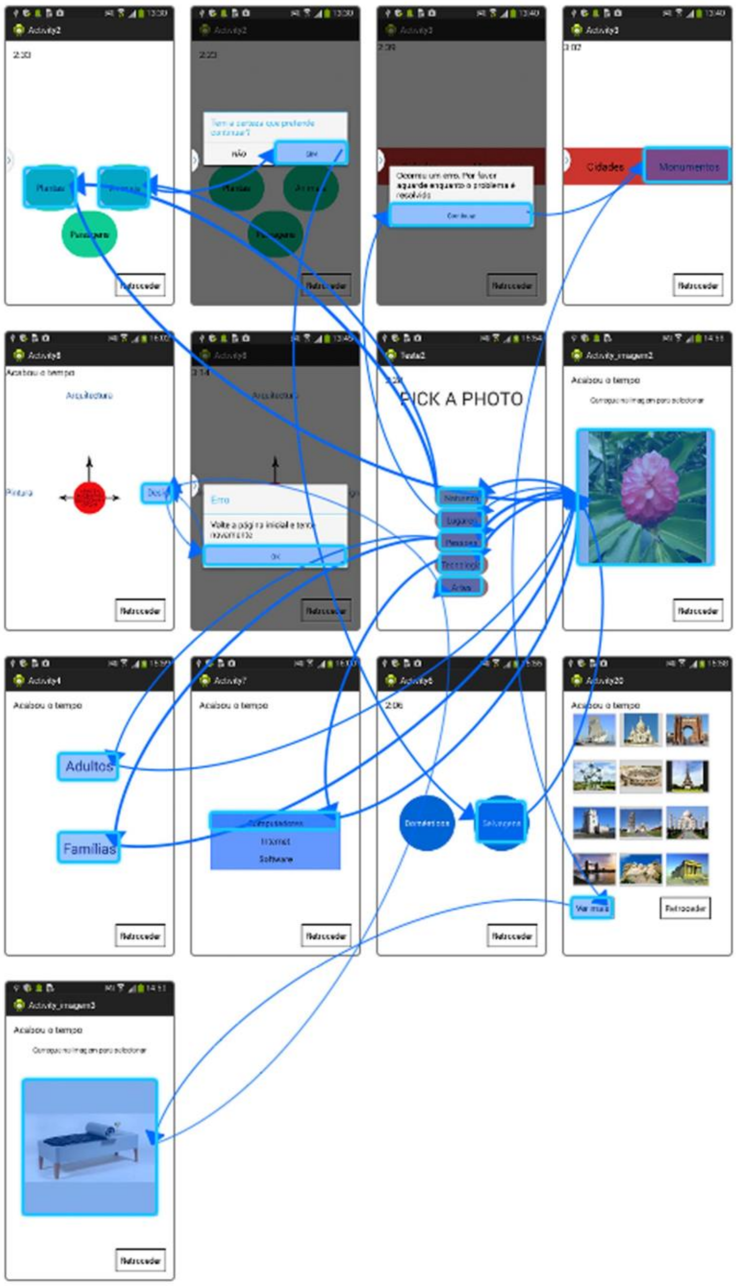

Figure 22- Example of a navigation pattern from FUSAMI for the tasks 1 to 6.

In the following figures we can see examples of patterns from task 7. The patterns that include only this task are more frequent and smaller compared with the other tasks.

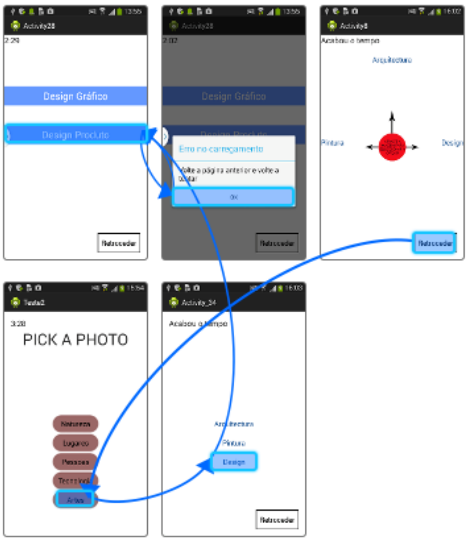

Figure 23- Example of a pattern with 6 steps for the task 7

Figure 24- Example of a pattern with 5 steps for the task 7

The pattern in Figure 23 has a size of 6 steps and shows the interaction between the first and the second error in task 7. Figure 24 shows a smaller pattern where only the second error appears, and has a size of 5 steps. Taking into account previous patterns we can find a common behaviour for multiple users. When an error appeared most users ignored the message and insisted on pressing the button again or searched in another category instead of going back, as the error message instructed. Figure 255 shows the transition that demonstrates this behaviour.

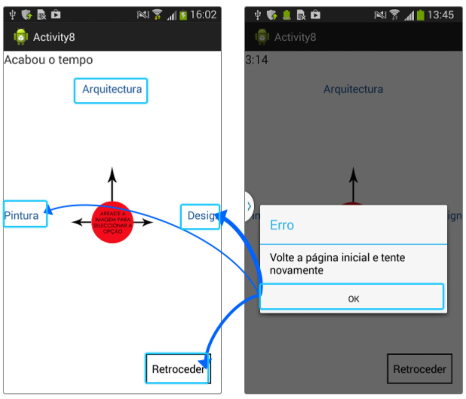

Figure 25- Transition pattern of the first error in task 7

4.6.5 Gesture location (gesture heat maps)

Due to the fact that the application created has a navigation based on menus with buttons, the heat maps extracted from the FUSAMI do not show any relevant information. As shown in the figure below, the heat maps show that the participants used the buttons on the interaction and did not perform any other type of gestures beyond *tap*.

Figure 26- Examples of heat maps from FUSAMI

4.6.6 Behavioural data and results

The data collected from the videos recorded can show different user behaviours. This information can be used in combination with the quantitative data described above to help us in getting at a more complete picture of users' behaviour which is not captured by FUSAMI. Some of them remained calm when faced with frustrating episodes, while others showed signs of nervousness and irritation. The most common behaviours among participants during the loading times were tapping their fingers or their fist on the table, and snapping their fingers. When something unexpected happened such as a mistake, or when they appeared to be confused, the most common signs among them were facial expressions like frowning or raising eyebrows, gesturing with hands or moving back in the chair. Some participants also expressed themselves verbally when they had problems with some task. The table below shows the results of these behavioural signs for each participant.

Participant	Level of frustration	Verbal expression	Hand gestures	Facial expression	Posture movements
1	2		X		
2	2		X		
3	4		X		
4	2		X	X	
5	4		X	X	
6	3				X
7	4		X	X	X
8	4		X		X
9	5		X		X
10	5	X	X		
11	3	X			
12	2				
13	1		X	X	
14	3		X	X	
15	3		X	X	
16	1		X		
17	4		X		

Table 4- Most common behaviours showed by participants

This behavioural information his described below for each of the tasks.

- Tasks 1 to 5: All participants, except one, were able to complete all tasks with ease and without problems. One participant gave up on task 3 because he was confused by the menu. The menu had three buttons with options, but he considered it to be a single button and always pressed the same option, never managing to reach the desired image.

- Task 6: All participants, except one, were able to complete the task. The participant who did not complete the task had trouble discovering the image category and gave up. Most users got a little frustrated with this task, because the pages took too long to load and time was running out.

- Task 7: This task was considered the most frustrating for all the users. Three participants dropped out and did not complete the task. When the error appeared, most people did not read the message saying to return to the previous page and try again, and continued to insist on the same option. After insisting on the option that originated an error message, some participants read the message and proceeded correctly, while others tried to search in the remaining categories that did not give the error message.

- Task 8: Six out of the seventeen participants did not complete the task, five of them for lack of time and the other one just looked for the picture on the first page, but the image was on page three so he gave up. Users had little difficulty in this task, just considered it a little annoying because of the constant pop-up windows that made them waste time.

- Task 9: Since it was impossible to finish this task correctly because the final image always appeared blank, it was considered that participants who found the correct image and tried to open it completed the task. Thus only two people did not complete the task, one due to lack of time and the other because he could not find the respective category and moved to the next image. The reactions to these images were distinct. When the image appeared blank, some people continued to press the image, others navigated back and tried to select the image again, while others just looked at the picture waiting for something to happen.

- Task 10: Only two people managed to reach this task but none managed to complete it.

4.7 Discussion

In this section all measures are analysed and interpreted in order to understand if frustration can be recognized through gestures. Although the current study is based on a small sample of participants, the results showed that some of the measures studied might be relevant in the study of frustration.

4.7.1 Number of gestures

The first measure that we are going to analyse is the number of gestures. In general, the number of gestures performed for each task increased when comparing the non-frustrating with the frustrating tasks. This number shows a considerable variation when compared at different levels of frustration. In the non-frustrating tasks the average difference between the minimum number and the performed number is 0.29. Analysing the same value according to the level of frustration, the value decreases to 0.05 for the users with frustration between 1 and 3, and 0.6 to the other group of users that reported frustration as 4 or 5. On the frustrating tasks, this average difference value is higher in all the previous cases. For all the users who completed all the tasks, the average number of gestures performed increases 4.36 when compared to the minimum number. The value decreases slightly compared to the group with the lowest level of frustration (3.25), and increases when compared with the other group (5.83). When analysing the data of the task 7 can we can see that, for users who have completed the task, the difference between the number of minimal gestures and the number of gestures made is high (13.29). For all the cases mentioned above, the number of gestures increases significantly in the tasks considered frustrating. This value increases significantly when compared to the average difference value of the frustrated users (19.33). For the other group of users (frustration level 1 to 5) this value was 8.75. Looking at the values mentioned above, we can observe that the number of gestures increases when comparing non-frustrating with frustrating tasks and also increases when the level of user frustration increases. Therefore, the number of gestures can be a meaningful measure to analyse and infer the user frustration.

4.7.2 Frequency of gestures

When analysing the frequency of gestures on the above cases, it can be seen that there is no significant difference between frustrating tasks when compared with non-frustrating. Neither when the users are distinguished by their level of frustration, the value of the frequency is very close to the two groups. Therefore, we cannot consider the frequency of the gestures performed as a relevant measure.

4.7.3 Time between gestures

As seen in Figure 15, the average time between gestures begins to decrease after the first task, and increases again in task 6, the first frustrating task. Then it decreases again significantly in task 7, and increases in the last two tasks. The successive decrease in the first tasks may reflect the adaptation of the user to the application user interface. Familiarity with the user interface causes the user to be faster in the interaction, which may explain the reduction of the time between gestures. Considering the values of this measure for task 7, we can observe that average time is shorter compared with other tasks, except for task 5. Since this task was considered to be the most frustrating, the decrease of time can mean that when the user is frustrated he or she makes more consecutive and faster gestures to try to solve a problem or to overcome an obstacle. The average value of time between gestures in task 9 shows a variation when comparing the two groups of users according to their level of frustration. Since the task 9 contained the image that always failed to open, a possible explanation for these values would be the fact that, when trying to solve the problem, less frustrated users reacted more calmly, while the other users performed more consecutive and faster gestures, which can be interpreted a sign of irritation and impatience.

4.7.4 Navigation patterns and transition results

The analysis of the navigation patterns suggests that in the easy and intuitive task, users behave in a similar way with slight differences in their interaction. In Task 7, the patterns are smaller and fragmented, suggesting that there was not a common behaviour during navigation between users and only a few steps are repeated.

4.7.5 Behavioural data

In most cases, people that showed these behaviours in a more effusive way, reported higher levels of frustration after the test. One of the most frequent behaviours among users was gesturing with his hands, like tapping the table with his fingers or fists or snapping fingers, while waiting. This signs suggested that participants might be slightly nervous or anxious. Other significant signs shown by users were their facial expressions, for example frown or raise their eyebrows, suggesting that they are confused or surprised.

4.7.6 General discussion

These findings suggest that in general, the most meaningful measure for the study was the number of gestures. Other measures such as the time between gestures, navigation patterns and transitions and the behaviour data also showed some results that should be considered relevant to this study. We could not fully compare the results with those obtained in previous work, described in Chapter 3, because the measures chosen for this work are generally different from what has been studied before. The most common measures used in the study of frustration were the pressure applied on the mouse or touchpad and physiological signals. However, in our study it was not possible to analyse these variables.

Amongst the studies analysed previously, in only one of them is there a measure that can be compared with those studied in our own study. There are similarities between the measures in this study and those described by Scheirer, Fernandez, Klein, & Picard (2002), who examined the mouse click behaviour during each frustrating episode. They distinguished four different types of behavioural responses according to the number of mouse clicks and suggested that superfluous clicking was a natural response for most users. They also suggest that a variety of typed patterns, such as repeating erroneous commands, could provide clues to the user's affective state. The present results show similar conclusions when analysing the number of gestures, which increases for the most frustrating tasks and when the user reports a higher level of frustration. Analysing the patterns of navigation can also help to identify different types of user behaviour that can be useful to identify their affective state.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

This dissertation aimed to extract meaning from users' gestures and their interaction in smartphones in order to study their behaviour for usability evaluation. This study set out to determine if it was possible to induce user frustration with a smartphone and if there were gesture patterns that could be associated with it. In order to achieve that a set of experiments were conducted, using an Android application with some intentional bugs and problems. All the test sessions were recorded in video and the data relating to the gestures were collected with FUSAMI.

For the first question, the results suggested that it is possible to induce user frustration. However, the level of frustration varies considerably from one person to the next. Faced with the same situation during the various episodes of frustration, some people react calmly, others show signs of nervousness and others become angrier, verbally expressing their frustration.

Regarding the second question, the results of this investigation show that the most meaningful measure for the study was the number of gestures performed, which shows significant variations according to the nature of the task and the level of frustration reported by the participants. The findings suggest that the number of gestures performed increases when the user encounters frustrating situations or when his/her level of frustration is higher. The time between gestures measure results also show some variations that should be taken into consideration. The decrease of time between gestures can be an indicator of impatience and irritation of the user, indicating that he/she has encountered a problem or obstacle during their interaction. The frequency measure shows no conclusive results.

Finally, a number of important limitations need to be considered. First, the number of participants in the tests was not large enough. In order to make the results more conclusive, the number of tests performed should have been higher. The main reason why the number of participants was smaller than expected was the difficulty in finding volunteers who agreed to participate in the study. Another limitation is the fact that the tests were undertaken within a laboratory environment and not in real-life cases of interaction. This might have had influence on participants' reaction.

5.2 Future work

The current study was unable to analyse the pressure variable. As seen in the previous work related to frustration, the pressure is a significant and common measure used to identify this state. Studying this measure can be a great asset to identify and measure frustration.

Since the application created has a navigation based on menus with buttons, the majority of gestures recorded were taps on the buttons. An application with different types of navigation where users could have more freedom in the gestures could provide more conclusive results.

Although the frequency and the time between gestures have not shown significant results in this study, these measures may turn out to be a relevant metric if explored in more detail.

References

Andrienko, G., Andrienko, N., Burch, M., & Weiskopf, D. (2012). Visual analytics methodology for eye movement studies. *Visualization and Computer Graphics, IEEE Transactions on*, *18*(12), 2889–2898.

Asteriadis, S., Tzouveli, P., Karpouzis, K., & Kollias, S. (2009). Estimation of behavioral user state based on eye gaze and head pose—application in an e-learning environment. *Multimedia Tools and Applications*, *41*(3), 469–493.

Bessiere, K., Ceaparu, I., Lazar, J., Robinson, J., & Shneiderman, B. (2003). Understanding Computer User Frustration: Measuring and Modeling the Disruption from Poor Designs. Retrieved from http://drum.lib.umd.edu//handle/1903/1233

Blignaut, P. (2010). Visual Span and Other Parameters for the Generation of Heatmaps. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 125–128). New York, NY, USA: ACM. doi:10.1145/1743666.1743697

Cartmill, E. A., Beilock, S., & Goldin-Meadow, S. (2012). A word in the hand: action, gesture and mental representation in humans and non-human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 129–143. doi:10.1098/rstb.2011.0162

References

Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J., & Shneiderman, B. (2004). Determining Causes and Severity of End-User Frustration. *International Journal of Human-Computer Interaction*, *17*(3), 333–356. doi:10.1207/s15327590ijhc1703_3

Chen, L. S.-H. (2000). *Joint processing of audio-visual information for the recognition of emotional expressions in human-computer interaction*. Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.76.1185

Dennerlein, J., Becker, T., Johnson, P., Reynolds, C., & Picard, R. W. (2003). Frustrating computer users increases exposure to physical factors. In *Proceedings of the International Ergonomics Association, Seoul, Korea*. Retrieved from http://www.researchgate.net/publication/2832780_Frustrating_Computers_Users_Increases/file/60b7d528f8535cac39.pdf

Frank, M., Biedert, R., Ma, E., Martinovic, I., & Song, D. (2013). Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *Information Forensics and Security, IEEE Transactions on*, *8*(1), 136–148.

Hudlicka, E. (2001). CAPTA: Cognitive affective personality task analysis. In *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomic Society. Minneapolis, MN, October*. Retrieved from http://www.researchgate.net/publication/228966850_CAPTA_Cognitive_affective_personality_task_analysis/file/32bfe50e1936a5bd83.pdf

Hudlicka, E. (2003). To feel or not to feel: The role of affect in human–computer interaction. *International Journal of Human-Computer Studies*, *59*(1), 1–32.

Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, *2*(3), 4.

References

Kapoor, A., Burleson, W., & Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies*, *65*(8), 724–736. doi:10.1016/j.ijhcs.2007.02.003

Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration: - Theory, design, and results. *Interacting with Computers*, *14*(2), 119–140. doi:10.1016/S0953-5438(01)00053-4

Kurzhals, K., & Weiskopf, D. (2013). Space-Time Visual Analytics of Eye-Tracking Data for Dynamic Stimuli. *Visualization and Computer Graphics, IEEE Transactions on*, *19*(12), 2129–2138.

Lazar, J., Jones, A., Bessiere, K., Ceaparu, I., & Shneiderman, B. (2005). User Frustration with Technology in the Workplace (2004). Retrieved from http://drum.lib.umd.edu//handle/1903/6515

Lazar, J., Jones, A., Hackley, M., & Shneiderman, B. (2006). Severity and impact of computer user frustration: A comparison of student and workplace users. *Interacting with Computers*, *18*(2), 187–207. doi:10.1016/j.intcom.2005.06.001

Lehman, B., Matthews, M., D'Mello, S., & Person, N. (2008). What are you feeling? Investigating student affective states during expert human tutoring sessions. In *Intelligent Tutoring Systems* (pp. 50–59). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-69132-7_10

Leitão, R. A. (2013). Creating mobile gesture-based interaction design patterns for older adults : a study of tap and swipe gestures with portuguese seniors. Retrieved from http://repositorio-aberto.up.pt/handle/10216/68413

McLaughlin, M., Chen, Y.-S., Park, N., Zhu, W., & Yoon, H. (2004). Recognizing user state from haptic data. In *International Conference on Information Systems Analysis and Synthesis*. Retrieved from http://www.academia.edu/download/30986333/CITSA04Paper2.pdf

References

Mentis, H. M., & Gay, G. K. (2002). Using TouchPad Pressure to Detect Negative Affect. In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces* (p. 406–). Washington, DC, USA: IEEE Computer Society. doi:10.1109/ICMI.2002.1167029

Pantic, M., & Rothkrantz, L. J. (2003). Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, *91*(9), 1370–1390.

Panwar, M., & Mehra, P. S. (2011). Hand gesture recognition for human computer interaction. In *2011 International Conference on Image Information Processing (ICIIP)* (pp. 1–7). doi:10.1109/ICIIP.2011.6108940

Pavlovic, V. I., Sharma, R., & Huang, T. S. (1997). Visual interpretation of hand gestures for human-computer interaction: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(7), 677–695. doi:10.1109/34.598226

Pernice, K., & Nielsen, J. (2009). *How to conduct Eyetracking Studies | Free Nielsen Norman Group Report*. Retrieved February 25, 2014, from http://www.nngroup.com/reports/how-to-conduct-eyetracking-studies/

Picard, R. W. (1999). Affective Computing for HCI. In *HCI (1)* (pp. 829–833). Retrieved from http://www.media.mit.edu/affect/pdfs/99.picard-hci.pdf

Poole, A., & Ball, L. J. (2005). Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future. In *Prospects", Chapter in C. Ghaoui (Ed.): Encyclopedia of Human-Computer Interaction. Pennsylvania: Idea Group, Inc*.

Prendinger, H., & Ishizuka, M. (2005). The Empathic Companion: A Character-Based Interface That Addresses Users Affective States. *Applied Artificial Intelligence*, *19*(3-4), 267–285. doi:10.1080/08839510590910174

Qi, Y., Reynolds, C., & Picard, R. W. (2001). The Bayes Point Machine for Computer-user Frustration Detection via Pressuremouse. In *Proceedings of the 2001 Workshop on Perceptive User Interfaces* (pp. 1–5). New York, NY, USA: ACM. doi:10.1145/971478.971495

References

Schaefers, K., Ribeiro, D., & de Barros, A. C. (2013). Beyond Heat Maps: Mining Common Swipe Gestures. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia* (pp. 3:1–3:4). New York, NY, USA: ACM. doi:10.1145/2541831.2541842

Scheirer, J., Fernandez, R., Klein, J., & Picard, R. W. (2002a). Frustrating the user on purpose: a step toward building an affective computer. *Interacting with Computers*, *14*(2), 93–118. doi:10.1016/S0953-5438(01)00059-5

Scheirer, J., Fernandez, R., Klein, J., & Picard, R. W. (2002b). Frustrating the user on purpose: a step toward building an affective computer. *Interacting with Computers*, *14*(2), 93–118. doi:10.1016/S0953-5438(01)00059-5

Scherer, S., Glodek, M., Layher, G., Schels, M., Schmidt, M., Brosch, T., … Palm, G. (2012). A generic framework for the inference of user states in human computer interaction. *Journal on Multimodal User Interfaces*, *6*(3-4), 117–141.

Shaker, N., Asteriadis, S., Yannakakis, G. N., & Karpouzis, K. (2013). Fusing visual and behavioral cues for modeling user experience in games. *Cybernetics, IEEE Transactions on*, *43*(6), 1519–1531.

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., & more, & 1. (2009). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (5 edition.). Boston: Prentice Hall.

Sonam, K., & Ubale, V. S. (n.d.). Gesture Recognition-A Review. *IOSR Journal of Electronics and Communication Engineering*, *5*, 19–26.

Tsang, H. Y., Tory, M., & Swindells, C. (2010). eSeeTrack--visualizing sequential fixation patterns. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 953–962. doi:10.1109/TVCG.2010.149

Van den Hoven, E., & Mazalek, A. (2011). Grasping gestures: Gesturing with physical artifacts. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *25*(Special Issue 03), 255–271.

References

Whitenton, K. (2013). *Minimize Cognitive Load to Maximize Usability | Nielsen Norman Group*. Retrieved June 25, 2014, from http://www.nngroup.com/articles/minimize-cognitive-load/

Yousefi, S., Abedan Kondori, F., & Li, H. (2013). Experiencing real 3D gestural interaction with mobile devices. *Pattern Recognition Letters, 34*(8), 912–921. doi:10.1016/j.patrec.2013.02.004

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *31*(1), 39–58.

Zeng, Z., Tu, J., Liu, M., Huang, T. S., Pianfetti, B., Roth, D., & Levinson, S. (2007). Audio-visual affect recognition. *Multimedia, IEEE Transactions on*, *9*(2), 424–428.

Zeng, Z., Tu, J., Liu, M., Zhang, T., Rizzolo, N., Zhang, Z., … Levinson, S. (2004). Bimodal HCI-related affect recognition. In *Proceedings of the 6th international conference on Multimodal interfaces* (pp. 137–143). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1027958