HORUS: an Emotional Recognition Tool

André Teixeira
Informatics Department/ Computer Science and Tech. Centre
University of Minho
Braga, Portugal
andre.f.teixeira@outlook.com

Manuel Rodrigues, Davide Carneiro
CIICESI, ESTG-IPP - School of Technology and Management
Polytechnic of Porto
Felgueiras, Portugal
{mfsr, dcarneiro}@estg.ipp.pt

Paulo Novais
Informatics Department/ Computer Science and Tech. Centre
University of Minho
Braga, Portugal
pjon@di.uminho.pt

Abstract--Emotion is an essential part of what means to be human, but it is still disregarded by most technical fields as something not to be considered in scientific or engineering projects. However, the understanding of emotion as an aspect of decision-making processes and of modelling of human behavior is essential to create a better connection between humans and their tools and machines. With this work we focus on the measurement of emotion of users through the use of non-intrusive methods, like measuring inputs and reactions to stimuli, along with the creation of a tool that measures the emotional changes caused by visual output created by the tool itself. Usage of the tool in a test environment and the subsequent analysis of the data obtained will allow for conclusions about the effectiveness of the method, and if it is possible to apply it to future studies on human emotions by investigators in the fields of psychology and computation.

I - INTRODUCTION

Emotion is a key aspect of the Human psyche. Although humans can be considered rational and logical creatures there are many options and attitudes people choose to follow that cannot be understood if the emotion of the individual in question is not taken into account. More than a detail or peripheral aspect of our mentality, emotion is a vital part of why we do what we do, why we are what we are. Emotion binds us together and breaks us apart, emotion can ravage the greatest of logical minds and break the greatest of strategists, and it can give rise to the best and the worst of any given individual. As such an essential aspect of the causes and reasons behind human behavior this phenomenon must be better understood. Initially relegated to the work of natural thinkers due to its immaterial nature and difficulty to measure, the emotive part of humans is now being researched by more precise and exact sciences, since there is now a possibility to use modern technology to try and measure this phenomenon, be it through behavior analysis be it through biological changes [1]. This phenomenon is defined by the American Psychological Association as a complex pattern of changes, including physiological arousal, feelings, cognitive processes, and behavioral reactions, made in response to a situation perceived to be personally significant. Emotion, including its type and intensity, is entirely dependent on stimuli to the individual, be it external like a smell or sound or internal like a memory. Also, that stimuli must have personal meaning for the subject. As such, emotion is something that can be studied in a non-intrusive way through the application of stimuli to individuals, and subsequent analysis of response and reaction to that stimuli. By including the measurement of the user's emotion, a system would be able to adapt to any single user in a way never seen until now, since most systems are completely blind in this regard. The idea is then to create a system based on existing psychology models of emotion that can induce emotive responses on the subject through visual stimuli and create a model of their interaction with the machine that can be used to measure the subject's emotion in different situations. The creation of such models would allow for a better understanding of emotional changes in individuals, therefore being an asset in both the field of psychology and informatics. If one could detect emotional changes in users through non-intrusive means, that is, without capturing live or past data on the user's being, interests or activities, based solely on the user's interaction with the machine itself in the current session of usage, then there could exist high potential to profile users and their actions and therefore respond to them in real time, seamlessly. This has already been done related to stress for instance [2,3]. Since it is however impossible to predict behavior without past cases the usage of images to provoke emotion and variation in the performance of the user in mundane tasks came forth. If subjected to an image that is expected to create a specific emotion the user will react in a certain way, changing its behavior in response to that emotion. If the system knows the emotion the image is associated with then it can detect the behavior detected when such emotion is present, therefore detecting the emotion itself. This work proposes an approach to essentially detect emotion using non-intrusive methods.

II - EMOTION

Emotion is a difficult thing to describe or define. All across human history it has been present as a vital component of our psyche, a part of what makes us human. "The only thing certain in the emotion field is that no one agrees on how to define emotion", [4]. In psychology, emotion is often defined as a complex state of feeling that results in physical and psychological changes that influence thought and behavior. Emotionality is associated with a range of psycho-logical phenomena, including temperament, personality, mood, and motivation.

Emotion can be hidden, and it can change, it can evolve and it manifests in different ways for different people, but it cannot be avoided and cannot be disabled. It stands as the basis of expressionism, of artistic creation and vision, even of scientific advancement, as it often is traits and desires of certain individuals that fuel discovery, not the benefits of the research itself.

Emotion as a psychic phenomenon of reaction to internal or external stimuli doesn't seem to change with time, at least not during recorded human history, but our ways of dealing with it and expressing said emotion are highly variable and almost completely social-oriented. There is very little done to integrate emotion in the more technical fields of study and research, which is understandable due to the difficulty of defining it as a palpable and usable variable. Only recently has emotion begun to be studied in fields like neuroscience and artificial intelligence, but it tends to be used more as something one wishes to evoke than something one wishes to work with or adapt to. There are still many models of study of emotion that are extremely relevant to anyone trying to better comprehend the subject at hand, even if most of these have no connection to any hard science [5].

There are many ways of exploring emotion, but most can be divided in one of three general categories, namely physiological, neurological and cognitive [6, 7].

Physiological theories suggest that responses within the body are responsible for emotions. **Neurological** theories propose that activity within the brain leads to emotional responses. Finally, **cognitive** theories argue that thoughts and other mental activity play an essential role in forming emotions. Table 1 summarizes some of the major theories and models related to emotions

Table 1 Emotion Related Theories

Evolutionary Theory of Emotion	emotions exist because they serve an adaptive role, they motivate people to respond quickly to stimuli in the environment, which helps improve the chances of success and survival.	does no contemplate the complexities of one's psyche and the variations that emotion can experience from person to person.
James-Lange Theory of Emotion	emotions occur as a result of physiological reactions to events. Witnessing an external stimulus leads to a physiological response. This connects emotion to bodily function in a structural level we experience emotions due to changes in our bodies.	fails to take into account self-inflected emotion, caused not by outside stimuli but by our own thoughts.
Cannon-Bard Theory of Emotion	It represents a neurobiological approach. people can experience physiological reactions linked to emotions without actually feeling those emotions emotion are not caused by physical reactions but are experienced in conjunction to them.	fails to explain the origins of the emotion itself, the causes for the impulse that starts the whole experience.
Schachter-Singer Theory of Emotion	physiological arousal occurs first, and then the individual must identify the reason for this arousal to experience and label it as an emotion. A stimulus leads to a physiological response that is then cognitively interpreted and labelled	the first theory explored until know that accepts the possibility of emotion being also a cause of actual personal choice and experience, albeit a subconscious one.

	which results in an emotion.	
Cognitive Appraisal Theory of Emotion	thinking must occur first before experiencing emotion.	it does not approach the subject on a manner scientific enough to justify enough consideration from the more technical part of the scientific community

III - THE IAPS

There are obviously many different models of studying, describing and measuring emotion, but none of these are directly related to computation or the proposed method of connecting images to specific emotional output in order to model an individual's emotional behavior.

The International Affective Picture System (IAPS) is a system currently being developed to provide ratings for a large set of emotionally-evocative, internationally-accessible, color pictures that includes contents across a wide range of categories. The IAPS is product and property of the NIMH Centre for Emotion and Attention (CSEA) at the University of Florida [8]. The associated model also allows for the definition of any emotion in a three-dimensional space, having in account values in three different parameters to define specific emotions. Its goal is to provide standardized materials that are available to researchers in the study of emotion and attention. The existence of these collections of normatively rated affective stimuli should allow better experimental control in the selection of emotional stimuli, facilitate the comparison of results across different studies conducted in the same or different laboratory, and encourage allowed exact replications within and across research labs who are assessing basic and applied problems in Psychological science.

This system was developed based on the assumption that emotion can be defined by a coincidence of values on a number of different strategic dimensions. This view is founded in Osgood's [9], ground-breaking work with the semantic differential, in which factor analysis conducted on a wide variety of verbal judgments indicated that the variance in emotional assessments were accounted for by three major dimensions. Dimensional views of emotion have already been advocated by a large number of theorists through the years, by the likes of [10], [11],or even going back to Wundt in 1898 [12].

The IAPS is then a collection of normatively rated affective stimuli, considering three dimensions: affective valence, ranging from pleasant to unpleasant, arousal, ranging from calm to excited, and dominance, ranging from controlled to in-control. These three dimensions, rated in a scale of points from 1 to 9, allow for the measuring of emotion in a more complex level than mere substantive denominations of "sad" or "afraid", since these are often highly subjective and simply not very scientific at all. This scale can be used to establish a three-dimensional model of a subject's emotional reactions to the stimuli produced by the researcher, a model that can afterwards be applied in the creation of personalized responsive systems that can benefit from the new knowledge of the user's state of mind at any given moment, knowledge gained through the user's actions and patterns of reaction. This way there is both no need to comment on the user's emotion using substantives nor to inquire about the meaning and purpose behind the actions the user takes. The machine knows the user has a certain emotional state in that particular moment due to that particular interaction and reacts to it accordingly without any need or intention of knowing what caused said reaction. The training method of exposing the subject to a collection of images is exactly that, a training method, unavoidable but temporary. The creation of the model necessitates knowledge of the causes of the emotional variations, and the image set of the IAPS model allows for that knowledge to be unobtrusive and standard. The user's habits will not be monitored, only their reaction to the aforementioned pictures, and even so all monitoring of content will end after the model is created. The use of this model allows for a dissociation between content and reaction to the content, necessary to prove the point of the detection of emotions through means that do not blatantly break the trust and privacy of the user. As such, these methods are invaluable if implemented in everyday products, even if the training phase can be a little overbearing for the casual consumer.

IV - EMOTIONAL RECOGNITION TOOL

The developed tool was nicknamed "Horus" in the name of the Egyptian god whose all-seeing eyes could see even a person's soul, relating to the intended purpose of measuring emotion through interaction. The tool's purpose is not to create models of emotion itself, but to collect data with which to build those models, as well as validate the data collected in order to ensure that a model construct with this data would be sufficiently exact.

The tool possesses a simple interface with clearly defined buttons that can be used to proceed to the subsequent window. When using the creation part of the tool, chosen through the main window, the user can select multiple images from a list based of a local image folder, where images can be dropped off as necessary. After selecting the chosen images and typing a name for the test, the test is created in the database and the user is returned to the main window. When using the test taking part of the tool, also accessible through the main window, the user can then choose a test to take from a list of previously created tests, taken from the connected database.

After selecting the test, one wishes to take the user goes through a series of windows with images selected for the test, and is asked to describe the image presented and evaluate the user's own emotional reaction to the image using the three dimensions of emotion defined in IAPS methodology.

This information is saved each time the user presses the next button to proceed to the next picture, until the end of the test. Unbeknownst to the user there are other types of data being saved, relating to the user's interaction with the system, which will be later explained in detail. As such, the developed tool can be described as a simple interface of visual stimuli testing, with a focus on allowing for easy and simple creation of tests and the even simpler taking of such tests. The data being recorded is not shown to the user and is saved in the database as individual test pieces related to the image shown and the test itself, allowing for simple access for posterior data parsing and analysis. The main objective of this application is therefore the collection and the treatment of data.

This tool is very dependent of the IAPS system of image evaluation with which it was built and is therefore unable to be used with studies that do not take this system in mind.

The test format itself is quite simple, consisting of a series of images (from IAPS database) being shown to the user in quick succession. The number of images and the particulars of which one depend only on the will of the test creator and can be manipulated to achieve different goals and study different aspects of human emotion, according to the IAPS three-dimensional view.

The user must describe the image in the text box in a quick but expressive manner and evaluate the way they feel about the image using the three-dimensional sliders. The objective of the test is always to measure the way the user interacts with the tool itself, through a number of non-intrusive methods not immediately obvious to the blind eye.

As such what the user writes in the text box is not nearly as important as the way the user writes it. In this same vein the levels assigned by the user in the sliders are not nearly as important as the way they move the mouse, the speed they use in their interactions, the time they take to complete each section of the test. This means that the basic explanation given to the user must be a slight misdirection, to allow for unadulterated data to be collected. The user must believe that what they are writing and evaluating is meaningful for the experience, if not only to ensure their focus on the task at hand.

A. Data Collected

There are many types or data being collected during the test. This data is what allows the study of the user's interactions, and obtaining it is therefore the main goal of the tool itself. Previous studies done in [13,14,15] included a tool to collect this kind of data non-intrusively. Each of the types of information listed below are collected for each of the images in a test, being directly related to the window they are collected from. Each time the user presses the" next" button the window's information is stored, and all of the variable created to store the data below are reset and readied for new insertions. As such these values are directly connected to the images, which is of course the point of this constructions since the relations between the values and the psychiatric evaluation of the images are what will be used to obtain conclusions that are hopefully useful. The data type of data collected are:

Text: Although the main objective of the test is not to evaluate what the user writes, since that would be a bit too intrusive according to the motives behind this whole venture, a lot of data can be collected through text processing that does not include the comprehension of the message in the text. As such, and in order to allow for different types of study, the text the test taker writes in the text box is saved in its entirety;

Valence Slider: The first slider refers to Valence and goes from unpleasant (1) to pleasant (9). This dimension encompasses how much the user likes the displayed image, and is the easiest to assess, be it by the investigator or by the test taker himself. The evaluation the user makes is not the most relevant, but this value is nevertheless stored each time;

Arousal Slider: The second slider refers to Arousal and goes from calm (1) to excited (9). This dimension encompasses the level of energy the image causes on the user. Arousal is classically associated with eroticism, but is in this case not limited by this view, being more of a evaluation of attractiveness of the image relating to the user personally. This is again not too relevant to the test but always saved;

Dominance Slider: The third and final slider refers to Dominance or Control and goes from controlled (1) to in control (9). This dimension encompasses the level of stress and discomfort the image causes, since human beings feel in control when they are comfortable with their surroundings or present situation and under stress when they are not. This data is saved with the rest of the sliders;

Time Elapsed: This data contains the time between the moment a new image window is loaded until the next button is pressed, encompassing the entirety of the time, in seconds, the user spends on a particular image. This information is crucial since it is simple to make connections between the time a user spends in a particular type of images with particular 3D emotion values compared to others with different parameters;

Time Between Key Presses: This data is stored as a list of millisecond periods, representing the time spent between key presses on the keyboard by the user. The time between these keyboard presses can be representative of the state of mind of the test taker, and as such is valuable information in the field of keyboard dynamics;
Time Between Mouse Clicks: Similarly, to the previous one, this type of data is composed of a list of millisecond periods between left mouse button clicks, which are usually connected to relevant actions when in the context of a computer application. This time between clicks is also extremely relevant in the vein of mouse dynamics;
Mouse Click Duration: Still in the way of mouse dynamics, this list of millisecond values represents the time the left mouse button is pressed down. This period begins when the button is pressed down and ends when it is released upwards. It can also be relevant to ascertain acuteness of though and emotion;
Key Press Duration: Same as the previous one but in the area of keyboard dynamics, this list of millisecond values represents the time between the moment one key of the keyboard is released and the moment a new one is pressed, being directly related to the behaviour of the individual writing on the text box;
Number of Backspace Presses: This single integer value represents the number of grammatical mistakes or errors done by the user during the typing process. These mistakes can have many causes, from the usual misspelling of a word to the rethinking of a whole sentence, which gives the possibility of creating a connection between the error and the image;
Number of Characters in Text: Easily obtained by counting the number of letters and symbols in the text, this integer value can be representative of the length of the message written and therefore of its potential as a carrier of emotional burden;
Time Spent Typing: This value comes from a timer that starts when the test taker begins typing and stops when the last character is inputted, representing the total time spent on the text box itself, which may vary greatly from picture to picture;
Mouse Movement Speed: This final mouse dynamic indicator is a list of pixels per millisecond values measured every fifty milliseconds by calculating the location of the mouse pointer and the time it took to get there from the previous location on the screen. These speed values might vary greatly, especially when the user is thinking, but may give important information about the user's readiness and assertiveness.

Each of these items are essential for the understanding of the interactions between the test subjects and the machine providing the test. The expectation is that through the study of this data some kind of pattern and model can be established that can allow for the creation of future ways to understand the user´s emotion through nothing more than their interactions with the peripherals and the computer's software itself.

B. Tests and Data Collecting

The main objective of the testing phase was to obtain sufficient data to articulate a hypothesis regarding the effectiveness of the tool and specifically of the usage of this kind of data collection in the identification of the emotions of users. In order to do so it was necessary to gather a study group large enough to produce a relevant amount of data. It was determined that at least twenty test subjects were necessary in order to generate enough data to create an effective study. It also was estimated that the test would have to include at least twenty images in order to create a strong enough impression on the user so that their emotional output fluctuated enough to detect.
However, since many aspects of emotion could be studied through the developed tool it was decided that the dimension of Valence would be the focus of the study, since Valence, or pleasantness in other words, is the dimension easier to influence through visual stimulation and therefore the one that would most likely create interesting results in this planned test. In order to obtain strong data regarding the Valence dimension two sets of images were chosen from the IAPS database, namely the ten images with the highest Valence values and the ten images with the lowest Valence values. The ten pictures with high values were to be shown first in order to ease the user into the test process, and the low value pictures were to be shown last, in order to create a lasting impression on the user strong enough to be expressed in their usage of the device.

The test data collected amounts to four hundred and eighty (480) test entry documents, which originates from tests taken by twenty-three (23) volunteers. Each of these documents involves thousands of records, especially in the mouse speed and time between clicks and presses categories.

Before a statistical treatment and subsequent definite conclusions could be made it was necessary to correct some initial issues with the data, as well as prepare it for analysis. The existence of 18 variables per document, some with hundreds of entries, make the data a bit larger, but still fragile in its ability to support decisive conclusions.

The data was processed using RStudio with its highly mathematical R programming language, a choice also motivated by its simple integration with MongoDB databases that would allow for seamless integration.

The first analysis done was a general summary of statistical information in the form of data like mean and median of each and every relevant data type, specifically the nine before mentioned measured dynamic values. the standard deviation and the KS Test [16] were also calculated for each of these. The deviation is a simple tool to understand and quantify the amount of variation or dispersion of the values present in the data set, since a low standard deviation indicates that the data points tend to be close to the mean of the set, while a high deviation shows that the existing data is spread out over a wider range of values. The Kolmogorov–Smirnov test, or KS test for short, was used in its two-sample facet, since it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples presented, in this case the high and low Valence collections. The test checks if two data sets come from the same distribution, and the important thing is to check its p-value, since one can reject the null hypothesis that the two samples were drawn from the same distribution if the p-value is less than the established significance level, in this case 0,05. These tests were performed for each and every of the nine main data types, for each of the two Valence collections, in order to gain a general understanding of the state of the data collected through the testing phase.

The data was compiled onto a table in order to allow for simpler analysis and comparison between values. The table contains the whole data given by the general summary, as well and the standard deviation and the p-value from the KS test. With this visualisation of the data we can already obtain some useful information about how the indicators are related, and take some conclusions about the relevance of said data types, table 2.

Table 2 – Compiled data

		Min	1st Qu.	Median	Mean	3rd Qu.	Max	S. Dev.	Ks Test p-value
Character Number	High Valence	4	14	21	26.14	34	79	17.21531	0.02884
	Low Valence	3	15.75	26	30.66	38.25	91	19.6656	
Backspace Presses	High Valence	0	0	1	2.048	3	10	2.543394	0.05069
	Low Valence	0	0	1	3.074	5	16	4.165282	
Time Elapsed	High Valence	10.92	24.01	34.34	38.56	49.67	93.22	18.39447	0.7668
	Low Valence	7.689	22.632	33.159	37.907	48.268	85.456	19.66392	
Type Duration	High Valence	532	4458	10562	12715	19079	42862	10083.26	< 2.2E-16
	Low Valence	263	5082	11058	14514	21770	50125	11800.24	
Click Duration	High Valence	1	132	133	421.8	617.5	2075	499.1917	6.00E-15
	Low Valence	11	121	133	235.5	181	1092	244.888	
Time Between Presses	High Valence	1	104	152	183.3	232	639	130.4182	2.89E-15
	Low Valence	1	72.25	143	167.88	219	595	124.3557	
Time Between Clicks	High Valence	11	425	1058	1439	1882	6689	1399.354	0.007886
	Low Valence	1	328	1080	1733	2100	9825	1990.433	
Press Time	High Valence	14	70	104	92.27	107	147	28.11155	0.05288
	Low Valence	13	70	105	92.62	107	147	29.43688	
Mouse Speed	High Valence	-0.0000071	0.03125	0.0662913	0.13517	0.1803281	0.7048763	0.1559297	< 2.2E-16
	Low Valence	0.01266	0.03226	0.08224	0.17407	0.23423	0.90253	0.2054756	

When it comes to the values of standard deviation, one can attain some conclusions on the dispersion of the data. First of all, in seven of nine cases the deviation is higher in the Low Valence collection that in the High Valence collections, something that cannot be ignored.

The lowest deviations can be found in the Mouse Speed values and the Backspace Presses values, although these can be justified due to the concrete nature of the values themselves, since the first is calculated in milliseconds and the second has a very small maximum and therefore a very small margin to verify divergence. One of the only conclusions that can be stated relative to the data types is that some data types like Click Duration and Time Between Clicks present significantly large differences between the two Valence collections, according to their means and standard deviation.

When it comes to the KS test, results are simpler to assess. Since the established significance level is 0,05 it can immediately be noted which data types possess p-values under this level. Once again seven in nine cases have a p-value well under the level, although not the same seven as before. Only Time Elapsed, Backspace Presses and Press Time have a p-value higher than the level, and in the case of the last two the difference is very small. Click Duration and Time Between Presses have values well under the level, with

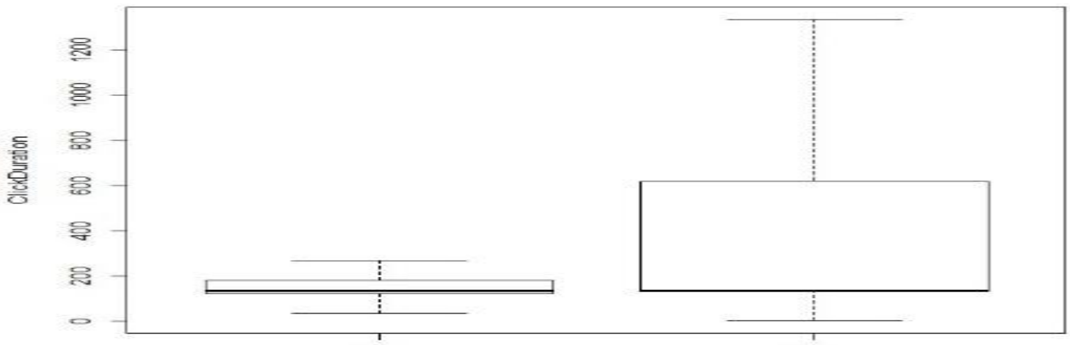


Figure 1 – Click duration box graph

ridiculously small values, and Type Duration and Mouse Speed have values so low that go beyond the minimum value calculated by the program, namely 2.2E-16. Although it has been stated that this type of scientific method of testing hypotheses by statistical analysis stands on a flimsy foundation due to the scientific community's over reliance on the p-value level of 0.05, it was still the best option to evaluate the data. As such these results tell us that the seven variables under the established level suffer significant differences between the two Valence collections, although the level of difference is not equal to the difference in p-value. Still, we can already tell that the three variables that have a higher p-value will most likely not be useful in the detection of emotions, since the difference between collections is so small. This goes hand in hand with the results obtained through the study of the rest of the elements in the table, even if not completely.

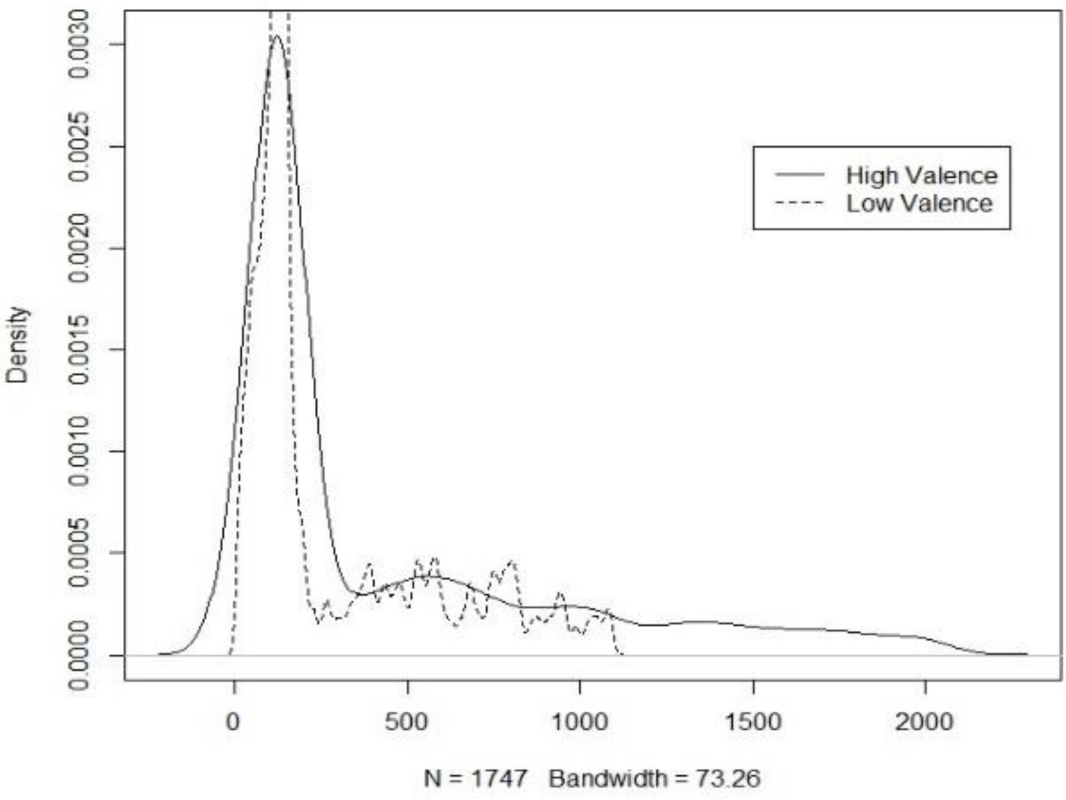
The detection of seven possible variables that provide relevant data is however encouraging, even if insufficient. Having done a general analysis, a more specific analysis individual to each data type was necessary. Therefore, that was an obvious next step.

V - RESULTS

The statistical analysis done before was useful but not determinant. Conclusions made about this matter need to be built on a strong foundation, a foundation built on individual analysis focused on each and every data type, in order to find if one or more of these are relevant in the detection and modulation of emotion in computer systems. As such a graphical comparison of the data types was necessary, since a graphical analysis can be critical to conclusions on comparisons of data collections.

The two types of graphs utilised were the Box Plot, or Box and Whisker Diagram, since it displays the distribution quite well with an attention to relevant outliers for as many collections as needed, and the Density Plot combination of graph, with the addition of lines for the collection not initially used. These choices were influenced by the necessity to observe the differences between data points on the same screen at the same time and for the high amount of easily obtainable information on the density and distribution of the data presented on these graphs.

All users that participated in the study were Portuguese citizens, with ages ranging from eighteen to fifty-one years old, and with considerable amounts of education. Their medium income privileged status as citizens of a democratic nation currently at peace means they have less resistance to disturbing imagery than someone used to it due to circumstance, necessity or other. These results are therefore fruit of the social and economic environment where the subjects reside and would be subject to significant change if the environment was a drastically different one. Even so, the data collected from the subject's interactions



was unadulterated and representative of their group, due to their willingness to comply with the test's necessities and to their status.

The evaluation of the variables has been done in three levels, namely Low, Medium and High relevance. From the nine variables considered relevant, three were proven mostly unchanged in the face of Valence changes, namely **Time Between Presses**, **Time Between Clicks** and **Press Time**. The analysis of these three indicators observed very little change, especially in the case of the third one, and due to their nature, the creation of a model using these values would be difficult. The third one of these also possessed a p-value higher than the expected level, which confirms its difficulty in establishing useful patterns. However, the small differences between Valence collections in these indicators might be enhanced if more test subjects were to be used in the test collection. Even so, these three are categorised as having Low relevance to the modelling of Valence emotional models. From the remaining six, four of the indicators were considered to have Medium relevance, namely **Backspace Presses**, **Character Number**, **Time Elapsed**, and **Type Duration**. Two of these, the first and the third, also had high p-values, and it is true that there are not many differences in an initial analysis, but both the amount of mistakes during writing and the amount of time spent in an image tend to vary sufficiently, after the graphical analysis. Even so, they might not be enough to confirm their usefulness, even if there is data to suggest it.

A higher amount of tests is mandatory in order to confirm their significance. As for the last two, namely **Mouse Speed** and **Click Duration**, they were evaluated as having High relevance, since there are very significant changes in the two collections, be it after the statistical analysis be it in the graphical individual one, figure 1 to 4. As such, we finish the analysis having confidence that two of the variables are highly likely to help detect changes in Valence, that four have a distinct possibility of being helpful in this

Figure 2 – Click duration plot

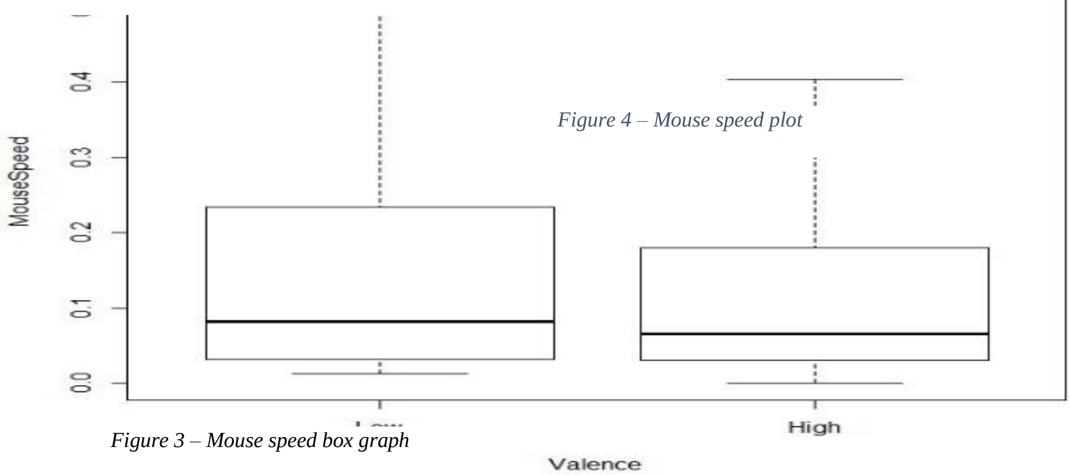
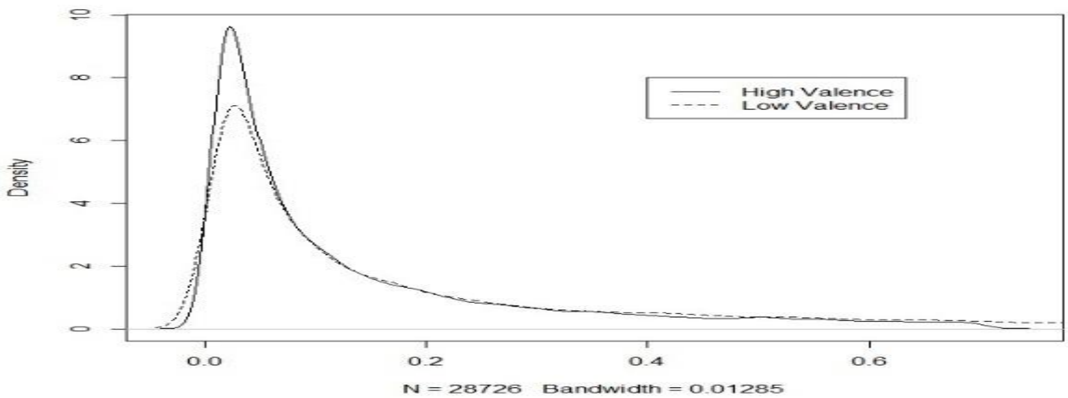


Figure 4 – Mouse speed plot

Figure 3 – Mouse speed box graph

endeavour and that three are most likely not as useful as the others. All these evaluations would benefit immensely from a second round of testing with a larger amount of data, but as it stands these are the results presented here through the analysis of the available data.

As a formulated hypothesis it is suggested that users tend to want to go through the disturbing images



at a faster rate but have difficulty in writing down their thoughts because of the nature of the images. Even though there are significant outliers there is enough data to support said hypothesis, even if this may only apply to situations where visual stimuli is presented. Keeping in mind that all these results come from exposing the users to visual stimuli, that the user's interactions were highly limited and monitored and that the amount of users themselves were rather small, the results are satisfactory when it comes in discovering patterns in interactions. The users emotional change, caused by the variation in Valence of the images, has indeed affected the way they interacted with the tool and with the computer, and has allowed for the identification of a group of variables that might help in the adaptation of systems to the emotional state of their users. As such, the experiment was successful in creating both a proof of concept and a starting point that can be used to continue studying the effects of emotions on interactions. The particulars of the Valence changes have also been sufficiently analysed to affirm that the method is sufficiently effective, even if not as much as one would want.

VI - CONCLUSIONS

This work had as its main objectives the better understanding of emotion and of its importance for intelligent systems, the creation of a different method of measuring emotion through non-intrusive means, the development of a tool that had the ability to do so and the exploration and treatment of data obtained through the tool, in order to judge the efficacy of the method. The study of human computer interaction and the role that emotion has and could have in the way people use and interact with technology in a daily basis has been very productive in both validating the use of emotion in intelligent systems and demonstrating the value of studying ways to implement that use. The idea to use non-intrusive methods to identify emotion, namely keyboard and mouse dynamics, was also validated as a valuable and possible answer to the dilemma, be it in its moral and intellectual facets be it its practical application. Respecting privacy is essential if we are to better integrate technology into our societies and doing so while adapting systems to their users' necessities is key in the vision of a better connected and more civil future. The designed tool was implemented successfully, according to the requisites and necessities recognized in early planning. It is completely functional, and it was capable of collecting data in the test phase, being utilized by a variety of users in a great variety of environments. The data collected was used to identify patterns and some indicators that could be used in the creation of models that are able to identify emotional Valence. This data was also used to validate an hypothesis on the Valence variations using non-intrusive methods, something that was essential in the context of what were the goals of the project from the beginning. The focus on a mixture of psychology and engineering as a way to identify and evaluate the methods and the data was essential in its success, and this association of different areas of study is something that must be more explored in the future, especially considering the way informatics engineering is sometimes not very available to do so. It is important to note that even if the data is enough to consider these hypothesis and methods as valid, and to take conclusion on the subject, the amount of test data is not enough to safely implement a system that depends on it, and therefore more in-depth testing with a larger pool of test subjects is required. The data types and indicators that were identified as having high relevance to the detection of emotion might also not be as useful in other dimensions of emotion that not Valence, and therefore testing dedicated to its other two dimensions is mandatory in order to have a decisive conclusion on the validity of the method as a whole. The developed tool would also benefit of more dedicated development and of an increase in functionality, in order to allow for more variety of users, tests and research projects to make use of it. Even so, this work is concluded with the certainty that a small but relevant step was given in the direction of better human computer interaction based on emotions, even with the amount of future work required to finally reach this so desired goal.

REFERENCES

[1] Picard, R. W. (2000). Toward computers that recognize and respond to user emotion. *IBM systems journal*, *39*(3.4), 705-719.

[2] Rodrigues, M. F., Gonçalves, S. M., Santos, R., Fdez-Riverola, F., & Carneiro, D. (2016). Intelligent Tutoring: Active Monitoring and Recommendation. In *Interdisciplinary Perspectives on Contemporary Conflict Resolution* (pp. 205-224). IGI Global.

[3] Carneiro, D., Rocha, H., & Novais, P. (2017, June). An environment for studying visual emotion perception. In International Symposium on Ambient Intelligence (pp. 238-245). Springer, Cham. [4] Parrott, W. G. (Ed.). (2001). *Emotions in social*

psychology: Essential readings. Psychology Press.

[4] Wierzbicka, A. (1992). Defining emotion concepts. Cognitive science, 16(4), 539-581.

[5] Wu, C. H., Huang, Y. M., & Hwang, J. P. (2016). Review of affective computing in education/learning: Trends and challenges. *British Journal of Educational Technology*, *47*(6), 1304-1323.

[6] Barrett, L. F., Lewis, M., & Haviland-Jones, J. M. (Eds.). (2016). *Handbook of emotions*. Guilford Publications.

[7] Myers, D. G. (2004). Theories of emotion. *Psychology: Seventh Edition, New York, NY: Worth Publishers*, *500*.

[8] Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, 39-58.

[10] Osgood, S., & Suci, G. J. (1988). Tannenbaum (1957). *The measurement of meaning*.

[11] Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. the MIT Press.

[12] Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological bulletin*, *98*(2), 219.

[13] Rodrigues, M., & Novais, F. F. R. P. (2012). An Approach to Assessing Stress in eLearning Students. *Proceedings of the 11th European Confernece on e-Learning: ECEL*, 461.

[14] Carneiro, D., Castillo, J. C., Novais, P., FernáNdez-Caballero, A., & Neves, J. (2012). Multimodal behavioral analysis for non-invasive stress detection. *Expert Systems* with Applications, 39(18), 13376-13389.

[15] Novais, P., Carneiro, D., Gomes, M., & Neves, J. (2012). Non-invasive estimation of stress in conflict resolution environments. In Advances on Practical Applications of Agents and Multi-Agent Systems (pp. 153-159). Springer, Berlin, Heidelberg.

[16] Goodman, L. A. (1954). Kolmogorov-Smirnov tests for psychological research. Psychological bulletin, 51(2), 160.