Universidade de Aveiro
2019

Departamento de Eletrónica, Telecomunicações e Informática

Daniel Barbosa de Oliveira

Utilização de um Sistema Multimodal para a Recolha de Informação Afetiva

A Multi-Modal Approach for Affective Data Gathering

Universidade de Aveiro
2019

Departamento de Eletrónica, Telecomunicações e
Informática

Daniel Barbosa de Oliveira

Utilização de um Sistema Multi-Modal para a Recolha de Informação Afectiva

A Multi-Modal Approach for Affective Data Gathering

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia de Computadores e Telemática, realizada sob a orientação científica do doutor Ilídio Castro Oliveira, Professor Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro e coorientação da doutora Susana Manuela Martinho dos Santos Baía Brás, Investigadora do IEETA, Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro.

Dedico este trabalho à minha família.

o júri

Presidente	**Professor Doutor António Joaquim da Silva Teixeira** Professor Associado c/ Agregação, Universidade de Aveiro
Vogal – Arguente Principal	**Doutora Maria Luísa de Morais e Sousa de Castro e Guedes** Investigadora Pós-Doc, Inesc Tec – Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência
Vogal – Orientador	**Professor Doutor Ilídio Fernando de Castro Oliveira** Professor Auxiliar, Universidade de Aveiro

agradecimentos

O presente trabalho deve a sua existência a quem me ajudou direta ou indiretamente, nomeadamente os meus mentores que me ajudaram quando encontrei dificuldades, os meus pais e irmão, pelo incansável e incondicional apoio que me deram, os participantes no estudo, que tornaram possível a existência de dados para serem analisados, a minha namorada, pelo apoio e pelas sugestões, o IEETA, pelo fornecimento do equipamento e de outros meios necessários, à Professora Cláudia Figueiredo (GOVCOP/DEP), pelo apoio prestado no desenvolvimento do estudo com participantes e ao Dr. Cláudio Teixeira.

Palavras-chave Computação Afetiva, Sistema Multimodal, Stress

Resumo

O reconhecimento, a interpretação e o processamento de emoções (*Affective Computing*) é uma área emergente das aplicações computacionais. Existem vários métodos de aquisição de dados e de classificação de emoções, com precisões distintas, em que os sistemas multimodais apresentam geralmente uma precisão mais elevada do que os unimodais.

Nesta dissertação, procuramos investigar os métodos atualmente usados para recolher informação afetiva bem como métodos para a análise da mesma, tendo em vista uma proposta de um sistema multimodal, com foco em métodos não-intrusivos, com potencial aplicação na monitorização de stress ocupacional.

O sistema desenvolvido tem como objectivo a recolha de informação afetiva, incluindo várias fontes de dados, como informação sobre utilização do rato e do teclado, dados ECG, vídeo da face e gravações de vídeo do ecrã do computador (para deteção de atividades). Para a classificação de emoções, foram utilizados os algoritmos de *Clustering* e de *Random Forest*.

Num estudo exploratório, usando o *dataset* de investigação SWELL, testámos o algoritmo de *Random Forest* e obtivemos uma precisão global de 89.97% na classificação, o que considerámos satisfatória, uma vez que é comparável com os resultados apresentados na literatura.

O sistema desenvolvido foi testado num conjunto de onze participantes. Globalmente, o algoritmo de *Random Forest* obteve uma taxa de erro de 65%. O algoritmo de *Clustering* testado não classificou acima de 3% dos dados na classe 2. Quando se avaliaram os questionários de avaliação do estado emocional (aplicados antes e depois do teste ao sistema), verificou-se que os participantes reportaram um decremento na ansiedade sentida depois da realização do estudo. O que pode indicar que o protocolo de recolha de dados apresentado pode não ter induzido os estados emocionais pretendidos (stress) nos participantes

O sistema multimodal encontra-se funcional e pode ser aplicado em outros estudos para recolha de marcadores de emoções.

Keywords Affective computing, multimodal system, stress

Abstract

Recognizing, interpreting and processing emotions (Affective Computing) is an emerging field of computer science. Multiple methods of data acquisition and emotion classification exist with different accuracy performances. Despite this, multimodal systems, generally have a higher accuracy than unimodal ones.

This dissertation's goal is to research the current methods of both affective data gathering and emotion classification while developing a multi-modal system, that focuses primarily on the utilisation of non-intrusive methods with potential application in cccupational stress.

The system has the purpose of collecting affective data including multiple data gathering methods such as mouse and keyboard utilisation data, ECG data, face and upper body video recordings and computer screen video recordings (for activity detection). For the emotion classification, the *Clustering* and *Random Forest* algorithms were utilised.

In the exploratory study with the already existent *SWELL* investigation dataset, we tested the algorithm of *Random Forest* and an overall accuracy of 89.97% was achieved, which we considered acceptable. In order to validate the final system, a study with eleven participants was conducted. An overall error rate of approximately 65% was achieved with the *Random Forest* algorithm. For the majority of the participants, the Clustering algorithm did not recognize most of the data above 3% in class 2. The participants also reported in the questionnaires an overall decrease in the stress felt. Therefore, it is possible that the proposed protocol did not induce the desired emotional state (stress) in the participants.

The developed multimodal system is functional and can be utilised in other studies with emotional markings gathering.

Contents

List of tables

List of illustrations

1 Introduction

1.1 Motivation

This dissertation was developed in the context of IEETA's research activities, related to the integration of the field of psychology with recent computational technology. Multiple works have already been done within this group, namely, a web platform to streamline psychology questionnaires [1], a mobile phone app that assists in studies of psychophysiology [2], among others.

Since the field of psychology is the science of behaviour and mind, a part of it is dedicated to the study of emotions. Emotions have a major impact on how an individual experiences a certain situation and influences how a person feels, thinks and acts. One of the major components in the study of emotions is the classification of the emotional state of an individual. However, according to Jaak Panksepp, it is difficult (and some even claim impossible) to measure directly the emotional state of any living being. This is due to currently not existing any measures that can unambiguously quantify changes in the emotional state of a living creature [3]. All the current methods of emotion classification rely on empirical data in order to predict the emotional state being felt by a given individual. Amongst others, user-filled questionnaires or ECG are some of the examples of the data acquisition methods currently used.

Multiple computer science fields have emerged due to the vast improvements in technology, predominantly computers, that have taken place in the last decades. One of these fields was Affective Computing, which consists of utilising computers to either detect or emulate emotions [4]. In order to detect emotions, the computational methods currently used rely on Machine Learning. We believe that the integration of Affective Computing in the field of psychology can add value by improving the quality of emotional data gathered. However, as with most emerging fields, few methods of affective classification are yet to be considered reliable. We also believe that further exploring, developing and analysing these methods can help in the advancements of this field.

The primary motivation of this dissertation is to develop a system in the field of Affective Computing and analyse its affective data gathering methods in the scenario of software development. Software development has two different interesting characteristics: the fact that it is considered by many a stressful job, which render the data collected valuable; and the fact that the developers have low mobility and have constant contact with sensors (such as the keyboard and the mouse), which facilitates the collection of accurate data. Two more aspects also weigh in the motivation: the focus on non-intrusive methods (such as data from keyboard or mouse usage) and designing a multi-modular system. The former is due to non-intrusive methods having a higher probability of the developer being willing to use them, while the latter having a potential higher accuracy than uni-modular systems [5].

1.2 Objectives

The primary goal of this dissertation is to develop a system that collects affective data using a multimodal approach that we believe enhances the precision of emotion classification algorithms. In order to achieve this, an investigation will be conducted beforehand to identify the current solutions being used as well as the corresponding benefits and issues. Additionally, another goal is to develop the system in a modular architecture in order to allow integration with potential external systems. Despite the system being able to be used in multiple different scenarios, the scenario for the proof of concept will be the affective data gathering from software developers. Since software developers have a considerable amount of contact with the mouse and keyboard, the dissertation will also have an increased focus on methods using these sensors.

1.3 Structure

The dissertation is divided into seven chapters. The **first** chapter is the Introduction (the current one), where an overview of the document is presented. The **second** chapter is the state of the art, where the first stage of the dissertation, research, is presented. This research presents the theme of affective computing and the current methods utilised as well as current literature. This chapter also includes the results of the processing of a pre-existent dataset. The **third** discusses the use cases and requirements for the system. The **fourth** chapter presents the system architecture and describing each subsystem. The **fifth** chapter reports which equipment and software were utilised or developed in order to implement the architecture along with the reasoning for selecting it over the existing alternatives. The **sixth** describes the study conducted to test and validate the system alongside with the results obtained from it. Also included in the chapter is the analysis of the results obtained. The last chapter, the **seventh,** presents the final thoughts on the dissertation and along with possible scenarios for future work.

2 State of the art

In this chapter, the basic concepts relevant to the dissertation will be reviewed alongside with the systems that have been utilised in the literature. Since the dissertation is situated in the field of *Affective Computing*, the chapter will begin by introducing this field and analysing its concepts in 2.1. In 2.2, the computational methods utilised in the literature for the recognition of emotions will be presented and discussed. Finally, in 2.3, an exploratory study to test emotion classification algorithms on a pre-existent dataset will be presented.

2.1 The emergent field of Affective Computing

Affective Computing is a branch of computer science which was introduced by Rosalind Picard in an MIT report titled "Affective Computing" [4]. In the report, Picard describes affective computing as computing that relates to, arises from, or influences emotions. Affective Computing is the simulation of empathy, which can be divided into two major areas: the detection of emotions and the emulation of emotions, with the former being the focus of this dissertation [4].

In the last decades, a significant improvement in computer science and information technologies has taken place, which led to rapid development in areas such as the internet, computers, portable devices, among others. Since the final purpose of these technologies is mostly to improve human lives, collecting and analysing affective data plays an important role in both improving the existing technologies and developing new ones. While still not being as widespread as other popular technologies, affective computing has been showing an increase in its popularity in the past decade. This is evidenced by the number of companies that have emerged in the marketplace such as *Kairos*[1] (specialises in face recognition), *CrowdEmotion*[2] (focuses on humanising technology), *Affectiva*[3] (provides insight into consumers emotional engagement), *Beyond Verbal*[4] (specialises in evaluating human voice), *PointGrab*[5] (provides smart sensor solutions to the building automation industry), and others[6].

Most affective classification systems follow a two-stage process which begins in the collection of data from one or multiple sensors and is followed by the extraction of meaningful patterns from the collected data.

2.1.1 Applications of affective computing

While researching for the applications of affective computing, we analysed both the applications suggested by and utilised in the literature as well as the applications being currently used

[1] https://www.kairos.com/
[2] https://www.crowdemotion.co.uk/
[3] https://www.affectiva.com/
[4] https://beyondverbal.com/
[5] https://www.pointgrab.com/
[6] https://www.predictiveanalyticstoday.com/what-is-affective-computing

by the companies. Affective methods are being applied in multiple domains such as intelligent tutoring systems, adaptive interfaces, gaming, software engineering, designing web-pages, cinematography, augmented and virtual reality and others [6].

One example that explains why Affective Computing is gaining popularity and becoming increasingly relevant is that some emotional states of employees can have a negative impact on the company. For example, it is estimated that work-related depressions cost €617 billion a year to European enterprises [7]. One of the factors that weigh in this number is that detecting emotions relies mostly on user-filled questionnaires and hormonal tests of samples (like saliva or blood). The former is very subjective, requires the full cognitive attention of the user and is affected by memory while the latter is heavily affected by circadian rhythms, intrusive and has a costly and slow analysis [8]. In comparison to the current alternatives, an affect-sensitive system has the potential to mitigate most of these issues by having the ability to: be objective (the data can be obtained directly from biometric sensors); be non-intrusive (sensors like webcams or keyboard and mouse are considered by many as non-intrusive [6], [7], [9]–[12]), more affordable than the conventional methods of diagnosis and having the ability to automatically quantify symptoms even before the patient reports them [8].

Another scenario that showcases the relevancy of Affective Computing lies in the field of stock market trading [13]. The psychiatrist Dr Alexander Elder mentioned in his book [13] that it is possible to predict whether or not the price of a particular stock would rise or fall based on the general mood of the crowd that engages on the trading process. In order to buy and sell stocks in the old days, one had to be physically present and, due to this, traders would utilise the noise from the crowd to evaluate the general mood and predict if the traders were *bearish* (i.e. stock price would fall) or *bullish* (i.e. stock prices would rise) on that day. With the development of the internet and computers, it is now impossible to evaluate those crowds like in the old days. However, nowadays traders have access to multiple data (such as the number of buy/sell orders, trade volume, price history, and more) and multiple algorithms that process that data have been developed in order to calculate the general mood of the crowds. These algorithms are used to assist in the buying/selling decisions.

In addition to these examples, some companies already provide multiple products in the affective computing field. The following list denotes a few examples of products from the most popular companies:

- Analysis of the human interacting with a system, with the purpose of assisting in its improvement;
- Affective data collection and analysis using facial features;
- Content optimisation;
- Emotional data gathering from the body language;
- Listening for arousal and stress (i.e. arousal and stress alerts);
- Affective data collection and analysis from the human voice;
- Analysis of affective data collected from workers of an industry in order to evaluate if some tasks are worth being automated;
- And others.

In the current literature, a vast number of scenarios and implementations have also been considered. In [14], a study was conducted to identify a relationship between mood disorders and mobile phone keyboard activity on subjects with bipolar disorders with the aim to assist in the detection and monitoring of mood disturbances. The authors concluded that *"mood states in bipolar disorder appear to correlate with specific changes in mobile phone usage"* [14] and that models similar to the ones presented may be useful in diagnosing and leading to a better understanding of those disorders. In [15], eye-tracking software was used in order to detect arousal from computer users through the pupillary response. The authors found a moderate correlation between the results obtained by the system and the data self-reported by the subjects. It is their opinion that despite not having a strong correlation, *"eye trackers can serve as a multi-sensory device for measuring arousal"* [15]. In [16], pressure data from typing in smartphones was captured in order to extract cues of stress in the user. A correlation between pressure and self-reported stress was found. In [17], a multimodal system was utilised that combined information from facial expressions recognition and speech prosody feature extraction. The system was deployed on a database containing audio and video data of simulated human-human discourse. In [18], the data collected from mouse movements and keystroke dynamics was processed in order to detect emotion in real-world learning scenarios.

Finally, and the most relevant application for this dissertation, is the scenario of stress detection using continuous data collection gathered from computer users, especially software developers. Software Development is a profession where the developers spend most of their time utilising a computer and interacting with its components, mainly the keyboard. This fact renders the continuous collection of data from the keystroke patterns able to produce a considerable amount of data while not being intrusive.

2.1.2 Emotion Recognition Systems

Before exploring how emotions are detected, it is important to understand how emotions are described and/or quantified. Currently, there are two main approaches to the description of emotions.

One was popularized by Paul Ekman in [19] and is a categorical approach consisting on defining a set of basic emotions, (emotions that can be combined to form another emotion, e.g. happiness or contempt) and complex emotions (emotions that are the combination of basic emotions, e.g. smugness, which is a combination of happiness and contempt).

The other is a dimensional approach [20] that describes the core affect (commonly described as *feeling*) as a "single integral blend of two dimensions, hence describable as a single point". The horizontal dimension measures the pleasure-displeasure (valence) where at the low extreme has emotions like agony and at the high extreme has emotions like ecstasy. The vertical dimension measures the arousal and ranges from emotions like sleep and drowsiness (low end) to emotions like alertness and frenetic excitement (high end). Figure 1 is a visual representation of these orthogonal axes.

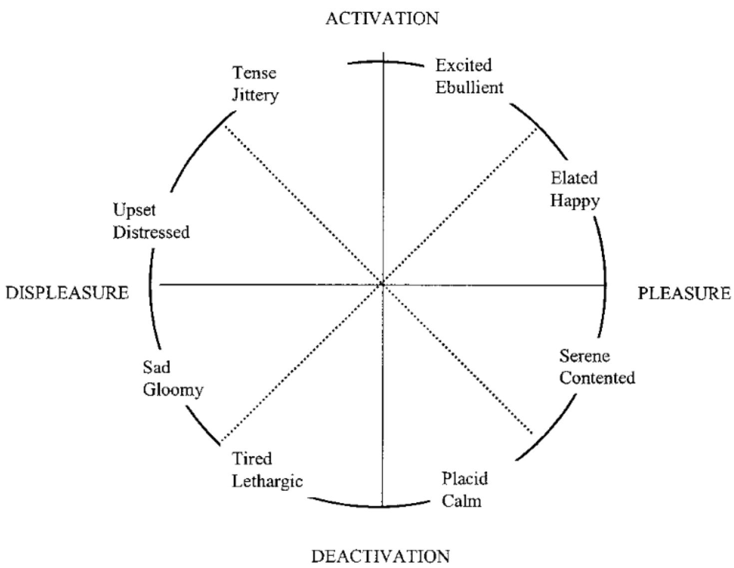

Figure 1 - Valence/Arousal Orthogonal Axes (source: [20])

2.2 Computational methods for affective data gathering

In this subsection, the methods currently used for affective data gathering will be explored and discussed.

2.2.1 Emotions and stress recognition based on physiology metrics

In the paper [6], the acquisition methods were categorised in the following categories: visual, audio, textual, physiological and input devices. The paper also included another category, that consisted of the combination of multiple methods, which will be further explored in 2.2.3.

Visual methods are mainly referred to as FER (Facial Emotion Recognition), which consists of the processing of image or video data. These methods' study has been increasing rapidly due to the development in areas such as human-computer interaction (HCI), virtual reality (VR), augmented reality (AR), advanced driver assistant systems (ADASs) and entertainment. Conventionally, FER has three main steps: face and facial expression detection (which consists on the detection of the facial area and the marking of facial components like the eyes and nose), feature extraction (which consists on the extraction of the spatial and temporal features of the facial components) and expression classification (which consists on the use of pre-trained classifiers such as a support vector machine (SVM), AdaBoost and random forest to produce the recognition results) [21]. While not being as popular, other visual methods apart from FER also exist, such is the case of the measure of one's posture.

Audio methods consist mainly on the processing of speech recordings. Most researchers have used global suprasegmental/prosodic features (such as stress tone, word juncture that accompanies or is added over consonants and vowels) whose mean, standard deviation, maximum and minimum of pitch contour and energy in the utterances (smallest unit of speech) would be calculated and used on the classifiers [22].

Textual methods consist in analysing the semantics of the speech and the written text [6].

Physiological methods consist of the measure of biological data. Most methods on this category require a sensor, thus being considered intrusive. Common methods are the measure of the heart rate, temperature, skin conductance, blood and saliva samples and fEMG (facial electromyography).

Input Devices' methods consist of any interaction with an input device. The most common devices used are the mouse, keyboard and touch screen but some other devices, like the driver wheel, can also be used.

2.2.2 Emotional states recognition based on keyboard and mouse usage

Since this dissertation focuses primarily on gathering affective data from software developers and other employees that use the computer for extended periods of time, the most interesting input devices are the mouse and the keyboard. Like most input devices, these are considered non-intrusive [11] while providing a substantial quantity of data due to being constantly used by the actor (the person whose affective data is being gathered).

Multiple studies ([6]–[8], [10], [11], [18], [23]) have been conducted using mouse and keyboard utilisation patterns to detect emotions, with mouse interaction not being as widely applied [18]. The general setup utilises a regular mouse and keyboard, and software that records the input from those devices, also known as *mouse loggers* and *key loggers*. In some scenarios (such as in [8]), a capacitive mouse and keyboard capable of detecting pressure were used. In order to validate the developed systems, the researchers conducted studies with participants. The majority of these studies consisted of the participants performing tasks in the computer while their interaction with the input devices was being recorded. Most of these tasks ranged from the transcription of an already existent text or audio tracks, developing software, writing about a specific topic and researching about a certain subject while writing the conclusions obtained.

The features selected and analysed in the studies are represented in Table 1 (keyboard) and Table 2 (mouse).

Keyboard Feature	Description
Keystroke Latency	The duration between a key release and the key press of the next keystroke.

Keystroke Duration	The duration between the key press and the key release of the same key.
Pressure	The amount of pressure applied in the key.
Typing Speed	The total number of keystrokes per a defined unit of time.
Error Rate	Percentage of error keys pressed (keys that indicate the writer made a mistake, e.g. backspace or delete).
Capitalization Rate	Capital to lower case ratio.
Special Key Rate	Special Keys ratio (e.g. spacebar or enter).
Key Frequency	Frequency of keystroke of all the keys.
Duration of sequences of strokes	The duration between the first keypress and the last key release of a given sequence (e.g. diagraphs, trigraphs, etc).

Table 1 - Keyboard Features utilised in the Literature (adapted from [18]*)*

Mouse Feature	Description
Position	The position over time.
Velocity	Speed and direction of movement.
Precision	The relation between the distance of two events and the actual covered distance.
Distance	Total distance covered.
Acceleration	Speed variation over time.
Mouse buttons interaction	Mouse buttons' press, release and scroll logs.
Inactivity	Rate of no mouse interaction.

Table 2 - Mouse Features utilised in the Literature (adapted from [18]*)*

The current literature supports that the mouse and keyboard utilisation can be a valid source of data to detect and classify emotions, with the keyboard having a greater focus. Despite not existing sufficient data to validate the detection of every emotional state, the emotional state of stress, in particular, shows good signs of being well detected within these methods. This statement is supported by, for example, the overall accuracy of 89.5% obtained in [23], all the algorithms in [7] having an accuracy above 70%, and the fact that in [8] above 83% of the participants had an increased typing pressure with increased levels of stress.

2.2.3 Multimodal approaches

Since there are multiple methods to gather affective data, it is an option to use a multimodal approach.

A review paper on methods based on mouse and keyboard patterns [6] stated that these methods do not have the required accuracy for real-life applications but they can be treated as an additional source of information to be paired with other methods creating "reliable emotion recognition systems".

In [22], a bimodal approach was developed using facial expressions and speech. The results reported an improvement of almost 5% (absolute) when comparing to the performance of the facial expressions. It is noteworthy that the facial expressions' methods had higher performance than the speech ones in emotion classification but still benefitted from the integration.

A review paper on multimodal systems [5] studied ninety contemporary multimodal approaches and performed an analysis comparing their accuracy to the unimodal alternatives. Over 85% of the studies incurred an improvement on the accuracy greater than at least 1%, supporting the premise that multimodal methods outperform unimodal ones. However, multimodal improvement was almost three times lower when trained on naturalistic data (4.59%) when comparing to trained data (12.7%) and it is the authors' opinion that the trained data is less relevant in comparison to the naturalistic due to "the ultimate goal of affect detection is to sense naturalistic affective expressions" [5] and that only the improvement of 4.59% should be considered. It was also stated that despite having higher complexity, multimodal approaches provide higher fidelity models (more accurate affective representation) while addressing missing data problems that impact greatly unimodal ones. Additionally, it was observed that the performance increase of multimodal methods whose different modalities conveyed similar information was lower than those whose modalities were more dissimilar.

After analysing the literature, it can be concluded that multimodal approaches, despite having the problem of increased complexity, have the following advantages:

- Increased rate of accuracy;
- Higher fidelity models of affect expression;
- Ability to address missing data problems.

2.3 An exploratory study of SWELL dataset

In order to select the best methods of data classification, an empirical study was conducted utilising an already produced, classified and analysed dataset. The dataset utilised was the *SWELL dataset* [24], [25], which was supported by the *Dutch national program COMMIT*. The decision to choose this dataset was based on: the already existent literature analysing it [7], providing more data to compare the results to; the fact that the dataset reflects real-life scenarios and stressors where the workers are performing their natural office work, which we believe renders the dataset more accurate in comparison with datasets not using real-life scenarios; and having multiple sensors (being multi-modal).

The dataset consisted of data collected from twenty-five participants, which were monetary compensated for study participation. Additionally, participants were told that the compensation would escalate with how well they performed during the study. Each participant was randomly assigned one task that involved writing an opinion about a topic and another task that involved researching and writing facts about another topic. The participants were subjected to different workloads and different stress levels and, in total, three conditions were simulated: neutral, which consisted in the participants having no stressors or time limit to perform the tasks; stressor "time pressure", which consisted in the participants having two-thirds of the time utilized in the neutral condition; and stressor "interruptions", which consisted in the participants receiving e-mails that could be relevant to the task, require response or both.

Data from multiple sensors was collected, namely **computer logging**, which utilised a key-logging application; **video from the participants' face and upper body**, which was recorded with a USB camera and analysed with a facial expression analysis software; **body posture**, which was recorded using a *Kinect 3D*; **body sensors**, which recorded two different types of data: ECG and skin conductance; and additional lab recordings, which consisted of ceiling camera recordings and microphones recordings. In addition, the dataset also included pre-processed data consisting of the minute-level feature extraction, which consisted of the sum of certain events per minute.

For the empirical study, we utilised the minute-level data of the keyboard, mouse and application usage. The variables from the dataset we analysed represent the number of times a particular event occurred within a minute and are represented in Table 3.

Variables	Event	Type
SnMouseAct	Mouse interactions.	Mouse
SnLeftClicked	Left mouse button clicks.	Mouse
SnRightClicked	Right mouse button clicks.	Mouse
SnDoubleClicked	Double mouse clicks.	Mouse
SnWheel	Mouse scrolls.	Mouse
SnDragged	Mouse movement while pressing a button.	Mouse
SnMouseDistance	Cursor distance.	Mouse
SnKeyStrokes	Keyboard keystrokes.	Keyboard

SnChars	Characters keyboard keystrokes.	Keyboard
SnSpecialsKeys	Special keys (e.g. *alt* or *shift*) keyboard keystrokes.	Keyboard
SnDirectionKeys	Direction keys keyboard keystrokes.	Keyboard
SnErrorKeys	Errors keys (e.g. *backspace* or *delete*) keyboard keystrokes.	Keyboard
SnShortcutKeys	Shortcut keys keyboard keystrokes.	Keyboard
SnSpaces	*Spacebar* keyboard keystrokes.	Keyboard
SnAppChange	Application changes.	Applications
SnTabfocusChanges	Tab focus changes.	Applications

Table 3 - Variables used in the Exploratory Study

To classify the data, we utilised the *random forest* algorithm [26] with three classes representing the three different conditions: C1 (neutral), C2 (time pressure) and C3 (interruptions). The algorithm was trained using the *train/test split* technique with data from twenty participants, which represented 80% of the data, to train the algorithm and the remaining five participants' data (20%) to be classified. The data was randomly selected. Afterwards, the average of the results obtained was calculated. The confusion matrix is represented in Table 4 and the performance variables are represented in Table 5.

	C1	C2	C3
C1	1345	13	5
C2	223	800	9
C3	206	59	763

Table 4 - SWELL Random Forest Results (Confusion Matrix)

Class	Accuracy	Fscore	Sensitivity
C1	86.94%	85.75%	98.68%
C2	91.12%	84.03%	77.52%
C3	91.85%	84.54%	74.22%
Global	89.97%	84.95%	84.95%

Table 5 - SWELL Random Forest Results (Performance Variables)

Globally, an accuracy of 89.97%, a FScore of 84.95% and a sensitivity of 84.95% were achieved. Despite this, the algorithm had a bias towards classification as C1 (neutral). However, it is possible that the participants might have had moments with a neutral emotional state under the scenarios where stress was being induced (C2 and C3). Since these results were satisfactory, we decided to use the *random forest* algorithm in the system as one of the classifiers of stressed emotional states.

3 Use cases for the affective data acquisition system

This section will present the requirements as well as the use cases defined before the implementation of the system we developed, which we named *IEETA Collecta*. Both the requirements and the use cases were collected and defined from the analysis of the state of the art and after discussing with the University's professors, students and software developers.

3.1 Measuring stress in software developers

One goal of the system is to be flexible in order to support multiple application scenarios. Examples of these scenarios include the gathering of stress data from software developers, the evaluation of the emotional response that films and video advertisings generate in the audiences, the gathering of affective data while testing online stores website interfaces, and others. These scenarios have in common the need to record data from a certain input, the need to extract features from the collected data and the need to perform an "affective classification" on the extracted features. Due to this, the system has to be implemented in a modular fashion where it is possible to integrate with different external systems by enabling data to be transferred between them or, in some cases, have the external system replacing a module of the system (e.g. different sensors, different emotion detection algorithms and others).

Despite supporting different scenarios, in this dissertation, the system will be defined and implemented taking into consideration the scenario of the detection of stress among software developers. Multiple factors weighted in our prioritisation of this scenario such as the fact that most software developers spend most of their working time in contact with input devices like the mouse, keyboard and staring at a computer screen, thus facilitating the collection of data; the better understanding of software development we already possess due to being related to our academic field; the ease of gathering participants from our department that are in the same field and that can validate the system; and the fact that all the sensors required were already available in our department.

Software Development is an activity heavily reliant on the utilisation of computers and thus sharing the same health issues as other computer-related activities. Computer stress is one of the most common disorders, along with carpal tunnel syndrome, computer vision syndrome and musculoskeletal problems [27]. Moreover, a study [27] has shown that there is a correlation between the prevalence of these disorders and the frequency of computer usage alongside the fact that individuals that are employees (such as software developers) experience almost all of the four health problems. Another fundamental aspect is the fact that prolonged stress affects both the health of the developer and the quality of the software they produce for the company, which emphasises the importance of addressing such emotional states [28].The first step towards preventing stress consists of being able to detect when a person is stressed. Afterwards, multiple approaches could be taken such as alerting the person in order to encourage behavioural changes or automatically informing a friend of the situation and asking them to send a funny message text, among others [29]. Another situation where the detection of stress is relevant is in assisting the diagnosis performed by health professionals. The system will make available extra data to the professional that can be used to obtain an enhanced precision in the classification. This data will have multiple advantages such as being collected in real-time (which does not rely on self-reporting methods that are reliant on memory and are subjective), having the possibility of being recorded through an extended period of time (which enables the collection of a vast amount of information), and also having the possibility of not interfering with the subject (non-intrusiveness) while the gathering process is taking place.

In order to develop software, the developers have to interact with a computer. The most common forms of interaction are by writing in a keyboard, utilising a mouse and viewing graphical elements in the monitor. Since the monitor is an output device, it can not act as a sensor in order to collect data from the developer. However, the mouse and the keyboard are input devices and it is possible to generate data from reading them while the subject is developing software. Another advantageous fact is that this data collection can be done non-intrusively because the subject does not have to interact differently with the devices as he normally would while developing. Developing software is also an activity where the user remains physically stationary which makes viable the recording of video from the face and upper body of the developer. However, some people regard this data acquisition method as intrusive whereas others consider it to be non-intrusive [11]. Finally, including ECG sensors can also add value to the system in the scenaros where an increased classification accuracy is required or when there is a need to valide other methods. It is noteworthy that we consider the utilisation of ECG sensors to be intrusive.

The data to be generated from the utilisation of the keyboard should be: what keys are being pressed; the delay between pressing and releasing the same key; and the delay between different keys. As it was evidenced in the literature [8], pressure information would also be beneficial but we do not have access to hardware capable of recording those metrics. The data gathered from the mouse should be: button interactions; mouse movement speed; and mouse movement direction. The video camera should generate data in video format representing the recording of the face and upper body. Finally, the data from the ECG sensors should be its values over time along with the Heart Rate (HR) and the interval between two R waves (RR).

In order to gather data in considerable amounts, the gathering process should be continuous. To achieve this, the system must use sensors that do not interfere with the process of software development, i.e. non-intrusive sensors. However, sensors considered intrusive can also be used, but the system must keep the possibility of non-intrusive data collection.

In some scenarios, the software developer may have the need to write or read sensitive data (such as write passwords or visualise sensitive e-mails, for example). It is important that the privacy of the developers is preserved. To achieve this, the system must provide the developer with a simple and fast way of toggling the screen and keyboard recording. Another solution for the keyboard data acquisition is not recording the context (i.e. which keys are being pressed) and only record the events (i.e. press and release of unspecified keys). The second solution could also be enhanced by providing the developer with the option to toggle the recording.

It is also required that the system supports integration with external systems. In order to achieve this, the relevant data generated by its components should be made available to external systems and the system must also have the ability to read data generated by an external system. This can be done by implementing the system in a modular fashion, where the interactions between its modules have interfaces that can be utilised by an external system. An example of this can be having a module from the system read/write from/to a separate file.

3.2 Use Cases of IEETA Collecta

Before the definition of the use cases, the entities that would interact with the system (actors) were defined. Four actors interact within the system and are described in the following list:

- **Software Developer**: user of the system whose behavioural and physiological data is being gathered while he/she is programming or performing another task in the computer. Despite being the most relevant actor, the software developer's interactions with the system are primarily passive;
- **Data Analyst**: user of the system that consumes (either by viewing through the display or by downloading) the data collected and processed by the system. The data analyst can either be a human analysing the data or an automated system such as a listener that sends an alert when the Software Developer is getting stress, for example. When consuming the data through the display, the Data Analyst can also provide additional data in the form of markers (notes that mark an event in a specified timestamp);
- **Administrator**: user of the system responsible for the management of the system, primarily the setup of the data gathering process and the maintenance;
- **External System**: any system that interacts with one or more parts of the system. The external system can both replace the functions of an existing subsystem or consume data without change the topology;

The system was also divided into two subsystems, each performing different tasks: **Data Logger**, which is responsible for the gathering of data from the sensors and its processing using computational methods; and **Data Viewer**, which performs the display of the data previously acquired to the Data Analyst as well as enabling the addition of extra data markers (e.g. markers identifying specific events that occurred during the data acquisition). It is also noteworthy that an External System can act as any of the previous subsystems by either replacing a part of the entirety of its tasks. We considered this division relevant in order to achieve a modular implementation and, consequently, each subsystem will be treated as a separate system while describing the use cases.

A visual representation of the Data Logger's use cases can be found in Figure 2, along with the use cases' description listed in Table 6.

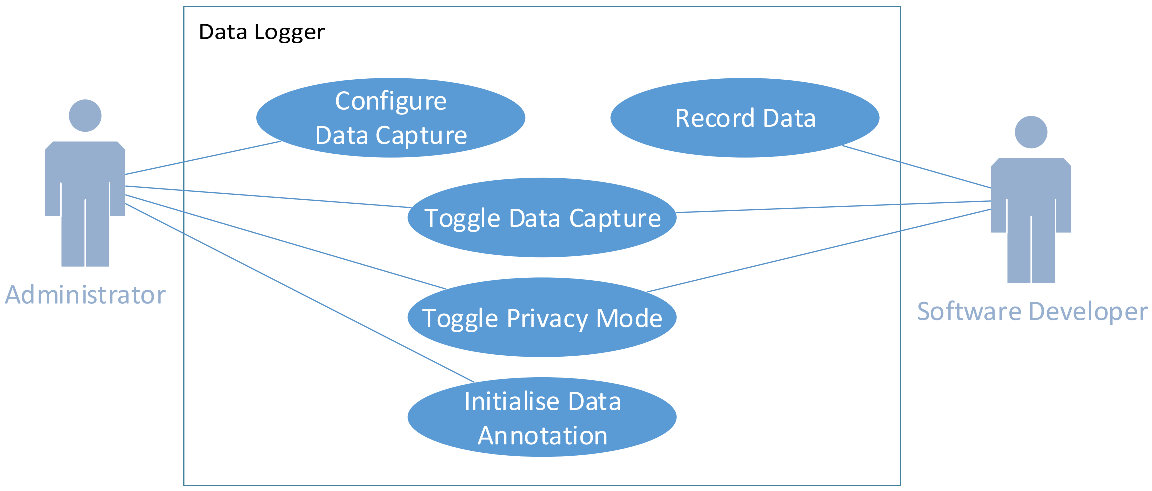

Figure 2 - Data Logger Subsystem Use Cases Diagram

Use Case	Description
Configure Data Capture	Set up both the hardware (apply and connect the sensors) and the software (configure the parameters) part of the gathering raw data process.
Record Data	Capture affective data from the Software Developer.
Toggle Data Capture	Turn on or off the recording of raw data.
Toggle Privacy Mode	Turn on or off the private mode. Privacy Mode consists of not recording any sensible data (e.g. passwords).
Initialise Data Annotation	Configure the parameters and start the post-processing and filtering of the raw data.

Table 6 - Data Logger Subsystem Use Cases.

A visual representation of the Data Viewer's use cases can be found in Figure 3, along with the use cases description listed in Table 7.

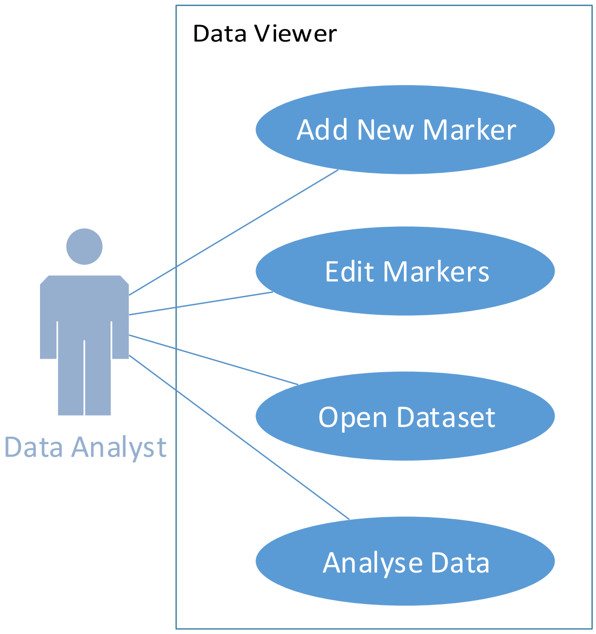

Figure 3- Viewer Subsystem Use Cases Diagram

Use Case	Description
Add New Marker	Add a marker in the desired point in time and save it to a repository.
Edit Markers	Remove or change the text and timestamp of one or multiple existent markers.
Open Dataset	Select a folder containing data to display originated from either the Data Logger subsystem or another external system.
Analyse Data	Perform all the tasks related to the consumption of data: view the data; hide irrelevant data (filtering); and manipulate the view (changing the focus of the view window).

Table 7 - Viewer Subsystem Use Cases Description

4 System Architecture

In this chapter, the overall architecture of the *IEETA Collecta* system will be presented including the modules that are part of it and how they interact with each other and possible external systems.

4.1 Architecture Overview

There are two different main functions that the system must perform: the acquisition and processing of data and the consumption/display of data. Due to this, the system was divided into two subsystems that perform each of the functions, respectively: **Data Logger** and **Data Viewer**.

The former can be further divided into two tasks: the acquisition of data, which incorporates the sensors (hardware that converts biometric signals to digital data) and the recorders (software that reads data from the sensors and outputs it in a format that can be read by the other modules); and the processing of data (i.e. data annotation) which performs the tasks of generating complex data (data obtained from processing existing data), injecting it into the raw data and filtering the undesired data for the scenarios defined in the parameters. The task of generating complex data consists of reading atomic variables (variables that are not generated from other variables) or, in some cases, complex variables and then calculating other variables. Examples of complex variables can be the velocity and direction of mouse movement, which is calculated using the atomic variables that represent the coordinates over time; the delay between key presses alongside the delay between pressing and releasing the same key in the keyboard, which is calculated from the atomic variables of key presses and key releases over time; or the Heart Rate (HR), which can be calculated from the atomic variables that represent the values of the ECG over time.

The latter reads the data (either from the system or another external system), displays it through the means of a visual interface and handles the operations related to annotations (i.e. markers).

Figure 4 is a visual representation of both the subsystems, the modules and the interactions between each other.

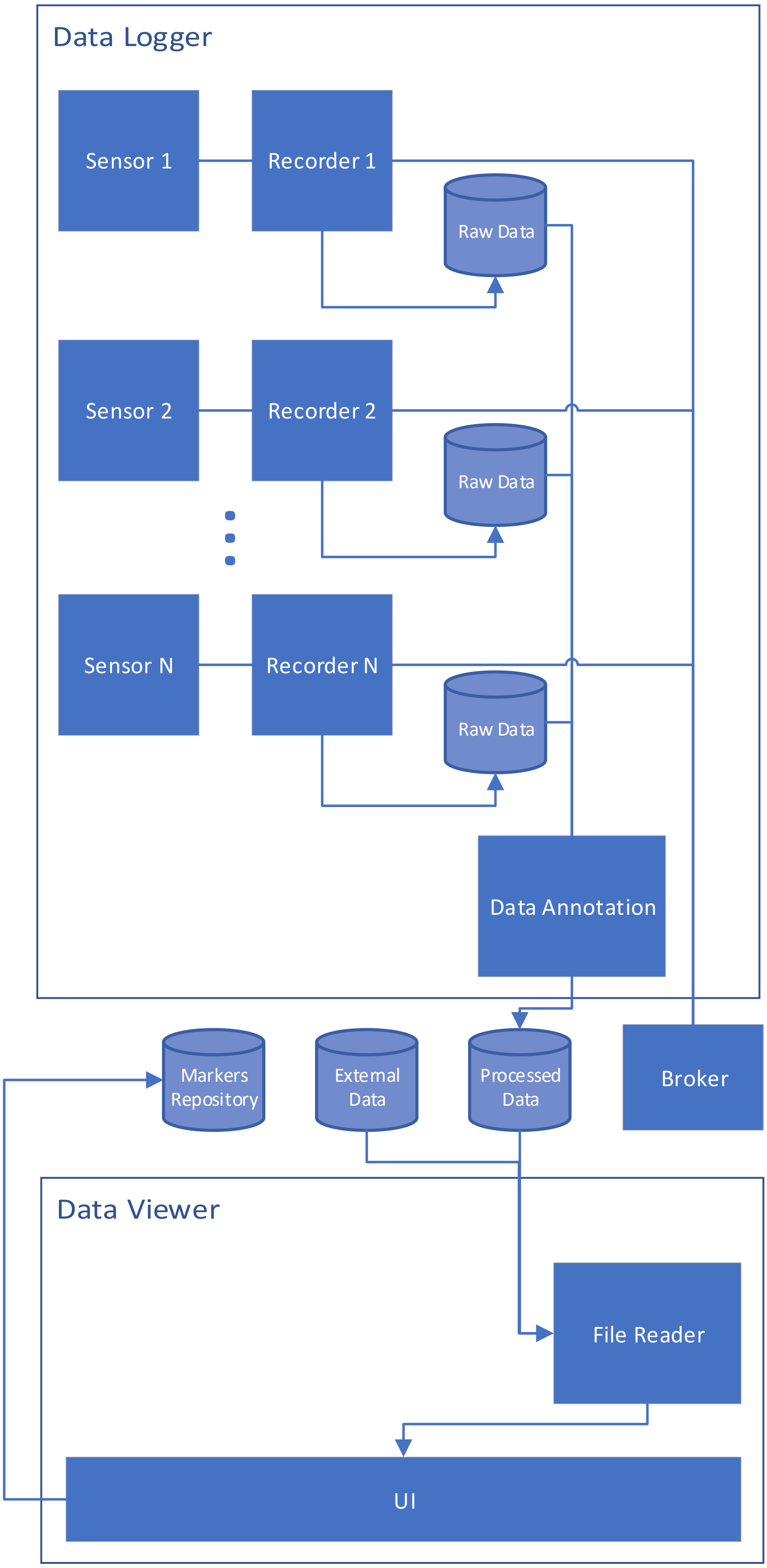

Figure 4 - IEETA Collecta's Architecture Overview

The general flow of information is the following: the **Sensors** convert the data from the subject into digital data, which is then accessed by the **Recorders** and pre-processed. Afterwards, the **Data Annotation** performs the task of calculating and injecting complex data, followed by the task of filtering undesired data and formatting into a format that is adequate for the parameters defined by the Administrator. This data is then read and synchronised by the **File Reader** of the Data Viewer and displayed in the **UI**, which also stores the data from the markers set by the Data Analyst.

Each of the modules present in the visual representation of the system are described with further detail in the following list:

- **Sensor**: Piece of hardware that measures a physical property (in this case, a biometric variable) and converts it to a digital format. Multiple sensors exist in the system, each belonging to a different acquisition method. Examples of sensors can be a mouse, keyboard, video camera, among others;
- **Recorder**: a Desktop application that reads data from a sensor, pre-processes it (e.g. format the data or add new data such as timestamps) and outputs Raw Data into a file or converts it to a message format and sends it to a broker. Multiple recorders exist, one for each sensor;
- **Data Annotation**: a Desktop application that reads Raw Data produced by a Sensor, calculates complex features, formats and filters into an adequate format (e.g. format adequate to process the data in Matlab), and outputs the Processed Data to a file. The Raw Data from every acquisition method is processed in a single Data Annotation module;
- **File Reader**: Software module designed to read and synchronise all the previously collected and generated data. This module also supports data from external systems, as long as the format protocols are respected. It is only in this module that the data from all the separate methods become unified and cease to be separate black boxes. This module also serves as a back-end to serve data to the UI;
- **UI**: Graphical interface that displays the previously synchronised data to the Data Analyst along with view manipulation features (e.g. zoom, change timeframe and others) and enables the manipulation of the markers. Changes performed in the markers are also stored in a separate file, Markers Repository, and can be accessed by an external system;
- **Broker**: Server that receives data in the message format from the publishers in the system and transmits them to the external systems that have subscribed to them. The Broker is the core of the interaction by messaging, which is a real-time alternative to file sharing.

4.2 Interactions between Modules

While the software is running, its modules require the sharing of information between each other. The flow begins with the raw data generated in the Data Logger module that consists of the atomic variables that were read from the Sensors by the Recorders. The raw data consists of the logs from the keyboard, mouse, ECG sensors and video recordings. This data is then received by the Data Annotation module (or, alternatively, an external system). Afterwards, the Data Annotation module utilises the raw data to calculate complex variables for each of the data types that require so (keyboard, mouse and ECG data) while filtering undesired data generating processed data. This data is then read by the File Reader which converts it into a format that can be displayed in the UI. When the Data Analyst is viewing the data, the marker database can be manipulated. This marker database consists of a text file containing a list of labels associated with a timestamp and only the UI reads and writes from it (apart from external systems that access the file). Figure 5 is a visual representation of this flow.

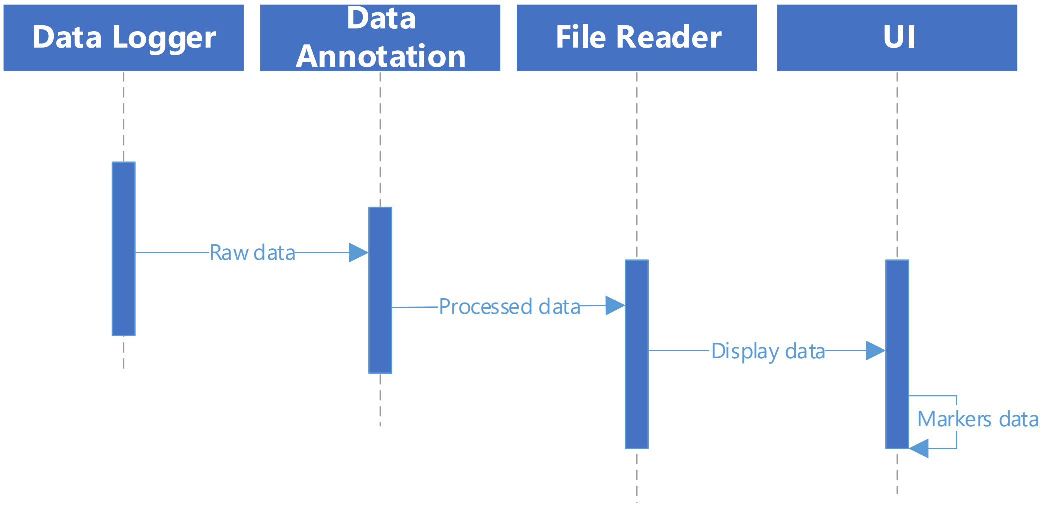

Figure 5 - UML sequence diagram of IEETA Collecta

The following list presents the data produced within the system:

- **Raw Data**: Output in file format produced by the recorder software after logging the data from the sensor(s);
- **Real-time Raw Data**: Stream of data, usually in message format, produced by the recorder software while recording data from the sensor(s). The most striking difference from the previous type is the fact that, as it is being collected, this data is provided in real-time. This type of data is not utilised within the system and is destined to be consumed by external systems;
- **Processed Data**: Output in file format produced by the Data Annotation module after processing the Raw Data;
- **Markers Data**: File containing the markers generated by the Data Analyst as well as the respective timestamps.

The types of communication that will take place are represented in Table 8.

Source	Type of data	Receiver(s)
Data Logger	Raw Data	Data Annotation, File Reader and External System
Data Logger	Real-time Raw Data	External System
Data Annotation	Processed Data	File Reader and External System
File Reader	Display Data	UI
UI	Markers Data	UI and External System

Table 8 – Types of communication in IEETA Collecta

In some scenarios, it is required to share data while it is still being collected (real-time) while in others, that requirement does not exist. Due to this fact, two alternatives for interaction between modules are to be implemented: Messaging, for when real-time data is required, and File Sharing, for any other situation. Having these two alternatives, each optimised for its situation, does increase the overall efficiency while not disregarding the requirements.

4.2.1 File Sharing

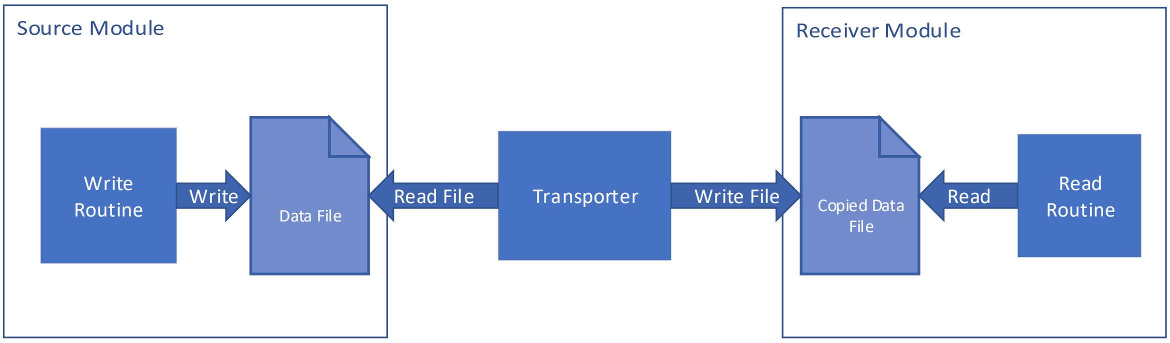

Figure 6 – Visual Representation of File Sharing

This method of interaction consists in having the source module writing a file with the information to be read by the Receiver. An entity is in charge of transporting (Transporter) the file to a location where it can be read by the Receiver. In some scenarios (e.g, where the source and the receiver have access to the same file system), a Transporter may not exist and the receiver can read the file directly from the location where the source stored the file. It is noteworthy that the Transporter could either be a computational system or a human manager.

Advantages:

- Lower computational resource usage;
- Not required to have multiple components running simultaneously;
- Faster communication and simplicity of implementation when the Sender and the Receiver are on the same device (due to the lack of need for a Transporter);

Disadvantages:

- No support for real-time data; *
- The Transporter has the potential to be complex;

* In the specific case where both the Sender and the Receiver are on the same device, it is possible to have both systems write and read from the same file and thus have real-time data. However, this approach is not efficient, more prone to errors and inferior to the Messaging alternative.

4.2.2 Messaging

This method of interaction consists in converting data to a message format in the Source, which is delivered to a broker who will forward to a Receiver. The message will then be converted into a format that can be used by the Receiver. The pattern publisher-subscriber is to be used.

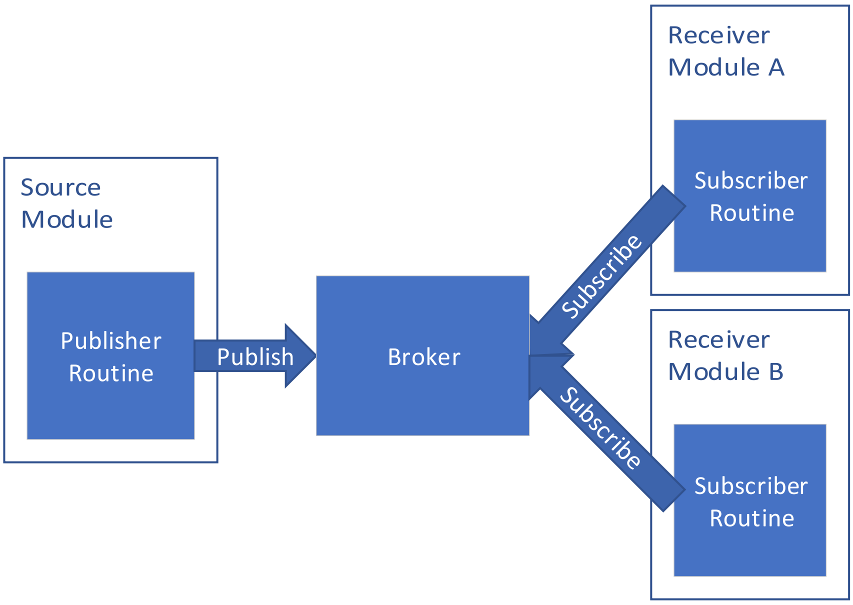

Figure 7 – Visual Representation of Messaging

Advantages:

- Support for real-time data;
- Broker solutions already available on the market;
  - Lower implementation complexity;
  - Lower maintenance learning curve.
- Faster communication and simplicity when the Sender and the Receiver are on different devices, networks or both.

Disadvantages:

- Higher resource usage due to:
  - Having a broker running (more noticeable when the Sender and the Receiver are on the same device);
  - Conversion of data from/to message format.
- At least two components must be running simultaneously.

4.3 Integration with External Systems

The system has three instances where data (raw, processed and markers) is produced and output so that can be read by another or, in the case of markers data, by the same system. Due to this, integration with External Systems is possible by allowing them to either access or provide the data in these particular instances. To access the data, the External System must either subscribe and receive from the broker via message, in the scenarios where real-time data is required, or read (or copy) the files directly from where they were output, for the remaining scenarios. In order to provide data to the system, the External System must provide it in a file that respects the structure defined in 5.2.

5 System Implementation

This chapter presents the technologies used to implement the architecture presented in the System Architecture as well as the justification of why it was selected over the alternatives. The Data Logger is presented in 5.1, which describes the Sensors and the Recorders for every acquisition method as well as the Data Annotation. The Data Viewer is described in 5.3 and the methods for the share of information between modules and external systems is described in 5.4. Finally, the description of the data format utilised in the system is described in 5.2.

5.1 Acquisition Methods

Acquisition methods are the methods utilised to acquire raw data from the subject. They have two major components: the **sensor** and the **recorder**. Since *IEETA Collecta* is a multi-modal system, multiple acquisition methods exist in the system and are categorised in the following way: key loggers, mouse loggers, screen loggers, webcam video recorders and ECG loggers. The following subsubsections describe how the sensors and the recorders were implemented for each acquisition method.

Key Logger

Figure 8 - Keyboard used in the System

This acquisition method consists on collecting data from the interaction with the keyboard. The **Sensor** used was the *DELL* keyboard portrayed in Figure 8. Since the only requirement was to have a keyboard that generates output in the ASCII [30] format with the Portuguese layout (most common keyboard in the location where the system was used), the choice of the keyboard was based on the availability in the University. The **Recorder** was implemented using Java with the library "system-hook" by the *GitHub* user *kristian* publicly available under the MIT License[7]. The reason for this choice was that the library is light-weight and thus has minimal or no impact on the performance of the system while the logging is taking place. An essential feature of the Recorder was the support for different keyboard layouts.

[7] https://opensource.org/licenses/MIT

Mouse Logger

Figure 9 - Mouse used in the System

This acquisition methods consists on collecting data from the interaction with the mouse. The **Sensor** used was the *Amazon Basics* mouse portrayed in Figure 9. The reason for this choice was the availability in the University due to not having any special requirement.

Similarly to the Key Logger, the **Recorder** was also implemented using Java with the library "system-hook" due to having the same requirement (having minimal or no impact on the system).

Screen Logger

This acquisition method consists of recording the computer screen. The **Sensor** used was the integrated graphics card available on the computer. The requirement was to be able to provide an output image with at least 720p resolution.

The **Recorder** was implemented using Java and is divided into two parts: the acquisition of images (frames) and the conversion of the frames to a video format. The former was implemented using native Java libraries, while the latter was implemented using the library "humble-video" by the *GitHub* user *artclarke* publicly available under the GNU[8] license.

Webcam Video Recorder

Figure 10 - Webcam used in the System

This acquisition method consists of recording the face and upper body of the subject. The **Sensor** used was the *Logitech HD* webcam portrayed in Figure 10. This model was chosen due to having 720p and fps (frames per second) rate over 15, which is the minimum recommended to run affective recognition by the major facial expressions methods.

The **Recorder** was implemented using Java and is divided into two parts: the acquisition of images and the conversion of the frames to a video format. The former was implemented using the library "webcam-capture" by the *GitHub* user *sarxos*, while the latter was implemented using the library "humble-video" by the *GitHub* user *artclarke* publicly available under the GNU[8] license.

ECG Logger

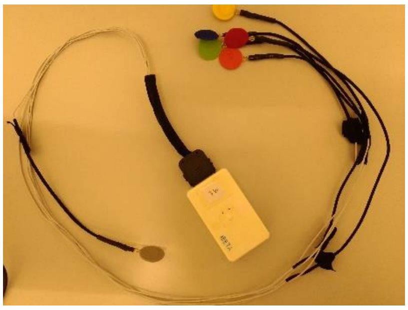

Figure 11 - ECG Sensor used in the System

This acquisition method consists on recording the subject's ECG data. The **Sensor** used was the VitalJacket from the company *Biodevices* portrayed in Figure 11. Since no special requirements exist, the reason for this choice was the availability in the University.

The **Recorder** was implemented in c++ using the SDK provided by the Sensor's manufacturer.

Data Annotation

The data annotation module was implemented in such a way that supports all the acquisition methods. It was implemented in its entirety in Java and is run by providing arguments denoting the operation to be performed and the location of the data to be processed alongside the destination for the output data. Only Standard JDK libraries were used, with the exception of video manipulation, in which the library *ffmpeg*[9] was utilised. The operations supported in this module are represented in the following list:

---

8 https://www.gnu.org/licenses/agpl-3.0.en.html

9 https://www.ffmpeg.org/

- Cut the data files into one or multiple files each representing specific intervals of time. (This operation was useful in dividing the data collected from the study described in chapter 6 into intervals of time representing each of the tasks performed by the participant);
- Calculation of complex features from the raw data and creation of processed data that consisted of the already existent raw data with the newly calculated features injected into it. The features generated for each data type are the ones described in chapter 5.2;
- Formating the processed data into a format that facilitates the importation into matlab. (This operation was also useful to the affective classification of the data obtained from the study described in chapter 6).

5.2 Data Description

The affective data generated by the system can be divided into two different categories: the raw data, consisting of the data generated by Data Logger subsystem, and the pre-processed data, consisting of the data generated by the Processing subsystem. Each can be further divided into multiple categories that represent one acquisition method each.

Raw Data

Five different types of raw data were gathered: data from **keyboard** utilisation (recorded key press and key release events), data from **mouse** interactions (recorded cursor movement or mouse key presses), video recording of the **computer screen**, video **recording of the face and upper body**, and **ECG** data. Each data type was recorded and stored in a different file. For every file, a specific naming criterion was defined: *"<data type ID><initial timestamp>.<extension>"*. *Data type ID* consisted of the ID of the data type, *initial timestamp* consisted of the Unix timestamp (milliseconds since 00:00:00 1st January 1970) of the moment when the recording started, and *extension* consisted of the file extension. The data variables collected from the sensors, their description and the file types can be consulted in Table 9.

Variable	Description	Acquisition Method (Data Type ID) / File Type (Extension)
time	Unix Timestamp.	Keyboard Utilisation (keyboard) / Tab Separated Values (.tsv)
eventType	The value is keyDown if the key is being pressed or keyUp if it is being released.	
keycode	The ASIIC code of the key pressed.	
controlPressed	The value is 1 if the control key is also being pressed alongside the current one, 0 if otherwise.	
menuPressed	The value is 1 if the menu key is also being pressed alongside the current one, 0 if otherwise.	
shiftPressed	The value is 1 if the shift key is also being pressed alongside the current one, 0 if otherwise.	
time	Unix Timestamp.	Mouse Utilisation (mouse) / Tab Separated Values (.tsv)
x	X coordinate of the cursor.	
y	Y coordinate of the cursor.	
left	The value is 1 if the left mouse button is pressed, 0 if otherwise.	
right	The value is 1 if the right mouse button is pressed, 0 if otherwise.	
middle	The value is 1 if the middle mouse button is pressed, 0 if otherwise.	
delta	If the mouse wheel is being rolled. If the value is negative, the roll direction is backwards.	
time	Unix Timestamp.	ECG Data (ecg) / Tab Separated Values (.tsv)
<unnamed>	50 columns of values consisting of the ecg sensor output value. (The sensor utilised sent a group of 50 values per message).	
<n. a.>	Video output.	Computer Screen Recording (print) / MP4 (.mp4)
<n. a.>	Video output.	Face and Upper Body Recording (webcam) / MP4 (.mp4)

Table 9 – Raw Data Description

Pre-processed Data

The pre-processing consisted of two different stages: calculation of new variables based on the raw data variables and formatting the already existent data. For every file, a specific naming criterion was defined: *“converted<data type ID><initial timestamp>.<extension>”*, with the *extension* always being *tsv*. The ECG data was formatted and both the keyboard and the mouse utilisation data had new variables added. The raw video data collected from both the Computer Screen Recording and the Face and Upper Body Recording was not pre-processed due to already being in the desired format for the emotion classification algorithms.

The ECG formatting process consisted of the unpacking of the 50 values included in one row, and giving a separate entry (row) for each value alongside a newly calculated timestamp and an ID that incremented iteratively.

The mouse and the keyboard utilisation data had new variables added. These added variables, alongside their description, are represented in Table 10 for the keyboard and in Table 11 for the mouse.

Variable	Description
keyDelay	The delay between the current keyUp and the last keyUp from another key (i.e. delay between two different keys). Only available in keyUp's.
releaseDelay	The delay between the current keyUp and the last keyDown after a keyUp of the same key (i.e. delay between pressing and releasing the same key). Only available in keyUp's.
numStrokes	Sum of keyUp's since the beginning of the recording to the current event.
numError	Sum of error key presses since the beginning of the recording to the current event.

Table 10 - Variables Added to Keyboard Data during Pre-processing

Variable	Description
hour	Hour part of the timestamp.
min	Minute part of the timestamp.
sec	Second part of the timestamp.
distance	Distance, in pixels, travelled by the cursor from the previous event.
accDistance	Accumulated distance, in pixels, travelled by the cursor since the beginning of the recording.
velocity	Velocity calculated with the distance and the delay from the previous event (velocity = distance/time).

Table 11 - Variables Added to Mouse Data during Pre-processing

In order to facilitate the import of the pre-processed data to *Matlab* so it could be used by the algorithms of emotion recognition, a filtering process took place and produced a new set of data. However, this data is less complete than the previously discussed and might have excluded important variables for other emotion classification methods. This process was applied on the keyboard utilisation data. The files generated had a specific naming criterion: *"matlabConverted<data type ID><initial timestamp>.<extension>"*, where the *data type ID* was *keyboard*. The file format changed from TSV to CSV, therefore the *extension* changed to *.csv* and the Header was also removed. In addition, the following variables were excluded: *time*, *eventType* and *keyCode*.

Markers Repository

The markers were stored in a TSV file named "*markers.tsv*" on the same folder of the data that they are associated with. The file consisted of two variables: the *timestamp*, which represented the timestamp and the *description*, which represented the text description of the marker.

5.3 Data Viewer

The Data Viewer was implemented using Java. The File Reader was implemented using strictly native Java libraries. For the UI, the library *vlcj*[10] was utilised to display video files while the library *JFreeChart*[11] was utilised to display data graphs. Since the main goal of this software was to allow the data analyst to view the data and mark events, the most relevant part is the display of visual data (i.e. videos). The Data Viewer UI contains two video panels that play the video files of the face and upper body and the recording of the screen in a synchronised fashion. There is a video bar on the bottom that represents the duration of the interval where the data can be synchronised. This bar can be interacted either by dragging or clicking in order to display data on the desired point in time. The user also has the possibility to play or pause the videos by pressing the Play/Pause button. Bellow the videos, the UI displays a graph with data from a file. As proof of concept, we decided to display ECG data in this graph. The graph also has the possibility to focus the view on a user-specified timestamp (by utilising the text field and the "Set Timestamp" button) or to automatically accompany the videos as they are being played (by toggling the button "Auto Mode"). Finally, on the left panel, the user can see the existent markers along with their timestamp and have the ability to add new markers or remove existent ones. Figure 12 is a screen capture from the UI displaying sample data from a gameplay session with an ECG recording.

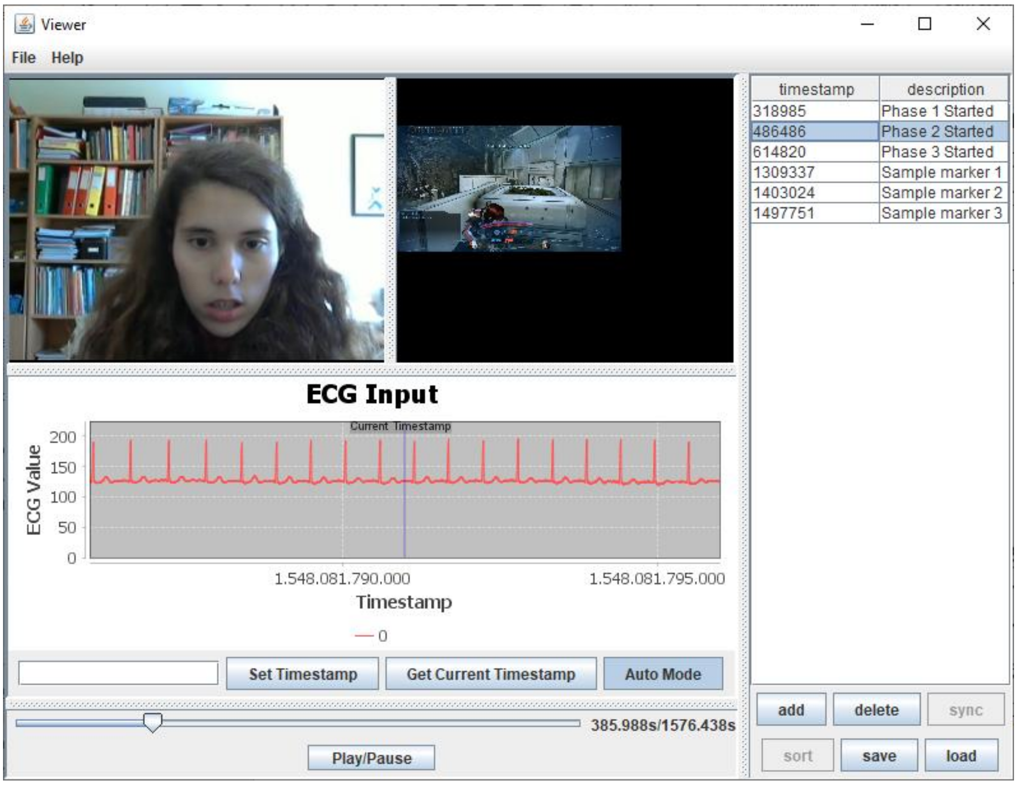

Figure 12 - Data Viewer UI

5.4 Communication

Communication methods are used to transfer data from one module to another and act as a bridge between them. They are divided into two different categories **File Sharing** and **Messaging**,

[10] http://capricasoftware.co.uk/projects/vlcj
[11] http://www.jfree.org/jfreechart/

each differing in its implementation. It is also noteworthy that external systems can differ in implementation as long as the interfaces are respected.

The **File Sharing** method was implemented with Java using only its native libraries. We preferred this solution over the use of an already implemented library because having a specialised implementation has less impact on the performance of the system than a broad one. Another decision that enhanced the performance was writing the file as the data was being generated thus enabling more memory to be freed while also facilitating error recovery.

The **Messaging** method was implemented in Java and utilised *RabbitMQ*[12]. This method was implemented alongside File Sharing to enable the possibility of real-time data, with the only exception being the writing of the markers' database (only File Sharing was supported). For the method to function correctly, a broker was also implemented. This choice of technology was based on a paper [31] where *RabbitMQ* was compared to *Apache Kafka*, another popular system that supports publish-subscribe messaging, and concluded that despite *Apache Kafka* being able to handle pub-sub messaging, *RabbitMQ* was more adequate.

[12] https://www.rabbitmq.com/

6 Results and validation

In order to validate the system and to probe for possible errors, a test scenario with frequent computer users was defined which included a study and a smaller-scale pilot study. The studies were conducted in the *Universidade de Aveiro* and utilised the same software and hardware already existent in the developed system. In addition to the hardware already present in the system, extra hardware (headphones) to output sound was also included in the studies. Software to display the tasks instructions, output the required media and receive text input from the participant was also developed.

6.1 Study with participants with frequent computer usage

Study Protocol

A protocol was designed in collaboration with the *Department of Education and Psychology from the University of Aveiro* and approved by the *Ethic's Council of the University of Aveiro*. Taking into account the protocol's requirements, the study's participants selected were students from the university that had frequent computer usage.

The **equipment** utilised was:

- One *VitalJacket*;
- One headphone;
- One computer;
- One webcam;
- One keyboard;
- One mouse;

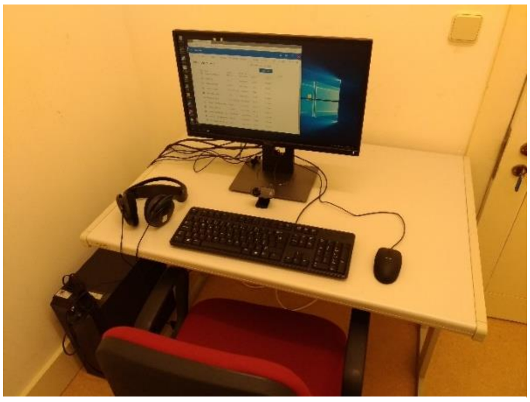

Figure 13 - Study Equipment Set-Up

All the equipment described in the previous list was already part of the system being tested. The only exception was the headphones, which were not previously incorporated in the system. The headphones utilised were *Sony Headphones*. Figure 13 represents the set-up of the equipment utilised in the study.

There were two separate moments where the participants were asked to fill out characterisation questionnaires: before and after performing the tasks. These questionnaires were implemented as a Google Form, as seen in Figure 14, and the software running the experiment would provide the link to them. All the questionnaires utilised can be found in the Appendix.

The characterisation questionnaires filled before performing the tasks consisted of:

- Dyslexia diagnosis;
- Hearing diagnosis;
- Vision diagnosis;
- State-Trait Inventory, utilising the *State-Trait Inventory for Cognitive and Somatic Anxiety* (STICSA) [32] questionnaire. Two different variants of the questionnaire were used in order to increase accuracy;
- Life satisfaction score, utilising the *Satisfaction With Life Score (SWLS)* [33] questionnaire;
- Positive and negative affect score, utilising the *Positive and Negative Affect Schedule (PANAS)* [34] questionnaire;
- Alexithymia (trouble identifying and describing emotions) score, utilising the *Toronto Alexithymia Scale (TAS-20)* [35] questionnaire.

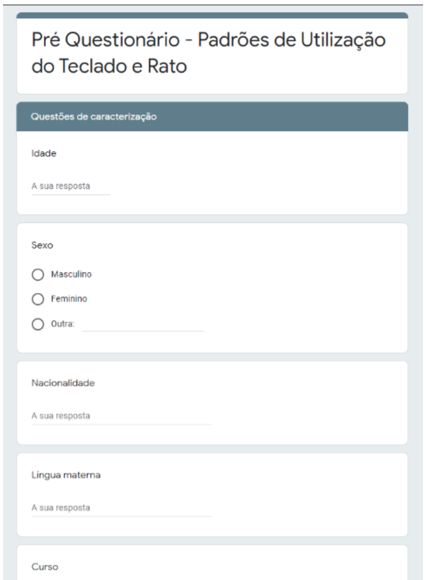

Figure 14 - Excerpt from a Google Forms' Characterisation Questionnarie

The characterisation questionnaires filled before performing the tasks consisted of:

- State-Trait Inventory, utilising the *State-Trait Inventory for Cognitive and Somatic Anxiety* (STICSA) [32] questionnaire;
- Positive and negative affect score, utilising the *Positive and Negative Affect Schedule (PANAS)* [34] questionnaire;

Upon arrival, the study was explained to the participants, and any doubts that might have arisen were clarified. The *informed consent* paper form was provided to the participant to read and sign in order to proceed. The participant was equipped with the needed sensors; the remaining required material to the experiment was set-up. The participant was sat on a chair, and an ID provided to him. The researcher would toggle the software to start capturing data. And finally, the researcher left the room, and the experiment would begin.

When the experiment began, the first stage was the filling of characterisation questionnaires. Afterwards, the participant performed five different tasks. The software running the experiment would display the instructions for each task before and during the execution to assist the participant in not forgetting them. In order to reset the emotional state from one task to another, the intervals between tasks had a fixed duration and displayed a neutral slide of images (images not linked with any particular emotion such as photographs of animals, plants, landscapes and others) while also playing a neutral sound (coffee shop background noise). The sequence of tasks is described in the following list:

1. **First Task**: The participant wrote two paragraphs about a **travel destination** he/she had visited or would like to visit that impacted him/her. There was no time restriction;
2. **Second Task**: The participant listened to an audio track of speech while transcribing it in the text area. The speed of the audio track was 0.7x. There was no time limit to finish the task;
3. **Interval**: Slideshow and background track played for 30 seconds;
4. **Third Task**: The participant listened to an audio track of speech while transcribing it in the text area. The speed of the audio track was 0.8x. After the track ended, there was a 3 second time limit to finish the task;
5. **Interval**: Slideshow and background track played for 30 seconds;

6. **Fourth Task**: The participant listened to an audio track of speech while transcribing it in the text area. The speed of the audio track was 1x. After the track ended, there was a 3 second time limit to finish the task;
7. **Interval**: Slideshow and background track played for 30 seconds;
8. **Fifth Task**: The participant wrote two paragraphs about a **film** that had impacted them. There was no time restriction;

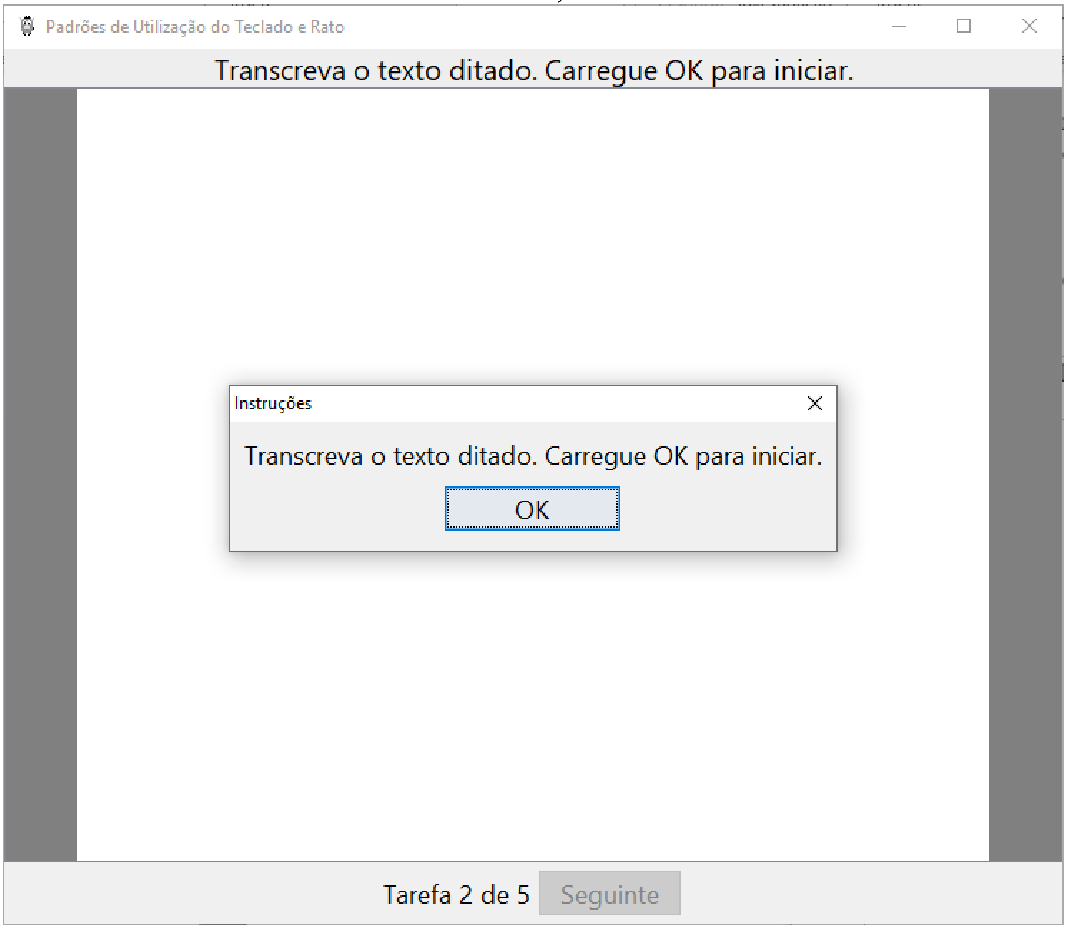

Figure 15 – Experiment Software Displaying Task Instructions

Finally, the participant would fill out the last set of characterisation questionnaires and finish the experiment. The researcher would then proceed to assist in the removal of the sensors, stop the data capturing and reward the participant with an edible gift.

The researcher would state at the beginning of the experiment that a competition was taking place and that the participant who had achieved the highest correct word count in the transcription tasks (tasks two, three and four) would be granted a reward. This was an attempt to influence the participant to take those tasks more seriously in order to both induce stress and reduce the probability of him/her giving up mid-task. When the experiment was over, the participant would be informed that no competition was taking place and that every participant would earn a reward.

Every participant would perform the tasks independently and asked not to talk about the tasks with the remaining participants that had yet to perform the study.

Pilot Study

The system utilised in the study was the result of improvements and bug fixes based on the feedback obtained from a pilot study performed beforehand with an earlier version. The system was improved incrementally between participants, i.e. the system would be improved from the last participant's feedback to the next. In total, four participants took part in the pilot study.

A list containing the feedback from the participants and the improvements they influenced on the system can be found in Table 12.

Feedback	Improvement
The participant had a mental block and skipped the first task without writing anything.	The topic was replaced from writing about a song to writing about a travel location.
The participant would not remember the instructions during the task or would skip them without reading.	A label containing the instructions of the active task was added to the interface.
The participant would not realise that the sound would be output by the headphones and did not equip them.	A message stating the need for headphones was added to the instructions. The researcher started to reinforce the requirement for headphones when describing the experiment.
The participant would mistakenly start the first task instead of filling the first questionnaire set.	Software changed so the browser automatically opens the questionnaire after the participant has written the ID.
The output from video recording did not have the required fps due to the computer system not being able to handle the recording.	The computer system was replaced by a more recent one.
The researcher would forget to perform the correct configuration on the recording software.	Correct configuration values were added to the defaults of the recording software.

Table 12 - Feedback from the Pilot Study

6.2 Results

Seven students participated in this study, with all the collected data being valid for processing and analysis. Each participant was given an ID in the format *Pxxxxxxx*, with the *x*'s representing numbers.

The data was collected from the mouse and keyboard utilisation patterns, the ECG, the face and upper body video recording and the computer screen video recording. Both the raw and the pre-processed data formats utilised in the study remained the same as the system (as described in Data Description). In the classification of emotions phase, two algorithms were used: the *Random Forest* algorithm [26]; and the *K-Means* algorithm [36], [37]. The format of the data produced by the *clustering algorithm* (*K*-Means) consisted of the percentage of participants classified in the class number 2. The data of the *random forest* algorithm was post-processed by comparing the classification obtained with the data manually annotated. The resulting data of this comparison was in the format of a confusion matrix.

For the majority of participants, the mouse had almost no data gathered or, in some cases, none at all. We believe that this was due to the tasks being mostly writing based. Consequently, the data gathered from the mouse was discarded in the analysis due to not being in sufficient quantity.

As inputs for the **clustering algorithm**, four different scenarios were considered and the keyboard data was utilised in all of them. The scenarios were the result of the combination of two different variables: if the data was normalised to 0 mean; and if ECG information (HR and RR) was included. The scenarios are the following:

1. Keyboard data **with** ECG information and **with** normalisation to 0 mean;
2. Keyboard data **without** ECG information and **with** normalisation to 0 mean;
3. Keyboard data **with** ECG information and **without** normalisation to 0 mean;
4. Keyboard data **without** ECG information and **without** normalisation to 0 mean.

Two classes were defined, one representing stress while the other not. Table 13 represents the percentage of samples classified as class number 2 generated from running the clustering algorithm.

	ECG/Mean	no ECG/Mean	ECG/no Mean	no ECG/no Mean
P0010207	2.34%	2.34%	97.66%	34.73%
P0010306	0.06%	0.45%	0.45%	0.06%
P0010307	0.06%	0.06%	0.64%	0.06%
P0010506	1.80%	1.80%	1.74%	43.96%
P0012506	0.71%	0.71%	0.71%	0.96%
P0020306	0.84%	1.22%	0.84%	0.13%

Table 13 - Clustering Results with K-means algorithm (Percentage in class 2)

For all scenarios, the percentage of samples classified as class number 2 was rarely above 3%. The results of the participant P0010207 differ significantly from the remaining participants. However, this could be due to a class inversion, which we considered to be not significant.

Another variant of the algorithm was also used, which considered three classes in order to have an increased similarity with the SWELL dataset analysis. However, the results were also associated with random responses. Since the results obtained from both variants differed from what we expected, we complemented the classification of the data with another algorithm, the random forest.

The **random forest algorithm** was used with three different classes, also in order to have an increased similarity with the SWELL dataset analysis. Class C1 represented the tasks with free writing (tasks one and five); class C2 represented the first transcription task (task two); class C3 represented the last two transcription tasks (tasks three and four). We believed that the class C1 denoted an unstressful situation, the class C2 a slightly stressful situation, and the class C3 a stressful situation. Table 14 represents the resulting confusion matrix while Table 15 represents the error rate of a sample of a certain class being classified in the correct class.

	C1	C2	C3
C1	14546	268	7476
C2	3711	46	2573
C3	8903	185	5866

Table 14 - Random Forest Results (Confusion Matrix)

Class	Error Rate
C1	34.74%
C2	99.27%
C3	60.77%

Table 15 - Random Forest Results (Error Rate)

The algorithm classified the majority of the class C1 samples correctly, with an error rate of 34.74%. However, the remaining classes, C2 and C3, had an error rate above 50% (99.27% and 60.77%, respectively).

Since the results differed from the expectations, it was decided to inspect the filled questionnaires by the participants, in order to verify if the implemented protocol correctly induced emotional state changes. The results of the **pre-tasks questionnaires** are from the translation of the STICSA questionnaire, designed to measure anxiety [32]. In the calculation of the score, the responses *Not at all*, *A little*, *Moderately* and *Very much so* were represented by the values 1, 2, 3 and 4, respectively. In the calculation of the percentages, it was assumed that the value 0 (*Not at all*) corresponded to 0% and the value 4 (*Very much so*) corresponded to 100%. Higher scores represent higher reported anxiety. Table 16 represents both the data and the score of each participant, and also the average (excluding participants with no or invalid data).

Participant	Not at all	A little	Moderately	Very much so	Score	Percentage
P0010306	6	9	5	1	2.048	34.92%
P0020306	8	11	2	0	1.714	23.81%
P0010506	14	7	0	0	1.333	11.11%
P0012506	16	5	0	0	1.238	7.94%
P0010207	0	0	0	0	<no data>	<no data>
P0010307	10	3	6	2	2.000	33.33%
P0072907	14	5	2	0	1.429	14.29%
				Average	1.627	20.90%

Table 16 - Pre-tasks Questionaire Results (STICSA)

Disregarding the participant with no reported data, the average score obtained was approximately 1.63 (20.90%), the minimum was approximately 1.24 (7.94%), and the maximum was approximately 2.05 (34.92%). These values can be translated to the participants reporting either a little or no stress at all.

Similarly to the pre-tasks questionaries, only the relevant data obtained from the **pos-tasks questionaries**, the STICSA, will be presented here. Besides, the same methodology of data representation will also be used. Table 17 represents both the data and score of each participant, as well as the average.

Participant	Not at all	A little	Moderately	Very much so	Score	Percentage
P0010306	9	6	0	2	1.706	23,53%
P0020306	8	12	0	0	1.600	20,00%
P0010506	16	5	0	0	1.238	7,94%
P0012506	16	5	0	0	1.238	7,94%
P0010207	0	0	0	10	4.000	100,00%
P0010307	12	2	0	1	1.333	11,11%
P0072907	12	8	0	0	1.400	13,33%
				Average	1.788	26,26%

Table 17 - Pos-tasks Questionaire Results (STICSA)

The total score final average is approximately 0.16 (5.36%) higher in the post-tasks questionnaires than in the pre-tasks questionnaires. However, one participant did not provide answers in the pre-tasks questionnaire and obtained a score of 4 (100%) which has a substantial disparity from the other participants. Consequently, this participant should not be utilised in the comparison.

Table 18 represents the **difference** between the results of the post and pre-tasks questionnaires for each participant. The average of the participants is also represented, excluding scenarios where no reported data was available.

Participant	Score Difference	Percentage Difference
P0010306	-0.342	-11.39%
P0020306	-0.114	-3.81%
P0010506	-0.095	-3.17%
P0012506	0.000	0.00%
P0010207	*<no data>*	*<no data>*
P0010307	-0.666	-22.22%
P0072907	-0.029	-0.96%
Average	-0.208	-6.93%

Table 18 - Differential between Pos and Pre-tasks Questionnaires Results (STICSA)

No participant reported higher stress after the tasks. The average is approximately -0.21 (-6.93%) and the values range from approximately -0.34 (-11.39%) to 0 (0%). Which may justify the inability of the classifier and clustering method to discriminate tasks. If the participants did not feel any anxiety or stress, it will not be reported on the physiological or behavioural metrics.

Analysis

When using the SWELL dataset in the exploratory study, a global accuracy of 89.97% was achieved using the same processing methods. Despite this fact, the same methods did not present any statistically significant results when analysing the data obtained from the study performed with participants with frequent computer usage. Due to this, we believe that the protocol designed did not correctly induce stress on the participants. To further support this hypothesis, the questionnaire data displays that the participants reported that the level of the perceived stress remained identical before and after the experiment, with a slight decrease after finishing the tasks.

We manually classified tasks one and five as non-stressful and tasks two, three and four as stressful. Since the task length was similar, we expected a classification of 40% (2/5) of the total data as non-stressful and 60% (3/5) as stressful. The Clustering algorithm produced results where the classification on class number 2 was rarely above 3% which differs greatly from the ideal expected results of classification of 60% in class number 1 and 40% in class number 2. These observations translate to the algorithm not classifying the data correctly and thus leading us to the decision of not analysing the results any further.

In the results of the Random Forest algorithm, it can be verified that identifying the C1 class was considerably more precise than the others. However, more than half of the classes C2 and C3 were classified as C1, 58.63% and 59.54% respectively, indicating a heavy bias towards classification as C1. These facts led us to believe that the high performance in classifying the C1 class is due to the free tasks containing broad affective data rather than the correct emotional state being induced on the participants.

In the questionnaires, the participant P0010207 did not fill the pre-tasks questionnaire, thus producing no data. Moreover, the data produced on the post-tasks questionnaires was significantly isolated from the other participants. Due to this, this participant's incomplete data has an impact on the result of the difference in the total average of 0.16 (5.36%). We believe that the average of the differences between the pos and pre-tasks questionnaires of each participant (-0.21 [-6.93%]) represents more accurately the real difference between stress reported.

7 Conclusions and future work

The main purpose of the dissertation was to develop a multi-modal system that gathered data with multiple acquisition methods while allowing integration with other systems. We believe that the system developed did fulfil its original objectives.

After the development, the system was tested in a study with participants that attempted to simulate the stress experienced by programmers while developing software. The analysis of the data gathered in the study had an accuracy below 50%, which is not sufficient to state that the system correctly classifies stress. However, we believe that this is due to the study protocol not inducing stress correctly because the participants reported in the pre and post-tasks questionnaires no significant stress level disparity, and also the *Random Forest* algorithm was previously utilised in the classification of another dataset (*SWELL*) and had an overall accuracy of 89.97%.

This work could be further expanded by conducting another study with a protocol that correctly induces stress, in order to validate the system. The implementation of additional sensors and acquisition methods, particularly, intrusive ones (since this dissertation focused primarily on non-intrusive) would also expand on the work performed. Another improvement would also be the support for other use cases apart from software development. We believe that Video Advertising, Cinematography and Online Shopping would be good use cases to explore since having affective data is beneficial due to Cinematography relying on emotions generated on the viewers to classifying if a film was good or not, and Video Advertising and Online Shopping relying on producing emotions on the potential buyers that generate buying behaviours [38], [39].

In this dissertation, I improved my programming skills in order to design and develop a multimodal system from scratch. I also learnt how the emerging field of Affective Computing has a promising future and, as its methods become more accurate and refined, can add tremendous value to the emotion classification. Additionally, non-intrusive methods allow for a continuous data gathering (due to not interfering with the subject tasks) which has the potential to prevent certain problems by detecting and warning about repeated harmful emotional states before they escalate any further and allow for specialist intervention earlier.

8 References

[1] P. S. R. Gomes, “Easy Psycho Study: web platform to streamline psychology questionnaires,” M.S. thesis, DETI, UA, Aveiro, PT, 2017.

[2] T. Bastos, “Vitals recorder: Sistema móvel para apoiar a realização de estudos de psicofisiologia,” M.S. thesis, DETI, UA, Aveiro, PT, 2018.

[3] J. Panksepp, *Affective Neuroscience: The Foundations of Human and Animal Emotions (Series in Affective Science)*. 2004.

[4] R. W. Picard, “Affective Computing,” 1995.

[5] S. K. D’Mello and J. Kory, “A Review and Meta-Analysis of Multimodal Affect Detection Systems,” vol. 47, no. 3, 2015.

[6] A. Ko, “A review of emotion recognition methods based on keystroke dynamics and mouse movements,” pp. 548–555, 2013.

[7] A. Alberdi, A. Aztiria, A. Basarab, and D. J. Cook, “Using smart offices to predict occupational stress,” *Int. J. Ind. Ergon.*, vol. 67, no. March, pp. 13–26, 2018.

[8] J. Hernandez, P. Paredes, A. Roseway, and M. Czerwinski, “Under Pressure: Measuring the Stress of Computer Users,” *Hum. Factors Comput. Syst.*, pp. 51–60, 2014.

[9] H. Lee, Y. S. Choi, S. Lee, and I. P. Park, “Towards Unobtrusive Emotion Recognition for Affective Social Communication,” *9th Annu. IEEE Consum. Commun. Netw. Conf. - Spec. Sess. Affect. Comput. Futur. Consum. Electron.*, 2012.

[10] C. Epp, M. Lippold, and R. L. Mandryk, “Identifying emotional states using keystroke dynamics,” p. 715, 2011.

[11] M. Wrobel, “Applicability of Emotion Recognition and Induction Methods to Study the Behavior of Programmers,” *Appl. Sci.*, vol. 8, no. 3, p. 323, 2018.

[12] S. Koldijk, M. A. Neerincx, and W. Kraaij, “Detecting Work Stress in Offices by Combining Unobtrusive Sensors,” *IEEE Trans. Affect. Comput.*, 2018.

[13] A. Elder, *The New Trading for a Living: Psychology, Trading Tactics, Risk Management, and Record-Keeping*. 2014.

[14] J. Zulueta *et al.*, “Predicting mood disturbance severity with mobile phone keystroke metadata: A biaffect digital phenotyping study,” *J. Med. Internet Res.*, vol. 20, no. 7, pp. 1–10, 2018.

[15] O. Matthews, M. Vigo, and S. Harper, “Sensing Arousal and Focal Attention During Visual Interaction,” no. 1, pp. 263–267, 2018.

[16] M. Exposito, J. Hernandez, and R. W. Picard, “Affective Keys: Towards Unobtrusive Stress Sensing of Smartphone Users,” *Proc. 20th Int. Conf. Human-Computer Interact. with Mob. Devices Serv. Adjun.*, pp. 139–145, 2018.

[17] G. Caridakis, L. Malatesta, L. Kessous, N. Amir, A. Raouzaiou, and K. Karpouzis, “Modeling naturalistic affective states via facial and vocal expressions recognition,” *ICMI’06 8th Int. Conf. Multimodal Interfaces, Conf. Proceeding*, pp. 146–154, 2006.

[18] S. Salmeron-Majadas, R. S. Baker, O. C. Santos, and J. G. Boticario, “A Machine Learning Approach to Leverage Individual Keyboard and Mouse Interaction Behavior from Multiple Users in Real-World Learning Scenarios,” *IEEE Access*, vol. 6, pp. 39154–39179, 2018.

[19] M. Kowalska and M. Wróbel, “Basic Emotions,” *Encyclopedia of Personality and Individual Differences*. pp. 1–6, 2017.
[20] J. A. Russell, “Core affect and the psychological construction of emotion.,” *Psychol. Rev.*, vol. 110, no. 1, pp. 145–172, 2003.
[21] B. C. Ko, “A brief review of facial emotion recognition based on visual information,” *Sensors (Switzerland)*, vol. 18, no. 2, 2018.
[22] C. Busso *et al.*, “Analysis of emotion recognition using facial expressions, speech and multimodal information,” *ICMI’04 - Sixth Int. Conf. Multimodal Interfaces*, pp. 205–211, 2004.
[23] S.-H. Lau, “Stress Detection for Keystroke Dynamics,” Carnegie Mellon University, 2018.
[24] S. Koldijk, M. Sappelli, S. Verberne, M. A. Neerincx, and W. Kraaij, “The SWELL Knowledge Work Dataset for Stress and User Modeling Research,” in *Proceedings of the 16th International Conference on Multimodal Interaction - ICMI ’14*, 2014, pp. 291–298.
[25] M. Sappelli, S. Verberne, S. Koldijk, and W. Kraaij, “Collecting a dataset of information behaviour in context,” in *ACM International Conference Proceeding Series*, 2014.
[26] Tin Kam Ho, “Random decision forests,” in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1995, vol. 1, pp. 278–282.
[27] A. Ellahi, M. Khalil, and F. Akram, “Computer users at risk: Health disorders associated with prolonged computer use,” *J. Bus. Manag. Econ.*, vol. 2, no. 4, pp. 171–182, 2011.
[28] A. B. Bakker, W. B. Schaufeli, M. P. Leiter, and T. W. Taris, “Work engagement: An emerging concept in occupational health psychology,” *Work Stress*, vol. 22, no. 3, pp. 187–200, Jul. 2008.
[29] P. Paredes and M. Chan, “CalmMeNow,” in *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA ’11*, 2011, p. 1699.
[30] A. A. Bruen and M. A. Forcinito, “ASCII,” in *Cryptography, Information Theory, and Error-Correction*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2011, pp. 445–446.
[31] P. Dobbelaere and K. S. Esmaili, “Kafka versus RabbitMQ,” in *Proceedings of the 11th ACM International Conference on Distributed and Event-based Systems - DEBS ’17*, 2017, pp. 227–238.
[32] D. F. Grös, M. M. Antony, L. J. Simms, and R. E. McCabe, “Psychometric Properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA): Comparison to the State-Trait Anxiety Inventory (STAI),” *Psychol. Assess.*, vol. 19, no. 4, pp. 369–381, 2007.
[33] E. Diener, R. A. Emmons, R. J. Larsen, and S. Griffin, “The Satisfaction With Life Scale.,” *J. Pers. Assess.*, vol. 49, no. 1, pp. 71–5, Feb. 1985.
[34] D. Watson, L. a. Clark, and A. Tellegan, “PANAS,” *J. Pers. Soc. Psychol.*, 1988.
[35] R. M. Bagby, J. D. A. Parker, and G. J. Taylor, “The twenty-item Toronto Alexithymia Scale--I. Item selection and cross-validation of the factor structure.,” *J. Psychosom. Res.*, vol. 38, no. 1, pp. 23–32, Jan. 1994.
[36] J. A. Hartigan and M. A. Wong, “Algorithm AS 136: A K-Means Clustering Algorithm,” *Appl. Stat.*, vol. 28, no. 1, p. 100, 1979.
[37] John A. Hartigan, *Clustering Algorithms*, 99th ed. John Wiley & Sons, Inc., 1975.
[38] R. P. Bagozzi, M. Gopinath, and P. U. Nyer, “The Role of Emotions in Marketing,” *J. Acad. Mark. Sci.*, vol. 27, no. 2, pp. 184–206, Apr. 1999.
[39] “A New Concept of Marketing: The Emotional Marketing,” *Brand Res. Accounting, Negoc. Distrib.*, 2010.

9 Appendix

9.1 *Informed Consent*

Consentimento Informado

A sua participação na presente investigação tem como objetivo recolher dados que permitam estudar os padrões de utilização do teclado e rato e sua associação com estados emocionais.

Este estudo insere-se no âmbito de um projeto de investigação que visa estudar a resposta fisiológica e comportamental das emoções, isto é usar informação do nosso corpo para interpretar emoções, possibilitando o desenvolvimento de sistemas que se adaptem à pessoa e suas necessidades. Um sub-tópico deste projeto pretende estudar a interação do utilizador com o teclado e rato. Neste contexto está a ser realizada uma dissertação do mestrado integrado em Engenharia dos Computadores e Telemática da Universidade de Aveiro, pelo aluno Daniel Oliveira, orientada pelo Professor Ilídio Oliveira e Doutora Susana Brás.

A sua participação é voluntária e implica a recolha de biosinais (ECG, EDA), imagens da face, e registo de utilização do teclado e rato durante a experiência. Em qualquer momento pode desistir da sua participação nesta experiência.

Todos os dados recolhidos destinam-se exclusivamente a fins de investigação e são confidenciais. Neste sentido, serão analisados de forma agregada, salvaguardando que as pessoas não possam ser identificadas.

Por favor, leia cuidadosamente cada uma das questões que se seguem e responda de acordo com o que melhor se aplica ao seu caso. **Note que as instruções, o tipo de questões e as escalas de resposta não são sempre iguais**

Desde já agradecemos a sua colaboração e disponibilizamos os nossos contactos para qualquer questão, dúvida ou interesse na temática.

Atenciosamente,

Ilídio Oliveira – DETI / IEETA – ico@ua.pt

Susana Brás – DETI / IEETA – susana.bras@ua.pt

Ao assinar este documento estará a autorizar a utilização dos seus dados para os fins de investigação indicados. Estará ainda a confirmar que leu e compreendeu a informação fornecida, tendo concordado com a mesma, garantindo que a sua participação é voluntária.

CONFIRMO QUE LI E CONCORDO COM A INFORMAÇÃO FORNECIDA, PELO QUE PRETENDO AVANÇAR COM A MINHA PARTICIPAÇÃO NO ESTUDO

Data: ______________

Assinatura: ____

9.2 *Pre-Task Questionaires*

I.Questões de caracterização

1. Idade_______anos
2. SEXO:
   Masculino ☐ Feminino ☐▪
3. Nacionalidade:___________________________________
4. Língua materna:_________________________________
5. Curso:____
   _____
6. Ano:________________
7. Ao nível da lateralidade como se avalia? Destro ☐ Esquerdino ou canhoto ☐ Ambidextro ☐▪
8. Possui, ou em algum momento possuiu, alguma forma de dislexia (disgrafia, disortografia ou discalculia - dificuldades que se podem manifestar por dificuldades ao nível da escrita)?
   Sim☐ Não ☐▪
   Especifique, pf:____
9. Tem algum problema de audição? Sim ☐ Não ☐▪
   Especifique, pf:____
10. Tem algum problema de visão? Sim ☐ Não ☐▪
    Especifique, pf:____
11. Encontra-se a tomar alguma medicação? Sim ☐ Não ☐▪
    Especifique, pf:____

STICSA-2

Abaixo encontra-se uma lista de frases que podem ser usadas para descrever como as pessoas se sentem. Ao lado de cada frase estão quatro números que indicam com que frequência cada frase é verdadeira para si (por exemplo, 1 – Nada, 4 – Muito).

Por favor leia cada frase atentamente e assinale o número que melhor indica **COM QUE FREQUÊNCIA, EM GERAL, A FRASE É VERDADEIRA PARA SI**.

	Nada	Um pouco	Moderadamente	Muito
1. O meu coração bate rápido	1	2	3	4
2. Os meus músculos estão tensos	1	2	3	4
3. Sinto-me agoniado com os meus problemas	1	2	3	4
4. Eu penso que os outros não me aprovarão	1	2	3	4
5. Eu sinto que me vou perdendo porque não consigo decidir-me atempadamente	1	2	3	4
6. Sinto-me tonto	1	2	3	4
7. Sinto fraqueza nos meus músculos	1	2	3	4
8. Sinto-me trémulo e instável	1	2	3	4
9. Eu perspetivo algumas desgraças futuras	1	2	3	4
10. Não consigo tirar alguns pensamentos da minha cabeça	1	2	3	4
11. Tenho dificuldade em relembrar coisas	1	2	3	4
12. Sinto a minha face quente	1	2	3	4
13. Eu penso que o pior vai acontecer	1	2	3	4
14. Sinto que os meus braços e pernas estão hirtos	1	2	3	4
15. Sinto a garganta seca	1	2	3	4
16. Eu esforço-me a evitar pensamentos desconfortáveis	1	2	3	4
17. Não me consigo concentrar sem a intrusão de pensamentos irrelevantes	1	2	3	4
18. A minha respiração é rápida e superficial	1	2	3	4
19. Preocupo-me por não conseguir controlar os meus pensamentos tão bem como eu gostaria	1	2	3	4
20. Tenho borboletas no estômago	1	2	3	4
21. Sinto as palmas das mãos húmidas	1	2	3	4

SWLS

(Diener, Emmons, Larsen, & Griffin, 1985; VP: Neto, Barros, & Barros, 1990)

Mais abaixo estão cinco afirmações, com as quais pode concordar ou discordar. Utilizando a escala de 1 a 7 abaixo indicada, refira o seu grau de acordo com cada afirmação selecionando o número apropriado:

1 - Totalmente em desacordo
2 - Desacordo
3 - Ligeiramente em desacordo
4 - Nem de acordo nem em desacordo
5 - Ligeiramente de acordo
6 - Acordo
7 - Totalmente de acordo

1. Em muitos aspectos a minha vida aproxima-se dos meus ideais.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
2. As condições da minha vida são excelentes.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
3. Estou satisfeito(a) com a minha vida.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
4. Até agora consegui obter aquilo que era importante na vida.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
5. Se pudesse viver a minha vida de novo, não mudaria quase nada.	**1**	**2**	**3**	**4**	**5**	**6**	**7**

PANAS

(Watson, Clark, & Tellegen, 1988; VP: Galinha & Pais-Ribeiro, 2005)

Esta escala consiste num conjunto de palavras que descrevem diferentes sentimentos e emoções. Leia cada palavra, e utilize a escala que apresentamos, assinalando o número que melhor indica em que medida sentiu cada uma das emoções **durante as últimas semanas**.

	Nada ou muito ligeiramente	**Um pouco**	**Moderadamente**	**Bastante**	**Extremamente**
1. Interessado	**1**	**2**	**3**	**4**	**5**
2. Perturbado	**1**	**2**	**3**	**4**	**5**
3. Excitado	**1**	**2**	**3**	**4**	**5**
4. Atormentado	**1**	**2**	**3**	**4**	**5**
5. Agradavelmente surpreendido	**1**	**2**	**3**	**4**	**5**
6. Culpado	**1**	**2**	**3**	**4**	**5**
7. Assustado	**1**	**2**	**3**	**4**	**5**
8. Caloroso	**1**	**2**	**3**	**4**	**5**
9. Repulsa	**1**	**2**	**3**	**4**	**5**
10. Entusiasmado	**1**	**2**	**3**	**4**	**5**
11. Orgulhoso	**1**	**2**	**3**	**4**	**5**
12. Irritado	**1**	**2**	**3**	**4**	**5**
13. Encantado	**1**	**2**	**3**	**4**	**5**
14. Remorsos	**1**	**2**	**3**	**4**	**5**
15. Inspirado	**1**	**2**	**3**	**4**	**5**
16. Nervoso	**1**	**2**	**3**	**4**	**5**
17. Determinado	**1**	**2**	**3**	**4**	**5**
18. Trémulo	**1**	**2**	**3**	**4**	**5**
19. Ativo	**1**	**2**	**3**	**4**	**5**
20. Amedrontado	**1**	**2**	**3**	**4**	**5**

TAS-20

(Bagby, Taylor, & Parker, 1994; VP: Prazeres, Parker, & Taylor, 2000)

Usando a escala fornecida como guia, indique o seu grau de concordância com cada uma das seguintes afirmações assinalando a opção correspondente. Dê só uma resposta por cada afirmação.

Use a seguinte chave:

1 – Discordo totalmente
2 – Discordo em parte
3 – Nem discordo nem concordo
4 – Concordo em parte
5 – Concordo totalmente

1. Fico muitas vezes confuso sobre qual a emoção que estou a sentir.	**1**	**2**	**3**	**4**	**5**
2. Tenho dificuldade em encontrar as palavras certas para descrever os meus sentimentos.	**1**	**2**	**3**	**4**	**5**
3. Tenho sensações físicas que nem os médicos compreendem.	**1**	**2**	**3**	**4**	**5**
4. Sou capaz de descrever facilmente os meus sentimentos.	**1**	**2**	**3**	**4**	**5**
5. Prefiro analisar os problemas a descrevê-los apenas.	**1**	**2**	**3**	**4**	**5**
6. Quando estou aborrecido, não sei se me sinto triste, assustado ou zangado.	**1**	**2**	**3**	**4**	**5**
7. Fico muitas vezes intrigado com sensações no meu corpo.	**1**	**2**	**3**	**4**	**5**
8. Prefiro simplesmente deixar as coisas acontecer a compreender por que aconteceram assim.	**1**	**2**	**3**	**4**	**5**
9. Tenho sentimentos que não consigo identificar bem.	**1**	**2**	**3**	**4**	**5**
10. É essencial estar em contacto com as emoções.	**1**	**2**	**3**	**4**	**5**

11. Acho difícil descrever o que sinto em relação às pessoas.	**1**	**2**	**3**	**4**	**5**
12. As pessoas dizem-me para falar mais dos meus sentimentos.	**1**	**2**	**3**	**4**	**5**
13. Não sei o que se passa dentro de mim.	**1**	**2**	**3**	**4**	**5**
14. Muitas vezes não sei porque estou zangado.	**1**	**2**	**3**	**4**	**5**
15. Prefiro conversar com as pessoas sobre as suas actividades diárias do que sobre os seus sentimentos.	**1**	**2**	**3**	**4**	**5**
16. Prefiro assistir a espectáculos ligeiros do que a dramas psicológicos.	**1**	**2**	**3**	**4**	**5**
17. É-me difícil revelar os sentimentos mais íntimos mesmo a amigos próximos.	**1**	**2**	**3**	**4**	**5**
18. Posso sentir-me próximo de uma pessoa mesmo em momentos de silêncio.	**1**	**2**	**3**	**4**	**5**
19. Considero o exame dos meus sentimentos útil na resolução de problemas pessoais.	**1**	**2**	**3**	**4**	**5**
20. Procurar significados ocultos nos filmes e peças de teatro distrai do prazer que proporcionam.	**1**	**2**	**3**	**4**	**5**

STICSA-1

(Ree, MacLeod, French, & Locke, 2000)

Abaixo encontra-se uma lista de frases que podem ser usadas para descrever como as pessoas se sentem. Ao lado de cada frase estão quatro números que indicam o grau com que cada frase pode descrever o seu humor ou o modo como se está a sentir neste momento (por exemplo, 1 – Nada, 4 – Muito).

Por favor leia cada frase atentamente e assinale o número que melhor indica **COMO SE SENTE NESTE MOMENTO,** neste preciso momento, mesmo que não seja a forma como se sente habitualmente.

	Nada	**Um pouco**	**Moderadamente**	**Muito**
1. O meu coração bate rápido	**1**	**2**	**3**	**4**
2. Os meus músculos estão tensos	**1**	**2**	**3**	**4**
3. Sinto-me agoniado com os meus problemas	**1**	**2**	**3**	**4**
4. Eu penso que os outros não me aprovarão	**1**	**2**	**3**	**4**
5. Eu sinto que me vou perdendo porque não consigo decidir-me atempadamente	**1**	**2**	**3**	**4**
6. Sinto-me tonto	**1**	**2**	**3**	**4**
7. Sinto fraqueza nos meus músculos	**1**	**2**	**3**	**4**
8. Sinto-me trémulo e instável	**1**	**2**	**3**	**4**
9. Eu perspetivo algumas desgraças futuras	**1**	**2**	**3**	**4**
10. Não consigo tirar alguns pensamentos da minha cabeça	**1**	**2**	**3**	**4**
11. Tenho dificuldade em relembrar coisas	**1**	**2**	**3**	**4**
12. Sinto a minha face quente	**1**	**2**	**3**	**4**
13. Eu penso que o pior vai acontecer	**1**	**2**	**3**	**4**
14. Sinto que os meus braços e pernas estão hirtos	**1**	**2**	**3**	**4**
15. Sinto a garganta seca	**1**	**2**	**3**	**4**
16. Eu esforço-me a evitar pensamentos desconfortáveis	**1**	**2**	**3**	**4**
17. Não me consigo concentrar sem a intrusão de pensamentos irrelevantes	**1**	**2**	**3**	**4**
18. A minha respiração é rápida e superficial	**1**	**2**	**3**	**4**
19. Preocupo-me por não conseguir controlar os meus pensamentos tão bem como eu gostaria	**1**	**2**	**3**	**4**
20. Tenho borboletas no estômago	**1**	**2**	**3**	**4**
21. Sinto as palmas das mãos húmidas	**1**	**2**	**3**	**4**

9.3 *Post-Task Questionaires*

De acordo com a escala proposta refira em que medida, **neste momento**, sente….

Medo	0	1	2	3	4	5	6	7	8	9	10
Alegria	0	1	2	3	4	5	6	7	8	9	10
Ansiedade	0	1	2	3	4	5	6	7	8	9	10
Stress	0	1	2	3	4	5	6	7	8	9	10

STICSA-1

(Ree, MacLeod, French, & Locke, 2000)

Abaixo encontra-se uma lista de frases que podem ser usadas para descrever como as pessoas se sentem. Ao lado de cada frase estão quatro números que indicam o grau com que cada frase pode descrever o seu humor ou o modo como se está a sentir neste momento (por exemplo, 1 – Nada, 4 – Muito).

Por favor leia cada frase atentamente e assinale o número que melhor indica **COMO SE SENTE NESTE MOMENTO,** neste preciso momento, mesmo que não seja a forma como se sente habitualmente.

	Nada	**Um pouco**	**Moderadamente**	**Muito**
1. O meu coração bate rápido	**1**	**2**	**3**	**4**
2. Os meus músculos estão tensos	**1**	**2**	**3**	**4**
3. Sinto-me agoniado com os meus problemas	**1**	**2**	**3**	**4**
4. Eu penso que os outros não me aprovarão	**1**	**2**	**3**	**4**
5. Eu sinto que me vou perdendo porque não consigo decidir-me atempadamente	**1**	**2**	**3**	**4**
6. Sinto-me tonto	**1**	**2**	**3**	**4**
7. Sinto fraqueza nos meus músculos	**1**	**2**	**3**	**4**
8. Sinto-me trémulo e instável	**1**	**2**	**3**	**4**
9. Eu perspetivo algumas desgraças futuras	**1**	**2**	**3**	**4**
10. Não consigo tirar alguns pensamentos da minha cabeça	**1**	**2**	**3**	**4**
11. Tenho dificuldade em relembrar coisas	**1**	**2**	**3**	**4**
12. Sinto a minha face quente	**1**	**2**	**3**	**4**
13. Eu penso que o pior vai acontecer	**1**	**2**	**3**	**4**
14. Sinto que os meus braços e pernas estão hirtos	**1**	**2**	**3**	**4**
15. Sinto a garganta seca	**1**	**2**	**3**	**4**
16. Eu esforço-me a evitar pensamentos desconfortáveis	**1**	**2**	**3**	**4**
17. Não me consigo concentrar sem a intrusão de pensamentos irrelevantes	**1**	**2**	**3**	**4**
18. A minha respiração é rápida e superficial	**1**	**2**	**3**	**4**
19. Preocupo-me por não conseguir controlar os meus pensamentos tão bem como eu gostaria	**1**	**2**	**3**	**4**
20. Tenho borboletas no estômago	**1**	**2**	**3**	**4**
21. Sinto as palmas das mãos húmidas	**1**	**2**	**3**	**4**

PANAS

(Watson, Clark, & Tellegen, 1988; VP: Galinha & Pais-Ribeiro, 2005)

Esta escala consiste num conjunto de palavras que descrevem diferentes sentimentos e emoções. Leia cada palavra, e utilize a escala que apresentamos, assinalando o número que melhor indica em que medida sentiu cada uma das emoções **durante as últimas semanas**.

	Nada ou muito ligeiramente	**Um pouco**	**Moderadamente**	**Bastante**	**Extremamente**
1. Interessado	**1**	**2**	**3**	**4**	**5**
2. Perturbado	**1**	**2**	**3**	**4**	**5**
3. Excitado	**1**	**2**	**3**	**4**	**5**
4. Atormentado	**1**	**2**	**3**	**4**	**5**
5. Agradavelmente surpreendido	**1**	**2**	**3**	**4**	**5**
6. Culpado	**1**	**2**	**3**	**4**	**5**
7. Assustado	**1**	**2**	**3**	**4**	**5**
8. Caloroso	**1**	**2**	**3**	**4**	**5**
9. Repulsa	**1**	**2**	**3**	**4**	**5**
10. Entusiasmado	**1**	**2**	**3**	**4**	**5**
11. Orgulhoso	**1**	**2**	**3**	**4**	**5**
12. Irritado	**1**	**2**	**3**	**4**	**5**
13. Encantado	**1**	**2**	**3**	**4**	**5**
14. Remorsos	**1**	**2**	**3**	**4**	**5**
15. Inspirado	**1**	**2**	**3**	**4**	**5**
16. Nervoso	**1**	**2**	**3**	**4**	**5**
17. Determinado	**1**	**2**	**3**	**4**	**5**
18. Trémulo	**1**	**2**	**3**	**4**	**5**
19. Ativo	**1**	**2**	**3**	**4**	**5**
20. Amedrontado	**1**	**2**	**3**	**4**	**5**